A thermodynamic atlas of carbon redox chemical space

Adrian Jinich[a,b,1], Benjamin Sanchez-Lengeling[a], Haniu Ren[a], Joshua E. Goldford[c], Elad Noor[d], Jacob N. Sanders[e], Daniel Segrè[c,f,g,h], and Alán Aspuru-Guzik[i,j,k,l,1]

[a]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; [b]Division of Infectious Diseases, Weill Department of Medicine, Weill-Cornell Medical College, New York, NY 10021; [c]Bioinformatics Program and Biological Design Center, Boston University, Boston, MA 02215; [d]Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland; [e]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; [f]Department of Biology, Boston University, Boston, MA 02215; [g]Department of Biomedical Engineering, Boston University, Boston, MA 02215; [h]Department of Physics, Boston University, Boston, MA 02215; [i]Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, ON, M5S 3H6 Canada; [j]Department of Computer Science, University of Toronto, Toronto, ON, M5T 3A1 Canada; [k]CIFAR AI Chair, Vector Institute, Toronto, ON, M5S 1M1 Canada; and [l]Lebovic Fellow, Canadian Institute for Advanced Research, Toronto, ON M5S 1M1 Canada

Edited by Pablo G. Debenedetti, Princeton University, Princeton, NJ, and approved November 17, 2020 (received for review April 9, 2020)

Redox biochemistry plays a key role in the transduction of chemical energy in living systems. However, the compounds observed in metabolic redox reactions are a minuscule fraction of chemical space. It is not clear whether compounds that ended up being selected as metabolites display specific properties that distinguish them from nonbiological compounds. Here, we introduce a systematic approach for comparing the chemical space of all possible redox states of linear-chain carbon molecules to the corresponding metabolites that appear in biology. Using cheminformatics and quantum chemistry, we analyze the physicochemical and thermodynamic properties of the biological and nonbiological compounds. We find that, among all compounds, aldose sugars have the highest possible number of redox connections to other molecules. Metabolites are enriched in carboxylic acid functional groups and depleted of ketones and aldehydes and have higher solubility than nonbiological compounds. Upon constructing the energy landscape for the full chemical space as a function of pH and electron-donor potential, we find that metabolites tend to have lower Gibbs energies than nonbiological molecules. Finally, we generate Pourbaix phase diagrams that serve as a thermodynamic atlas to indicate which compounds are energy minima in redox chemical space across a set of pH values and electron-donor potentials. While escape from thermodynamic equilibrium toward kinetically driven states is a hallmark of life and its origin, we envision that a deeper quantitative understanding of the environment-dependent thermodynamic landscape of putative prebiotic molecules will provide a crucial reference for future origins-of-life models.

origins of life | prebiotic chemistry | thermodynamics | redox biochemistry | systems chemistry

Redox reactions are fundamental to biochemistry. Recent work has uncovered quantitative thermodynamic principles that influence the evolution of carbon redox biochemistry (1–3). This line of work has focused on the three main types of redox reactions that change the oxidation level of carbon atoms in molecules: reductions of carboxylic acids (–COO) to aldehydes (–C=O); reductions of aldehydes and ketones to alcohols (hydroxycarbons) (C–O); and reductions of alcohols to hydrocarbons (C–C). The "rich-get-richer" principle states that more reduced carbon functional groups have higher standard redox potentials (1–3). Thus, alcohol reduction to a hydrocarbon is more favorable than aldehyde/ketone reduction to an alcohol, which in turn is more favorable than carboxylic acid reduction to an aldehyde. This explains why, across all six known carbon fixation pathways, adenosine triphosphate is invested solely (with ribulose-5P kinase as the single exception) to drive carboxylation and the reduction of carboxylic acid functional groups (2, 4). Quantitative analysis of biochemical redox thermodynamics has also explained the emergence of NAD(P) as the universal redox cofactor. With a standard redox potential of −320 mV, NAD(P) is optimized to reversibly reduce/oxidize the vast majority of central metabolic redox substrates (3). In addition, since its standard potential is ∼100 mV lower than that of the typical aldehyde/ketone functional group, it effectively decreases the steady-state concentration of potentially damaging aldehydes/ketones in the cell (3). Finally, other physicochemical properties like hydrophobicity and charge act as constraints that shape the evolution of metabolite concentrations (5).

Although the emergence of early self-reproducing systems requires the underlying chemistry to be out of thermodynamic equilibrium, a quantitative and comprehensive understanding of the underlying energy landscape would be very valuable. For example, high abundance of a given compound in a mixture of thermodynamically equivalent molecules could be ascribed to a kinetics-enabled, energy-driven process. However, the most likely scenario is that even at equilibrium, some compounds may be significantly more favorable than others, establishing the initial conditions for subsequent out-of-equilibrium processes. In addition, kinetics of catalysis and thermodynamics are highly intertwined, jointly contributing to effective reaction rates (6–8). Therefore, a comprehensive and quantitative understanding of the underlying thermodynamic landscape could help inform kinetic models of prebiotic redox chemistry.

Significance

Carbon redox chemistry plays a fundamental role in biology. However, the thermodynamic and physicochemical principles underlying the rise of metabolites involved in redox biochemistry remain poorly understood. Our work introduces the theory and techniques that allow us to quantify and understand the global energy landscape of carbon redox biochemistry. We analyze the space of all possible oxidation states of linear-chain molecules with two to five carbon atoms and generate a detailed atlas of the thermodynamic stability of metabolites in comparison to nonbiological molecules. Although the emergence of life required the underlying chemistry to bootstrap itself out of equilibrium, a quantitative understanding of the environment-dependent thermodynamic landscape of prebiotic molecules will be extremely valuable for future origins of life models.

Author contributions: A.J., B.S.-L., J.E.G., D.S., and A.A.-G. designed research; A.J., B.S.-L., H.R., and J.E.G. performed research; E.N., J.N.S., and D.S. contributed new reagents/analytic tools; A.J., B.S.-L., H.R., and J.E.G. analyzed data; and A.J., J.E.G., D.S., and A.A.-G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: adj2010@med.cornell.edu or aspuru@utoronto.ca.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2005642117/-/DCSupplemental.

First published December 21, 2020.

Downloaded by guest on January 23, 2021

Despite these motivations, the thermodynamic and physicochemical principles underlying the rise of carbon redox biochemistry remain very poorly understood. Here, we combinatorially generate the chemical space of all possible redox states of linear-chain n-carbon compounds (for $n = 2$ to 5). We partition each n-carbon linear-chain redox chemical space into biological metabolites and nonbiological compounds and systematically explore whether metabolites involved in biochemical redox reactions display features that would be unexpected elsewhere in redox chemical space. To compare physicochemical and thermodynamic properties of the biological and nonbiological molecules, we use cheminformatic tools and a recently developed quantum-chemical approach to estimate standard reduction potentials ($E^{o\prime}$) (3) of biochemical reactions.

In addition to generating a molecular energy landscape of broad applicability to the study of biochemical evolution, our analysis provides specific insight on redox biochemistry, which we summarize in the following five major conclusions: 1) the oxidation level and asymmetry of aldose sugars makes them unique in that they have the highest possible number of connections (reductions and oxidations) to other molecules; 2) biological compounds (metabolites) tend to be enriched for carboxylic acid functional groups and depleted for aldehydes/ketones; 3) metabolites tend to have, on average, higher solubilities and lower lipophilicities than the nonbiological molecules; 4) across a range of pH and electron-donor/-acceptor potentials, metabolites tend to have, on average, lower Gibbs energies relative to the nonbiological compounds; and 5) by adapting Pourbaix phase diagrams—an important conceptual tool in electrochemistry—to the study of redox biochemistry, we find that the n-carbon linear-chain dicarboxylic acids and fatty acids (e.g., succinate and butyrate in four-carbon redox chemical space) are the local minima in the energy landscape across a range of conditions and thus may have a spontaneous tendency to accumulate. Our results suggest that thermodynamics may have played an important role in driving the rise of dominant metabolites at the early stages of life and yield insight into the principles governing the emergence of metabolic redox biochemistry.

