



Open camera or QR reader and scan code to access this article and other resources online.

Quantifying Global Origin-Diagnostic Features and Patterns in Biotic and Abiotic Acyclic Lipids for Life Detection

Denise K. Buckner,^{1,2,3} Morgan J. Anderson,^{3,4} Sydney Wisnosky,^{4,5} Walter Alvarado,^{3,6} Michel Nuevo,³ Amy J. Williams,¹ Antonio J. Ricco,^{3,7} Anamika,⁸ Sara Debic,⁹ Lauren Friend,⁴ Trinh Hoac,⁴ Linda Jahnke,³ Leslie Radosevich,⁴ Ross Williams,¹⁰ and Mary Beth Wilhelm³

Abstract

Lipids are a geologically robust class of organics ubiquitous to life as we know it. Lipid-like soluble organics are synthesized abiotically and have been identified in carbonaceous meteorites and on Mars. Ascertaining the origin of lipids on Mars would be a profound astrobiological achievement. We enumerate origin-diagnostic features and patterns in two acyclic lipid classes, fatty acids (*i.e.*, carboxylic acids) and acyclic hydrocarbons, by collecting and analyzing molecular data reported in over 1500 samples from previously published studies of terrestrial and meteoritic organics. We identify 27 combined (15 for fatty acids, 12 for acyclic hydrocarbons) molecular patterns and structural features that can aid in distinguishing biotic from abiotic synthesis. Principal component analysis (PCA) demonstrates that multivariate analyses of molecular features (16 for fatty acids, 14 for acyclic hydrocarbons) can potentially indicate sample origin. Terrestrial lipids are dominated by longer straight-chain molecules (C₄-C₃₄ fatty acids, C₁₄-C₄₆ acyclic hydrocarbons), with predominance for specific branched and unsaturated isomers. Lipid-like meteoritic soluble organics are shorter, with random configurations. Organic solvent-extraction techniques are most commonly reported, motivating the design of our novel instrument, the Extractor for Chemical Analysis of Lipid Biomarkers in Regolith (ExCALiBR), which extracts lipids while preserving origin-diagnostic features that can indicate biogenicity. Key Words: Fatty acid—Lipid—Hydrocarbon—Mars analog—Biomarker—Carbonaceous chondrite. Astrobiology 24, 1–35.

1. Introduction

IN THE SEARCH for life beyond Earth, molecular biosignatures can provide compelling evidence for ancient or extant life. Molecular biosignatures include biogenic organic compounds produced by life and subsequently preserved in the rock record, serving as “molecular fossils” of the organisms from which they originated. However, since organic molecules are synthesized through both biotic and abiotic

processes, identifying evidence for life, or its absence, is contingent upon the ability not only to detect organics but also to diagnose their origin, using information contained within individual molecular structures and overall distributions within a sample. These origin-diagnostic features and patterns can indicate whether the organics in question were likely to have been synthesized biotically versus abiotically—a potential key to determining whether life arose on other bodies in the Solar System.

¹Department of Geological Sciences, University of Florida, Gainesville, Florida, USA.

²Blue Marble Space Institute of Science, Seattle, Washington, USA.

³Space Science & Astrobiology Division, NASA Ames Research Center, Moffett Field, California, USA.

⁴Axient Corporation, Huntsville, Alabama, USA.

⁵Department of Biology, University of Miami, Coral Gables, Florida, USA.

⁶Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois, USA.

⁷Electrical Engineering—Integrated Circuits Laboratory, Stanford University, Stanford, California, USA.

⁸Department of Space Studies, University of North Dakota, Grand Forks, North Dakota, USA.

⁹Department of Biology, Johns Hopkins University, Baltimore, Maryland, USA.

¹⁰Civil & Environmental Engineering & Earth Sciences, Notre Dame University, Notre Dame, Indiana, USA.

This molecular biosignature-based approach to life detection was first proposed in 1965 by James Lovelock, who posited that physical parameters within sets of organic molecules can indicate a biotic origin if they differ from those observed in abiotic scenarios or expected within thermodynamic constraints (Lovelock, 1965). Specifically, evidence for order and/or evidence for non-equilibrium chemistry can signal life. This framework, which formed the basis for organics-focused life-detection experiments in subsequent decades, is frequently cited as a key method for astrobiologists (Dorn *et al.*, 2011; Georgiou and Deamer, 2014) leveraged by numerous current missions to Mars (*e.g.*, the Curiosity and Perseverance rovers), and recommended by the Decadal Survey for astrobiology missions prioritized in the coming decade (*e.g.*, Enceladus flyby and Orbilander, Mars Life Explorer, ExoMars, Europa lander) (National Academies of Sciences, Engineering, and Medicine, 2022). Lovelock's approach is supported by tens of thousands of biomarker analyses on Earth, where organic biogeochemists utilize known molecular structures to track evolution, ecosystem dynamics, environmental change, and diagenesis recorded in organics preserved in geologic samples (Brocks *et al.*, 2005; Brocks and Schaeffer, 2008; Lee and Brocks, 2011; Vinnichenko *et al.*, 2020). The results of these analyses demonstrate that life has evolved mechanisms to preferentially synthesize molecules that support cell structures, drive metabolic processes, and pass genetic information.

The same properties and approaches can be key tools for astrobiologists to apply to samples from other Solar System bodies (Georgiou and Deamer, 2014; Mißbach *et al.*, 2018). With pattern-based analyses of geologically preserved organics, identifying whole cells or complex biomolecules (*e.g.*, DNA, proteins, carbohydrates, membrane lipids) unique to Terran organisms is not a requirement for detecting extraterrestrial life. Instead, distributions of the monomers (*e.g.*, nucleobases, amino acids, sugar derivatives, carboxylic acids, hydrocarbons) that make up those larger structures can provide origin-diagnostic information indicating whether life exists or existed on another planet (McCullom *et al.*, 1999; Dorn *et al.*, 2011; Mißbach *et al.*, 2018). Critically, some of these building blocks of life, namely lipids, are both geologically recalcitrant (*e.g.*, Peters and Moldowan, 1993; Brocks and Schaeffer, 2008; Eigenbrode, 2008; Lee and Brocks, 2011) and amenable to detection with analytical methods utilized for characterizing extraterrestrial organics both *in situ* and with returned sample analysis (*e.g.*, mass spectrometry-based) (*e.g.*, Mahaffy *et al.*, 2012; Martins, 2020). Few molecule classes fulfill all these criteria, but one class that does is acyclic lipids.

Broadly defined, lipids are a diverse class of organic molecules that are soluble in organic solvents (IUPAC, n.d.; Ratnayake and Galli, 2009; Summons *et al.*, 2022). On Earth, a large diversity of lipids are biosynthesized and fulfill life-enabling functions by building the cell membranes that are universal to life as we know it (Summons *et al.*, 2008; Georgiou and Deamer, 2014). The lipid central cores of bacterial and eukaryotic cell membranes are made up of ester-linked monocarboxylic acids, *i.e.*, fatty acids, while archaeal membranes comprise ether-linked acyclic isoprenoids, *i.e.*, hydrocarbon chains composed of 2-methyl-1,3-butadiene (isoprene) units (Fig. 1). Although "lipid" traditionally refers to biosynthesized molecules, we leverage

the term to describe soluble hydrocarbons of both biotic and abiotic origin, since the same classes of molecules (*e.g.*, carboxylic acids and hydrocarbons described in this paper) are synthesized both biogenically and abiogenically and are found in carbonaceous meteorites. Similarly, while "fatty acid" traditionally refers to biosynthesized longer-chained (*e.g