Results

Aldose Sugars Have the Maximal Number of Redox Connections. We combinatorially generated all possible redox states of n-carbon linear-chain compounds (for $n = 2$ to 5 carbon atoms per molecule) and studied the properties of the resulting chemical spaces (Fig. 1). For every molecule in n-carbon redox chemical space, each carbon atom can be in one of four different oxidation levels: carboxylic acid, ketone or aldehyde, hydroxycarbon (alcohol), or hydrocarbon (Fig. 1*A*). Molecules in redox chemical space are connected to each other by three different types of two-electron reductions (or the reverse oxidations) that change the oxidation level of a single carbon atom: reduction of a carboxylic acid to an aldehyde; reduction of an aldehyde/ketone to a hydroxycarbon; and reduction of a hydroxycarbon to a hydrocarbon. In order to make the redox chemical space model system tractable to analysis, we decreased its complexity by excluding carbon–carbon bond cleavage/formation reactions (e.g., reductive carboxylations or oxidative decarboxylations), keto–enol tautomerizations, intermediate carbon–carbon double-bond formation, intramolecular redox reactions, or different stereoisomers for a given molecular oxidation level (see *SI Appendix*, Table S1 for how the number of compounds increases when these reactions are included). In what follows, we focus the majority of our analysis on the properties of the four-carbon linear-chain redox chemical space (see *SI Appendix*, Figs. S7–S14 for corresponding results in two-, three-, and five-carbon linear-chain redox chemical space).

The four-carbon linear-chain redox chemical space contains 78 molecules connected by 204 reactions. The molecules span 11 different molecular oxidation levels, from the fully oxidized 2,3-dioxosuccinic acid (2 carboxylic acids and 2 ketones) to the fully reduced alkane butane (Fig. 1*A*); 84 reactions reduce aldehydes/ketones to hydroxycarbons (or oxidize hydroxycarbons to aldehydes/ketones), and the same number reduce hydroxycarbons to hydrocarbons (or oxidize hydrocarbons to hydroxycarbons). Since carboxylic acids are restricted to carbon atoms at the edges of a molecule (i.e., carbon nos. 1 and 4 in 4-carbon linear-chain molecules), only 36 reactions reduce carboxylic acids to aldehydes (or oxidize aldehydes to carboxylic acids) (Fig. 1*C*).

The number of reactions that connect a molecule to its oxidized or reduced products—the redox degree of a molecule—ranges from 2 to $2n$ (Fig. 1*B*). In $n = 4$-carbon redox chemical space, we find that only a single molecule in the network, the aldose sugar erythrose (and its stereoisomers), has the maximal degree value of $2n = 8$. This holds true for all redox chemical spaces regardless of the number of carbon atoms: only the corresponding aldose sugars in the two-, three-, five-, and six-carbon redox chemical spaces have the maximal-degree value, $2n$ (Fig. 1*B*). This is explained by the fact that the n-carbon aldose sugar satisfies the two constraints required to have the maximal number of redox connections: 1) each atom must be in an "intermediate" oxidation level that can be both oxidized and reduced. Therefore, all inner carbon atoms (i.e., atom nos. 2 and 3 in four-carbon linear-chain molecules) must be in the hydroxycarbon oxidation level, while carbon atoms at the edges (i.e., atom nos. 1 and 4) can be either in the aldehyde or hydroxycarbon oxidation level. 2) The molecule must not be symmetric under a 180° rotation along its center. Thus, the two edge atoms must be in different oxidation levels. This leads uniquely to the aldose sugar molecular redox-state configuration.

Biological Compounds Are Enriched in Carboxylic Acids and Depleted of Aldehyde/Ketone Groups. What distinguishes the subset of compounds in redox chemical space that appear in cellular metabolism from those that do not? To address this question, we subdivided the 78 molecules from the full 4-carbon redox chemical space into 30 biological compounds (also referred to from here onward simply as metabolites or "natural" compounds), which were identified based on matches with Kyoto Encyclopedia of Genes and Genomes (KEGG) database entries (9, 10), and the remaining 48 "nonbiological" compounds (Fig. 2*A*). Compounds in KEGG that correspond to molecules in redox chemical space but have alcohol groups substituted by amines or phosphates were considered a match, as these functional groups have the same oxidation level (see *Materials and Methods* for further details). For example, the metabolites oxaloacetate and aspartate have the same oxidation level at every carbon atom but differ by the substitution of an alcohol into an amine; both are considered a match to the corresponding molecule in our network. Similarly, we consider metabolites with carboxylic acid groups that are activated with either thioesters or phosphates groups as matches to molecules in redox chemical space.

As a first comparison between metabolites and nonbiological compounds, we analyzed the enrichment or depletion of functional groups (i.e., carbon atom oxidation levels) in the two categories (Fig. 2*B*). Specifically, we counted the number of times that each functional group appears in the set of metabolites and compared it to analytically derived expected null distributions for random sets of compounds (*Materials and Methods*). We found that in four-carbon linear-chain redox chemical space, metabolites are significantly enriched in carboxylic acids ($P < 0.001$) while being significantly depleted for ketones ($P < 0.001$) (Fig. 2*B*). We find similar trends in three- and five-carbon redox chemical spaces (*SI Appendix*, Figs. S8–S11). Since all but one molecule in two-carbon redox chemical space are biological metabolites, this space is not amenable to such statistical analysis. Furthermore, after normalizing for observed single functional-group statistics (see *Materials and Methods* for further details), we computed the null

Downloaded by guest on January 23, 2021

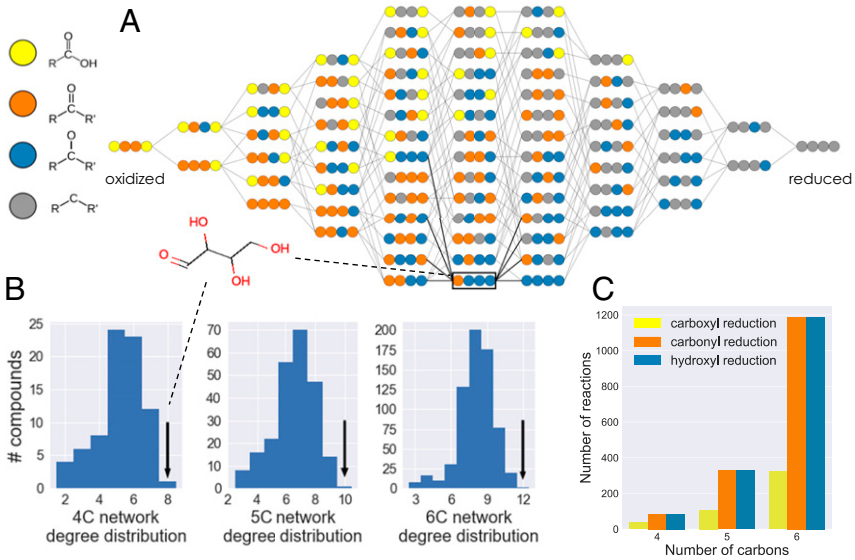

Fig. 1. The structure of *n*-carbon linear-chain redox chemical space. (*A*) The redox chemical space defined by the set of all possible four-carbon linear chain molecules that can be generated from three different types of redox reactions: reduction of a carboxylic acid to an aldehyde group; reduction of an aldehyde/ketone group to a hydroxycarbon (alcohol); and reduction of a hydroxycarbon to a hydrocarbon (and corresponding oxidations). Carbon atoms are represented as colored circles, with each color corresponding to an oxidation state: yellow, carboxylic acid; orange, aldehyde/ketone; blue, hydroxycarbon; and gray, hydrocarbon. Compounds within each column have the same molecular oxidation state and are organized from most oxidized (left) to most reduced (right). (*B*) The degree distributions for the four-, five-, and six-carbon linear-chain redox chemical spaces. In all cases, the aldose sugar is the only compound with the maximal number of possible reductions and oxidations (black arrows). (*C*) Number of reactions in the four-, five-, and six-carbon linear-chain redox chemical spaces that belong to each of the three types of redox reactions considered.

distributions for higher-order functional-group patterns, i.e., pair (2mer) and triplet (3mer) patterns (*SI Appendix*, Fig. S1). According to our analysis, only the 2mer pattern with a hydroxycarbon next to a hydrocarbon is depleted in the metabolites, albeit not significantly ($P = 0.05$). The number of times that all other 2mer and 3mer functional-group patterns appear in metabolites—including the highly uncommon dicarbonyl pattern—can be explained by the underlying single functional-group statistics.

We then asked whether the observed functional-group enrichments and depletions translate to differences in physicochemical properties of the metabolites and the nonbiological compounds. In particular, the significant enrichment of carboxylic acids is likely expected to result in higher average solubility and lower lipophilicity for the natural relative to the nonnatural compounds. However, given the fact that solubility and lipophilicity depend on the combination of all functional groups in a given molecule—with polar groups like carboxylic acids increasing solubility and nonpolar groups like hydrocarbons decreasing it—and that there is a enrichment (albeit not statistically significant) of hydrocarbons in the natural compounds, the combined effect of these two trends may be nontrivial. We therefore explicitly computed the solubility (logS) and lipophilicity (as captured by the octanol–water distribution coefficient [logD]) at pH 7. Indeed, we find that, in correlation with their enrichment for carboxylic acid functional groups, metabolites have significantly higher solubilities ($P < 0.005$) (Fig. 2*C*) and significantly lower logD values ($P < 0.01$) (*SI Appendix*, Fig. S2) than the set of nonbiological compounds. We observe similar trends in three- and five-carbon redox chemical spaces (*SI Appendix*, Figs. S9 and S13).

Metabolites Have on Average Lower Gibbs Energies than Nonbiological Compounds. We next focused on estimating the energy landscape of our redox compounds, with special attention to the question of whether metabolites and nonbiological compounds display different patterns in this landscape. We used a recently developed calibrated quantum-chemistry approach (3) to accurately predict the apparent standard redox potentials $E^{o'}$(pH) of all reactions in *n*-carbon linear-chain redox chemical space ($n = 2$ to 5). Previous work has shown that the calibrated quantum-chemistry method achieves significantly better accuracy than the group-contribution method (GCM) (3), the most commonly used approach to estimate thermodynamic parameters of biochemical compounds and reactions (11–14). Briefly, the quantum-chemistry method relies on density functional theory with a double-hybrid functional (15, 16) to compute the differences in molecular electronic energies and utilizes a two-parameter calibration against available experimental data. We computed the energies of several geometry-optimized conformations of the fully protonated species of each compound. We then estimated the standard redox potential E^o of the fully protonated species as the difference in electronic energies of the products and substrates, $\Delta E_{\text{electronic}}$. Using cheminformatic pK_a estimates (Marvin 17.7.0, 2017; ChemAxon) and the Alberty Legendre transform (17, 18), we converted the standard redox potentials to transformed standard redox potentials $E^{o'}$(pH), which depend on pH. Finally, in order to correct for systematic errors in the quantum-chemistry calculations and the cheminformatic pK_a estimates, we calibrated—via linear regression—the transformed standard redox potentials $E^{o'}$(pH) against a dataset of available experimental values (see *Materials and Methods* for further details).

We note that the improvement in accuracy of the quantum-chemical approach over GCM is particularly striking for the linear-chain compounds in our redox chemical spaces. This is most apparent for the set of aldehyde/ketone to hydroxycarbon reductions (*SI Appendix*, Fig. S3): while GCM prediction is no better than an average value predictor ($R^2 = -0.04$), the redox potentials predicted with the calibrated quantum-chemistry method correlate linearly with experimental values (Pearson $r = 0.50$). GCM accounts only for the difference in group energies of products and substrates to estimate redox potentials. Thus, for redox reactions, it effectively ignores the molecular environment surrounding the reduced/oxidized carbon atom, collapsing all of the potentials associated to aldehyde/ketone functional-group reductions to two

Downloaded by guest on January 23, 2021

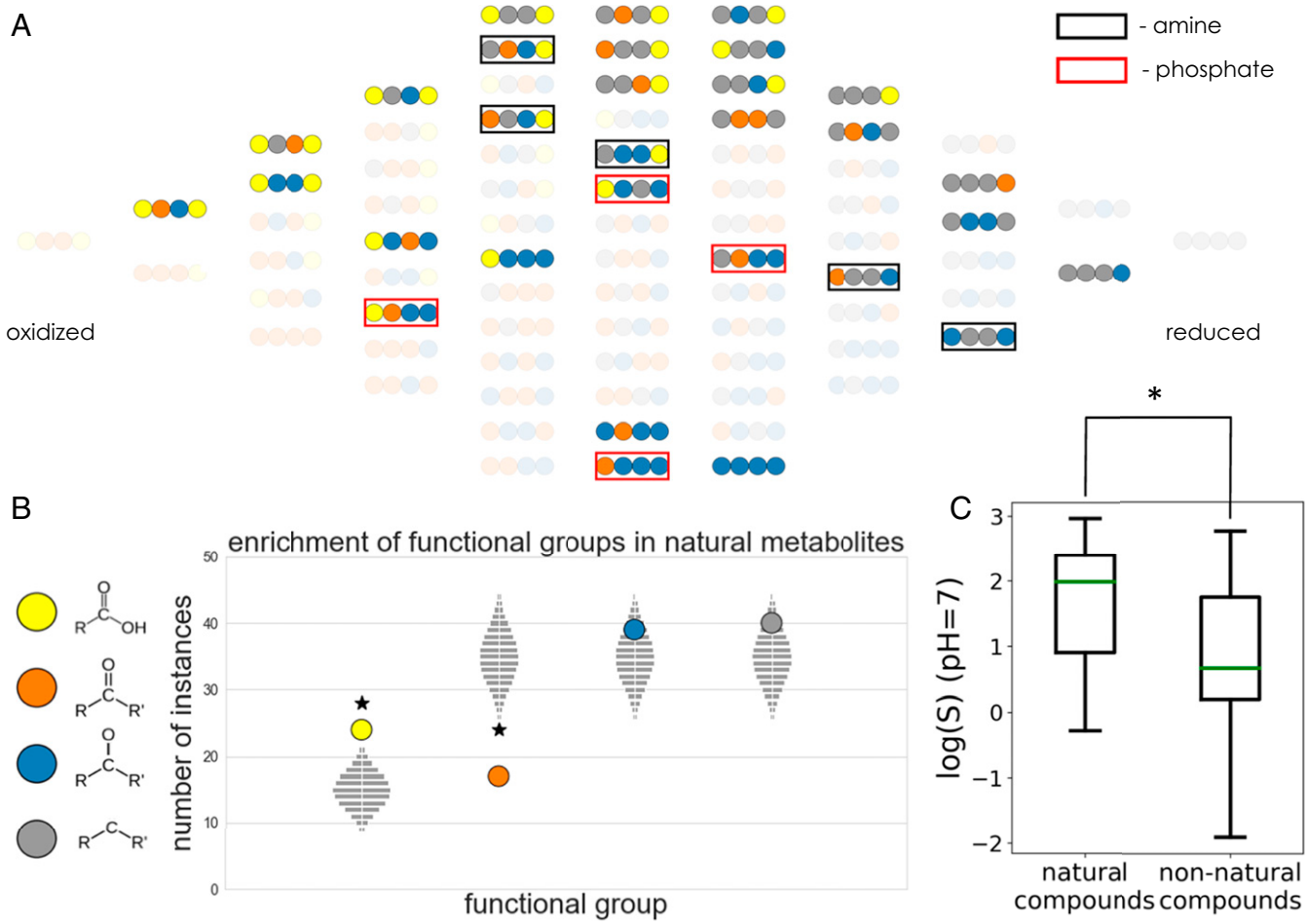

Fig. 2. Functional-group statistics and aqueous solubilities of biological compounds in the four-carbon linear-chain redox chemical space. (*A*) The subset of molecules in four-carbon linear-chain redox chemical space that matches biological metabolites in the KEGG database. Compounds that match KEGG metabolites but with alcohol groups substituted by either amines or phosphates are marked with black and red squares, respectively. (*B*) Enrichment and depletion of functional groups in the set of biological compounds. The vertical position of each colored circle corresponds to the number of times each functional group appears in the set of biological compounds. The light gray rectangles show the corresponding expected null distributions for random sets of molecules sampled from redox chemical space. See *SI Appendix*, Fig. S1 for statistical analysis of functional-group pairs and triplets. (*C*) Comparison of predicted aqueous solubility log(S) at pH 7 for biological and nonbiological compounds in the four-carbon linear-chain redox chemical space. Biological compounds have significantly higher solubilities than the nonbiological set ($P < 0.005$).

values (the average aldehyde and ketone reduction energies), thus lowering its prediction accuracy (*SI Appendix*, Fig. S3). Therefore, the use of our calibrated quantum-chemical method is essential in order to accurately predict and analyze the energetics of *n*-carbon linear-chain redox chemical spaces.

We used the predicted $E^{o'}$(pH) values to generate the energy landscape of redox chemical space. To do this, we assumed that each compound is coupled to an electron donor/acceptor with a given steady-state redox potential, *E*(electron donor). This potential could represent that set by a steady-state ratio of NAD^+/NADH or other abundant redox cofactor inside the cell (19). Alternatively, in the context of prebiotic chemistry, it could represent the potential associated with a given concentration of molecular hydrogen in an alkaline hydrothermal vent or different iron oxidation states in prebiotic oceans (20, 21). Given a value of *E*(electron donor), we convert the $E^{o'}$(pH) of each reaction into a Gibbs reaction energy, using $\Delta G_r(\mathrm{pH}) = -nF(E^{o'}(pH) - E(\text{electron donor}))$ (where n is the number of electrons and F is Faraday's constant). The set of Gibbs reaction energies for all redox transformations—as a function of pH and electron-donor potential—defines the energy landscape of our redox chemical space (Fig. 3*A* and *SI Appendix*, Figs. S5 and S6).

A notable finding of this analysis is that, across a range of cofactor potentials, metabolites in four-carbon linear-chain redox chemical space have, on average, significantly lower relative Gibbs energies than the nonbiological compounds (Fig. 3*B*). We find a similar trend for metabolites in three- and five-carbon linear-chain redox chemical space (*SI Appendix*, Figs. S10–S14); however, because of the few number of nonnatural compounds in three-carbon redox chemical space, the trend there is not statistically significant (*SI Appendix*, Fig. S10). An important exception to this general trend (and one that is conserved across spaces with different numbers of carbon atoms) is that the aldose sugars (e.g., erythrose), the ketose sugars (e.g., erythrulose), and the sugar alcohols (e.g., threitol) have a higher relative Gibbs energy than all compounds in redox chemical space across a large range of pH and electron-donor potential.

A Pourbaix Phase Diagram of Redox Chemical Space Maps Local Minimal-Energy Compounds. In addition to the trends observed for average energy differences between biological and nonbiological compounds, the relative energies of individual compounds change as a function of pH and *E*(electron donor) (*SI Appendix*, Figs. S5 and S6). To further investigate the detailed

Downloaded by guest on January 23, 2021

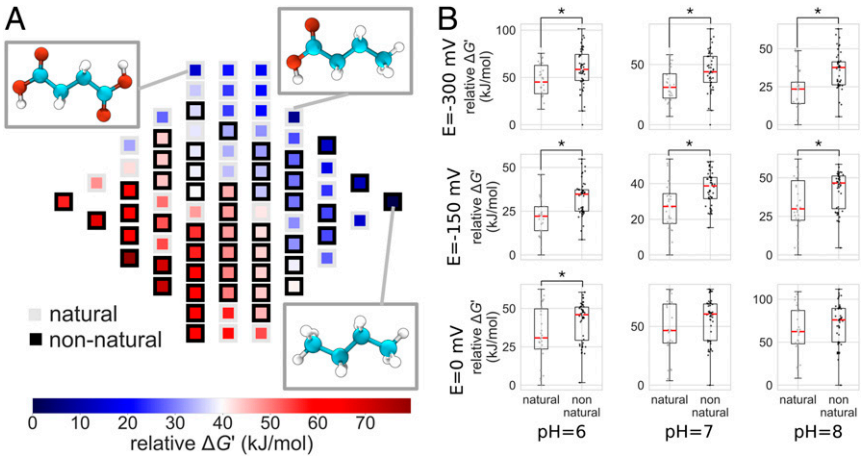

Fig. 3. Thermodynamic landscape of the four-carbon linear-chain redox chemical space. (*A*) Relative Gibbs energies of metabolites at pH 7 and *E*(electron donor/acceptor) = −300 mV. Gibbs energies are normalized relative to the metabolite with the lowest energy. Compounds within a column (same molecular oxidation state) are sorted from highest (bottom) to lowest (top) relative energies. The structures of the three compounds that are local minima in the thermodynamic landscape are shown: succinate (*Top Left Inset*), butyrate (*Top Right Inset*), and butane (*Bottom Right Inset*). These compounds have lower Gibbs energies than all of their neighboring molecules accessible by a reduction or oxidation. (*B*) Relative Gibbs energies of biological and nonbiological compounds for a range of pH and *E*(electron donor/acceptor) values. At each value of pH and *E*(electron donor/acceptor), Gibbs energies are normalized relative to the compound with the lowest energy. Asterisks indicate statistically significant differences of average values (Welch's *t* test; $P < 0.05$).

structure of the thermodynamic landscape, we set out to map which molecules are local minima at each value of pH and *E*(electron donor). A molecule is a local minimum in redox chemical space if its Gibbs energy is lower than that of all of its neighbors with whom it is connected through a reduction or an oxidation. We adapt Pourbaix phase diagrams, a powerful standard visualization tool in the field of electrochemistry (22), to the problem of mapping out regions of pH, *E*(electron donor) phase space in *n*-carbon linear-chain redox chemical space. In a Pourbaix diagram, the predominant equilibrium states of an electrochemical system and the boundaries between these states are mapped out as a function of two-phase space parameters. Fig. 4 shows a Pourbaix phase diagram representation of four-carbon linear-chain redox chemical space (see *SI Appendix*, Figs. S7, S8, and S12 show the corresponding Pourbaix diagrams for two-, three-, and five-carbon linear-chain redox chemical spaces).

At the lower left corner of the diagram in Fig. 4, in the region corresponding to more acidic pH values and more negative electron-donor potentials, the fully reduced four-carbon alkane butane is the only local (and the global) energy minimum (Fig. 4). Thus, assuming all compounds are kinetically accessible, butane would be expected to accumulate in these conditions. The structure of redox chemical space becomes richer as pH and *E*(electron donor) increase. Succinate and the four-carbon short-chain fatty acid (SCFA) butyrate—two biologically important metabolites—emerge as two additional local minima at more oxidative regions of the phase diagram (Fig. 4). Both succinate and butyrate consist of inner carbon atoms in the hydrocarbon (fully reduced) oxidation level and edge carbon atoms in either the hydrocarbon or the carboxylic acid (fully oxidized) state. Notably this pattern—where the *n*-carbon linear-chain dicarboxylic acid (oxalate, malonate, succinate, and glutarate for two-, three-, four-, and five-carbon atoms, respectively), the fatty acid (acetate, propionate, butyrate, and valerate), and the alkane (ethane, propane, butane, and pentane) emerge as the only local minima in a large region of phase space—is conserved in redox chemical spaces with different number of carbon atoms (*SI Appendix*, Figs. S7, S8, and S12). Further increases in either pH or electron-donor potential result in the emergence of additional compounds, both metabolites and nonbiological molecules, as local energy minima in the landscape (Fig. 4 and *SI Appendix*, Figs. S7, S8, and S12).

Can we predict from simple physicochemical principles the identity of the local minimal-energy compounds? A simple mean-field toy model (Fig. 4*C*) that focuses on the average standard redox potentials $\langle E^{o'}(\mathrm{pH})\rangle$ of the different carbon functional groups can help intuitively predict which metabolites accumulate at given values of pH and *E*(electron donor). Fig. 5, *Upper* shows the distributions of standard potentials at a fixed pH (pH 7) for all compounds in four-carbon redox chemical space categorized by the type of functional group undergoing reduction. Given a fixed value of *E*(electron donor), the average redox potentials for each functional-group category can be used to compute average Gibbs reaction energies for each

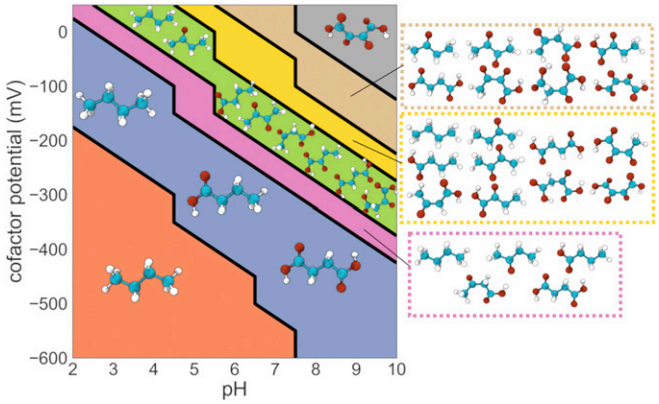

Fig. 4. Pourbaix phase diagram for the four-carbon linear-chain redox chemical space. Molecules that are local minima in the energy landscape at each region of pH, *E*(electron donor/acceptor) phase space are shown. At low pH and *E*(electron donor/acceptor) values, butane is both the global and the only local minimum-energy compound. At intermediate values of pH and *E*(electron donor/acceptor), several metabolites emerge as local minima and would thus tend to accumulate. For example, the metabolites oxaloacetate, acetoacetate, and α-ketobutyrate emerge as local energetic minima in the region of phase space shown in green. Finally, in the upper right corner of the phase diagram, characterized by higher values of both pH and *E*(electron donor/acceptor), the fully oxidized four-carbon compound 2,3-dioxosuccinic acid emerges as the only local (global) minimum.

Downloaded by guest on January 23, 2021

type of carbon redox transformation via the following equation: $\Delta G_r(\mathrm{pH}) = -nF(< E^{o'}(\mathrm{pH}) > - E(\text{electron donor}))$ (where n is the number of electrons and F is Faraday's constant). We use this to generate Fig. 5, *Lower*, which schematically shows the relative average Gibbs energies of the four different carbon oxidation levels at different values of E(electron donor) at pH 7. The boundaries delimiting different regions of the E(electron donor) axis mark the values where the rank-ordering of relative average Gibbs energies for the four carbon oxidation-level changes.

As an illustrative example, we focus on region III in Fig. 5, which is approximately delimited by the values $-360\ \mathrm{mV} \leq E(\text{electron donor}) \leq -180\ \mathrm{mV}$. In this region, the reduction of carboxylic acids to aldehydes (yellow to orange) is highly unfavorable (alternatively, the oxidation of aldehydes to carboxylic acids is highly favorable). On the other hand, the reduction of ketones to hydroxycarbons (orange to blue), as well as the reduction of hydroxycarbons to hydrocarbons (blue to gray), are thermodynamically favorable. Thus, at pH 7 and for this range of E(electron donor) values, the edge carbon atoms (atom nos. 1 and 4 in the four-carbon linear-chain compounds) are driven to either the most-oxidized (carboxylic acid) or the most-reduced (hydrocarbon) oxidation level, while the inner carbon atoms (atom nos. 2 and 3)—which cannot exist in the carboxylic acid oxidation level—are driven to the hydrocarbon oxidation level. This corresponds precisely to the molecular oxidation levels of dicarboxylic acids, fatty acids, and alkanes (e.g., succinate, butyrate, and butane), the local and global minimal-energy compounds in this region of pH, E(electron donor) phase space.

Discussion

In this work, we introduced the chemical spaces of all molecular oxidation levels of n-carbon linear-chain compounds and analyzed their structural, physicochemical, and thermodynamic properties. In some respects, our work is related to that of Morowitz et al. (23) in that it similarly explores the chemical properties of a portion of chemical space in search for possible traces of their relevance to the rise of biological systems. However, our work does not rely on chemical databases—which may be biased toward certain categories of compounds (24)—but rather analyzes a portion of chemical space in a way that is general, unbiased, and potentially arbitrarily extendible. In addition, while Morowitz et al. imposed a priori solubility, oil–water partition coefficients and combustion energies as criteria for the selection of metabolism-like organic compounds, our work systematically compares metabolites to the rest of the molecules in redox chemical space and demonstrates significant differences in these physicochemical parameters.

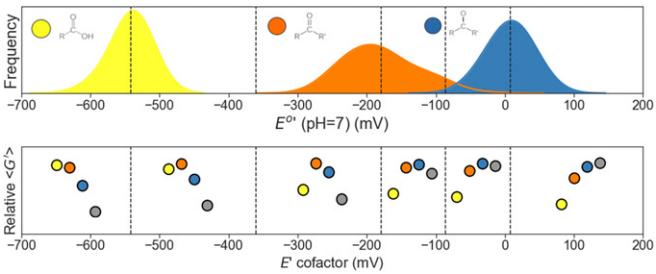

Fig. 5. A mean-field toy model explains the identity of molecular oxidation states of local minima. The schematic diagram illustrates how the average standard redox potentials of different carbon functional groups dictate the identity of the minimal-energy compounds. *Top* shows the distributions of standard redox potentials (pH 7) of all reactions in the four-carbon linear-chain redox chemical space, grouped according to the functional group that is reduced during the transformation: carboxylic acid (yellow), aldehyde/ketone (orange), and hydroxycarbon (blue). *Bottom* shows—for different values of E(electron donor/acceptor)—the resulting relative average Gibbs energies of the functional groups. For example, in the region where E(electron donor/acceptor) is between about −360 and −190 mV, aldehydes/ketones (orange) have, on average, the highest relative Gibbs energy, followed by hydroxycarbons (blue), carboxylic acids (yellow), and hydrocarbons (gray). Therefore, minimal-energy compounds will have inner carbon atoms (atom nos. 2 and 3 in four-carbon molecules) that equilibrate to the hydrocarbon oxidation state and edge carbon atoms (atom nos. 1 and 4) that equilibrate to either the carboxylic acid or the hydrocarbon oxidation state.

Our results are applicable and agnostic to "genetics-first" (25–28), or "metabolism-first" (29–32), scenarios for the early stages of life. Thermodynamics is a fundamental constraint in chemistry, and comprehensively and accurately mapping out the energy landscape of well-defined portions of redox chemical space should be of value to a wide variety of prebiotic chemistry models involving carbon redox chemistry. Thus, our capacity to estimate the thermodynamic landscape of molecules in a key region of the chemical space relevant to living systems could be equally valuable for metabolism-first studies aimed at characterizing possible avenues toward primitive organized autocatalytic networks, as well as for genetics-first studies exploring conditions conducive to the formation of nucleotides.

Examining the connectivity of redox chemical space, we found that aldose sugars—e.g., glyceraldehyde (n = 3), erythrose (n = 4), ribose (n = 5), and glucose (n = 6) and their corresponding stereoisomers—are unique in that they are the only compounds with the highest possible number of oxidative and reductive connections ($2n$) to neighboring molecules. Whether this maximal number of connections, coupled to their relatively high Gibbs energies, played a role in the emergence of aldose sugars as key players in cellular metabolism remains to be explored. One hypothesis is that a high degree value in redox chemical space could influence their role as central metabolic hubs in metabolic networks, optimally positioning them as high-energy substrates with many biochemical roads leading to a wide variety of compounds (33). However, a preliminary statistical analysis revealed that the connectivity of aldose sugars in the KEGG compound of metabolic reactions is not significantly high in comparison to other compounds (Dataset S1). It is also possible that, among the compounds with high energy, those accessible through multiple routes (i.e., with high connectivity) would have a higher chance of becoming incorporated into a rising metabolism. Although these interpretations are purely speculative, the quantitative knowledge provided by our work provides the basis for further exploration.

We found that the set of biological compounds is significantly enriched for carboxylic acid functional groups and have, on average, significantly higher solubilities (logS at pH 7) than the set of nonbiological compounds. In addition to an increase in aqueous solubility, other reasons for why carboxylic acids may have been selected during the evolution of metabolism potentially include a decrease in permeability across lipid membranes (34). This is reflected in the predicted values of octanol–water distribution coefficients, logD(pH 7) for the biological and nonbiological compounds (*SI Appendix*, Fig. S2). In addition, the enrichment for carboxylates may have enhanced the ability of enzymes to recognize small molecule substrates. Our analysis also showed that biological compounds are significantly depleted in aldehyde/ketone functional groups. Notably, in the four-carbon network, only one biological compound, diacetyl—which appears in the metabolic networks of yeast and several bacterial species (35, 36)—contains two ketone functional groups. This is consistent with the fact that aldehyde/ketone groups are significantly more reactive than carboxylic acids or hydroxycarbons and can cause oxidative damage, spontaneously cross-link proteins, inactivate enzymes, and mutagenize DNA (37).

Our analysis brings to the field of origin of life a recently developed quantum-chemical method for calculating biochemical redox potentials (3) that can have multiple applications for

Downloaded by guest on January 23, 2021

studying early biomolecules irrespective of any specific hypothesis on putative early pathways or on whether genes or metabolism emerged first. Our thermodynamic calculations, which rely on a recently developed calibrated quantum-chemistry approach that was shown to have better accuracy than cheminformatic methods (3), revealed that metabolites have, on average, lower Gibbs energies than the nonbiological set of compounds across a range of pH and electron-donor potentials. One could speculate on the relevance of these results in the context of the very early stages of the emergence of metabolism. In particular, there may have been a stage, prior to the rise of efficient catalysis, in which prebiotic carbon redox chemistry had not bootstrapped itself completely yet out of the equilibrating forces of thermodynamics. Under these circumstances, the equilibration process would have led to accumulation of a pool of compounds, which could serve as initial substrates or replenishing intermediates for downstream, kinetically controlled processes such as autocatalytic cycles or pathways that synthesized monomers for replicative polymers. Importantly, the identity of these thermodynamically favored and metabolite-enriched molecules depends on the values of the two environmental parameters (pH and environmental redox-couple potential), as shown in our Pourbaix phase diagrams. For example, low potential and acidic regions of phase space—e.g., E(electron donor) = −350 mV, pH 7—lead to accumulation of few metabolites (e.g., succinate and butyrate in four-carbon chemical space), whereas other regions of phase space—e.g., E(electron donor) = −200 mV, pH 7—would accumulate a more complex ensemble of compounds. Future work could explore whether particular ensembles of metabolite-enriched compounds (i.e., regions of phase space) are better primed to evolve into prebiotic pathways.

The resulting thermodynamic landscape also revealed in detail which compounds—both biological and nonbiological—are energetic local minima as a function of pH and E(electron donor) and would thus tend to accumulate. We find that some regions of phase space—such as those leading to the production of large quantities of volatile short hydrocarbons like butane and other alkanes—would be problematic for prebiotic chemistry. Indeed, although bacterial alkane production has been described (38, 39), the high volatility of these compounds likely limits their role in metabolism. However, as mentioned above, other regions of phase space would be expected to lead to complex metabolite-enriched mixtures of molecules that could provide a rich source of starting substrates for downstream kinetically controlled processes. Furthermore, spatial or temporal gradients of the environmental parameters—which have been proposed in the context of origins of life models (40, 41)—would open up the possibility of complex dynamic behavior, shifting the balance between different local minima. Finally, although our work dissects the energy minima of each isolated n-carbon redox chemical space, these are, in reality, embedded in a larger chemical space. Recent developments will help dissect and understand the properties of such expanded chemical networks (42, 43).

In the context of known living organisms, in the Pourbaix phase diagram representation of four-carbon linear-chain redox chemical space, the biological metabolites succinate and butyrate are local minima across a range of physiologically relevant pH and E(electron donor) values. Succinate is a key intermediate in the tricarboxylic acid cycle (TCA) cycle, with numerous recently elucidated signaling functions (44, 45). Interestingly, succinate accumulation occurs in a number of different organisms, including bacteria such as *Escherichia coli* (46), *Mycobacterium tuberculosis* (47), as well as several bacterial members of the human gut microbiome (48–51) and the bovine rumen (52–55); fungi such as the yeast *Saccharomyces cerevisiae* (56) and members of the genus *Penicillium* (57); green algae (58); parasitic helminths (59); the sleeping sickness-causing parasite *Trypanosoma brucei* (60); marine invertebrates (61); and humans (52–55). More specifically, our observations are consistent with the behavior of the TCA cycle under anaerobiosis and hypoxia (62–65). In these conditions, the reactions of the TCA cycle operate like an "incomplete fork," with a portion of the pathway running in a reductive ("counterclockwise") modality, i.e., oxaloacetate sequentially reduced to malate, fumarate, and succinate. Thus, despite the fact that in these examples, succinate is part of biochemical networks of higher complexity than our redox chemical space, its empirically observed accumulation is consistent with its identity as a local energy minimum. We also note that the SCFA butyrate accumulates to high (millimolar) levels in the gut lumen as a product of bacterial fermentation (66, 67). We found that this pattern—where the n-carbon linear-chain dicarboxylic acid and the fatty acid emerge as the only local minima in a large region of phase space—is conserved in the Pourbaix diagrams for redox chemical spaces with different numbers of carbon atoms (*SI Appendix*, Figs. S7, S8, and S12). Therefore, in analogy to succinate accumulation, it would be reasonable to search for evidence of n-carbon linear-chain dicarboxylic acid accumulation in different biological systems under physiological conditions matching the relevant region of phase space. Studying glutarate accumulation in hypoxic and/or acidic conditions would be particularly enticing, since pathways for its biosynthesis (e.g., as part of lysine metabolism) are conserved across many species.

There are several caveats and limitations associated with our analysis. The first one is that our redox chemical-space analysis is based solely on thermodynamics and does not account for kinetics. Thus, we assume that all molecular oxidation levels are accessible, effectively ignoring kinetic constraints. While it is clear that the emergence of early self-reproducing systems requires the underlying chemistry to bootstrap itself out of equilibrium, our quantitative and comprehensive analysis of the underlying energy landscape of redox reactions should be of value to prebiotic chemistry models. Furthermore, it may be interesting for future work to explore the relative role of thermodynamic and kinetic control in establishing early metabolism and ask whether kinetically controlled pathways that kick-started life were preceded by a prebiotic chemical era predominated by a strongly biased thermodynamically dominated mixture of compounds.

In our study, we explicitly did not want to bias our results toward biologically relevant molecules and redox reactions, and therefore the set of redox transformations considered here is not restricted to only those catalyzed by present-day enzymatic reaction mechanisms. For instance, the reduction of a hydroxycarbon (alcohol) functional group to a hydrocarbon occurs enzymatically through a C=C double-bond intermediate (for example, the reduction of malate to succinate occurs via fumarate). Therefore, a hydroxycarbon functional group that is incapable of undergoing elimination cannot undergo such a reduction using known enzymatic mechanisms. Since it is plausible that these transitions could have occurred through nonenzymatic mechanisms in prebiotic chemistry, we purposefully did not exclude these instances from our analysis.

In addition, our redox chemical space ignores further biochemical details: we do not include intramolecular redox transformations (where an electron transfer within a molecule changes the oxidation level of two different carbon atoms) or keto–enol tautomerizations; we do not account for non–linear-chain carbon compounds nor the different possible stereoisomers of a given molecular oxidation level (e.g., L-malate vs. D-malate), which may differ in energy; and we do not consider functional-group activation chemistry (e.g., the conversion of carboxylic acids to thiols), which has an important effect on thermodynamics. Finally, our partitioning of molecules into metabolites and nonbiological compounds relies on what is found in the KEGG database, which is only a proxy for the

Downloaded by guest on January 23, 2021

absolute set of compounds that partake in nature's redox biochemistry.

However, despite these caveats, we propose that our simplified redox chemical space is rich enough to serve as a baseline for a better understanding of the underlying thermodynamic and physicochemical principles of carbon redox biochemistry. In future work and following recent exciting developments in the field of heuristically aided quantum chemistry (42, 43, 68, 69), our chemical space model could be expanded to include the additional types of biochemical transformations mentioned above and begin to account for kinetic accessibility. It would be particularly interesting to include carboxylation and decarboxylation reactions (both reductive/oxidative and nonreductive/nonoxidative), which would effectively connect the different *n*-carbon redox chemical spaces to each other but would significantly increase the complexity of the analysis. Including additional types of reactions such as aldol/retro–aldol reactions and hydrations/dehydrations would fully map the chemical space model to experimentally tractable reaction networks (21, 70–72).

Finally, given the importance of redox chemistry at the early stages of life's history, it is possible to think of our landscape as a generalization of the space of metabolites found in current living systems (9, 10, 73). By taking into account this extended space, future models for the rise and evolution of biochemistry (42, 69, 74) could more specifically compare the evolutionary trajectory of life as we know it to alternative paths potentially involving transiently relevant molecules and reactions (31, 75).

Materials and Methods

To generate the reactions, we used the RDKit cheminformatics software to design simplified molecular-input line-entry system (SMILES) reaction templates (reaction strings), which, when applied to a compound, will reduce it according to the functional groups detected. In order to classify compounds in the full redox networks as biological or nonbiological, we looked for matches in the KEGG database of metabolic compounds using the RDKit toolbox. To compute the null distribution for the expected number functional-group patterns in *n*-carbon molecules, we derived the analytical solution for the probability of observing *m* instances of a given functional group in a sample size *N* (*SI Appendix*). For functional-group patterns of larger size, we computed the null distributions empirically by sampling compounds from redox chemical space. For all tests of statistical significance (i.e., differences in solubilities, *n*-gram counts, octanol–water partition coefficients, Gibbs energies of biological vs. nonbiological compounds), we performed Welch's unequal variance *t* test.

We used the cheminformatics software ChemAxon (Marvin 17.7.0, 2017; ChemAxon) to predict the pH-dependent solubility, logS(pH), and octanol–water distribution coefficients, logD(pH). To predict standard redox potentials with quantum chemistry, we computed the electronic structure and energy of the fully protonated species of each metabolite using the Orca quantum-chemistry software (*SI Appendix*). We then calibrated predictions against available experimental data using linear regression. Our group-contribution calculations to estimate redox potentials rely on the method as implemented by Noor et al. (14).

Data Availability. All study data are included in the article, *SI Appendix*, and Datasets 1–4.

ACKNOWLEDGMENTS. We thank Arren Bar-Even for invaluable discussions and Ron Milo, Manuel Razo-Mejia, Jennifer Wei, and Dmitrij Rappoport for feedback. We thank Harvard Research Computing for their support in using the Odyssey cluster. A.A.-G., A.J., and B.S.-L. thank Anders G. Frøseth for his generous support. J.E.G. was partially supported by grants from NASA, NSF, and the Human Frontiers Science Program. D.S. acknowledges support from the Directorates for Biological Sciences and Geosciences at NSF and NASA (Agreements 80NSSC17K0295, 80NSSC17K0296, and 1724150). A.A.-G. acknowledges the generous support from the Canada 150 Research Chairs Program.

1. A. L. Weber, Chemical constraints governing the origin of metabolism: The thermodynamic landscape of carbon group transformations under mild aqueous conditions. *Orig. Life Evol. Biosph.* **32**, 333–357 (2002).
2. A. Bar-Even, A. Flamholz, E. Noor, R. Milo, Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta* **1817**, 1646–1659 (2012).
3. A. Jinich *et al*., Quantum chemistry reveals thermodynamic principles of redox biochemistry. *PLOS Comput. Biol.* **14**, e1006471 (2018).
4. A. Bar-Even, A. Flamholz, E. Noor, R. Milo, Rethinking glycolysis: On the biochemical logic of metabolic pathways. *Nat. Chem. Biol.* **8**, 509–517 (2012).
5. A. Bar-Even, E. Noor, A. Flamholz, J. M. Buescher, R. Milo, Hydrophobicity and charge shape cellular metabolite concentrations. *PLOS Comput. Biol.* **7**, e1002166 (2011).
6. E. Noor, A. Flamholz, W. Liebermeister, A. Bar-Even, R. Milo, A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Lett.* **587**, 2772–2777 (2013).
7. D. A. Beard, H. Qian, Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS One* **2**, e144 (2007).
8. A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, R. Milo, Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10039–10044 (2013).
9. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
10. M. Kanehisa *et al*., KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
11. M. D. Jankowski, C. S. Henry, L. J. Broadbelt, V. Hatzimanikatis, Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* **95**, 1487–1499 (2008).
12. E. Noor *et al*., An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics* **28**, 2037–2044 (2012).
13. A. Flamholz, E. Noor, A. Bar-Even, R. Milo, eQuilibrator–The biochemical thermodynamics calculator. *Nucleic Acids Res.* **40**, D770–D775 (2012).
14. E. Noor, H. S. Haraldsdóttir, R. Milo, R. M. T. Fleming, Consistent estimation of Gibbs energy using component contributions. *PLOS Comput. Biol.* **9**, e1003098 (2013).
15. T. Schwabe, S. Grimme, Towards chemical accuracy for the thermodynamics of large molecules: New hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.* **8**, 4398–4401 (2006).
16. S. Grimme, Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **124**, 034108 (2006).
17. R. A. Alberty *et al*., Recommendations for terminology and databases for biochemical thermodynamics. *Biophys. Chem.* **155**, 89–103 (2011).
18. R. A. Alberty, *Thermodynamics of Biochemical Reactions* (John Wiley & Sons, 2005).
19. R. A. Alberty, Thermodynamics and kinetics of the glyoxylate cycle. *Biochemistry* **45**, 15838–15843 (2006).
20. W. Martin, M. J. Russell, On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 1887–1925 (2007).
21. K. B. Muchowska, S. J. Varma, J. Moran, Synthesis and breakdown of universal metabolic precursors promoted by iron. *Nature* **569**, 104–107 (2019).
22. M. Pourbaix, *Atlas of Electrochemical Equilibria in Aqueous Solutions* (National Association of Corrosion Engineers, 1966).
23. H. J. Morowitz, J. D. Kostelnik, J. Yang, G. D. Cody, The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7704–7708 (2000).
24. L. E. Orgel, The implausibility of metabolic cycles on the prebiotic Earth. *PLoS Biol.* **6**, e18 (2008).
25. W. Gilbert, Origin of life: The RNA world. *Nature* **319**, 618 (1986).
26. D. Sievers, G. von Kiedrowski, Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224 (1994).
27. L. E. Orgel, Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
28. N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, The origin of RNA and "my grandfather's axe". *Chem. Biol.* **20**, 466–474 (2013).
29. M. Ralser, The RNA world and the origin of metabolic enzymes. *Biochem. Soc. Trans.* **42**, 985–988 (2014).
30. D. Segré, D. Ben-Eli, D. W. Deamer, D. Lancet, The lipid world. *Orig. Life Evol. Biosph.* **31**, 119–145 (2001).
31. G. Wächtershäuser, Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 200–204 (1990).
32. W. F. Martin, Hydrogen, metals, bifurcating electrons, and proton gradients: The early evolution of biological energy conservation. *FEBS Lett.* **586**, 485–493 (2012).
33. A. L. Weber, Sugars as the optimal biosynthetic carbon substrate of aqueous life throughout the universe. *Orig. Life Evol. Biosph.* **30**, 33–43 (2000).
34. A. Walter, J. Gutknecht, Monocarboxylic acid permeation through lipid bilayer membranes. *J. Membr. Biol.* **77**, 255–264 (1984).
35. N. García-Quintáns, G. Repizo, M. Martín, C. Magni, P. López, Activation of the diacetyl/acetoin pathway in Lactococcus lactis subsp. lactis bv. diacetylactis CRL264 by acidic growth. *Appl. Environ. Microbiol.* **74**, 1988–1996 (2008).
36. S. R. Swindell *et al*., Genetic manipulation of the pathway for diacetyl metabolism in Lactococcus lactis. *Appl. Environ. Microbiol.* **62**, 2641–2643 (1996).
37. T. Miyata, Y. Izuhara, H. Sakai, K. Kurokawa, Carbonyl stress: Increased carbonyl modification of tissue and cellular proteins in uremia. *Perit. Dial. Int.* **19** (suppl. 2), S58–S61 (1999).
38. Y. J. Choi, S. Y. Lee, Microbial production of short-chain alkanes. *Nature* **502**, 571–574 (2013).
39. M. Harger *et al*., Expanding the product profile of a microbial alkane biosynthetic pathway. *ACS Synth. Biol.* **2**, 59–62 (2013).
40. N. Lane, Proton gradients at the origin of life. *BioEssays* **39**, (2017).

Downloaded by guest on January 23, 2021

41. N. Lane, J. F. Allen, W. Martin, How did LUCA make a living? Chemiosmosis in the origin of life. *BioEssays* **32**, 271–280 (2010).
42. D. Rappoport, C. J. Galvin, D. Y. Zubarev, A. Aspuru-Guzik, Complex chemical reaction networks from heuristics-aided quantum chemistry. *J. Chem. Theory Comput.* **10**, 897–907 (2014).
43. D. Rappoport, A. Aspuru-Guzik, Predicting feasible organic reaction pathways using heuristically aided quantum chemistry. *J. Chem. Theory Comp.* **15**, 4099–4112 (2019).
44. L. Tretter, A. Patocs, C. Chinopoulos, Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis. *Biochim. Biophys. Acta* **1857**, 1086–1101 (2016).
45. M. P. Murphy, L. A. J. O'Neill, Krebs cycle reimagined: The emerging roles of succinate and itaconate as signal transducers. *Cell* **174**, 780–784 (2018).
46. E. Maklashina, D. A. Berthold, G. Cecchini, Anaerobic expression of Escherichia coli succinate dehydrogenase: Functional replacement of fumarate reductase in the respiratory chain during anaerobic growth. *J. Bacteriol.* **180**, 5989–5996 (1998).
47. H. Eoh, K. Y. Rhee, Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6554–6559 (2013).
48. C. Serena *et al*., Elevated circulating levels of succinate in human obesity are linked to specific gut microbiota. *ISME J.* **12**, 1642–1657 (2018).
49. F. De Vadder *et al*., Microbiota-produced succinate improves glucose homeostasis via intestinal gluconeogenesis. *Cell Metab.* **24**, 151–157 (2016).
50. P. Kovatcheva-Datchary *et al*., Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of prevotella. *Cell Metab.* **22**, 971–982 (2015).
51. G. Jakobsdottir, J. Xu, G. Molin, S. Ahrné, M. Nyman, High-fat diet reduces the formation of butyrate, but increases succinate, inflammation, liver fat and cholesterol in rats, while dietary fibre counteracts these effects. *PLoS One* **8**, e80476 (2013).
52. J. Li *et al*., Succinate accumulation impairs cardiac pyruvate dehydrogenase activity through GRP91-dependent and independent signaling pathways: Therapeutic effects of ginsenoside Rb1. *Biochim. Biophys. Acta Mol. Basis Dis.* **1863**, 2835–2847 (2017).
53. E. T. Chouchani *et al*., Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature* **515**, 431–435 (2014).
54. E. T. Chouchani *et al*., A unifying mechanism for mitochondrial superoxide production during ischemia-reperfusion injury. *Cell Metab.* **23**, 254–263 (2016).
55. P. W. Hochachka, K. B. Storey, Metabolic consequences of diving in animals and man. *Science* **187**, 613–621 (1975).
56. H. Muratsubaki, Regulation of reductive production of succinate under anaerobic conditions in baker's yeast. *J. Biochem.* **102**, 705–714 (1987).
57. M. Gallmetzer, J. Meraner, W. Burgstaller, Succinate synthesis and excretion by Penicillium simplicissimum under aerobic and anaerobic conditions. *FEMS Microbiol. Lett.* **210**, 221–225 (2002).
58. J. J. Van Hellemond, A. G. Tielens, Expression and functional properties of fumarate reductase. *Biochem. J.* **304**, 321–331 (1994).
59. M. H. Roos, A. G. Tielens, Differential expression of two succinate dehydrogenase subunit-B genes and a transition in energy metabolism during the development of the parasitic nematode Haemonchus contortus. *Mol. Biochem. Parasitol.* **66**, 273–281 (1994).
60. S. Besteiro *et al*., Succinate secreted by Trypanosoma brucei is produced by a novel and unique glycosomal enzyme, NADH-dependent fumarate reductase. *J. Biol. Chem.* **277**, 38001–38012 (2002).
61. A. de Zwaan, R. H. M. Eertman, Anoxic or aerial survival of bivalves and other euryoxic invertebrates as a useful response to environmental stress—A comprehensive review. *Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol.* **113**, 299–312 (1996).
62. D. Amador-Noguez *et al*., Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in Clostridium acetobutylicum. *J. Bacteriol.* **192**, 4452–4461 (2010).
63. S. Watanabe *et al*., Fumarate reductase activity maintains an energized membrane in anaerobic Mycobacterium tuberculosis. *PLoS Pathog.* **7**, e1002287 (2011).
64. X. Chen, A. P. Alonso, D. K. Allen, J. L. Reed, Y. Shachar-Hill, Synergy between (13)C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in E. coli. *Metab. Eng.* **13**, 38–48 (2011).
65. T. Hartman *et al*., Succinate dehydrogenase is the regulator of respiration in Mycobacterium tuberculosis. *PLoS Pathog.* **10**, e1004510 (2014).
66. J. R. Jass, Diet, butyric acid and differentiation of gastrointestinal tract tumours. *Med. Hypotheses* **18**, 113–118 (1985).
67. D. R. Donohoe *et al*., The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol. Cell* **48**, 612–626 (2012).
68. D. Rappoport, Reaction networks and the metric structure of chemical space(s). *J. Phys. Chem. A* **123**, 2610–2620 (2019).
69. D. Y. Zubarev, D. Rappoport, A. Aspuru-Guzik, Uncertainty of prebiotic scenarios: The case of the non-enzymatic reverse tricarboxylic acid cycle. *Sci. Rep.* **5**, 8009 (2015).
70. S. J. Varma, K. B. Muchowska, P. Chatelain, J. Moran, Native iron reduces CO_2 to intermediates and end-products of the acetyl-CoA pathway. *Nat. Ecol. Evol.* **2**, 1019–1024 (2018).
71. K. B. Muchowska *et al*., Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* **1**, 1716–1721 (2017).
72. M. A. Keller, A. V. Turchyn, M. Ralser, Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol. Syst. Biol.* **10**, 725 (2014).
73. J. E. Goldford, H. Hartman, T. F. Smith, D. Segrè, Remnants of an ancient metabolism without phosphate. *Cell* **168**, 1126–1134.e9 (2017).
74. J. E. Goldford, D. Segrè, Modern views of ancient metabolic networks. *Curr. Opin. Syst. Biol.* **8**, 117–124 (2018).
75. E. Smith, H. J. Morowitz, Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13168–13173 (2004).

Downloaded by guest on January 23, 2021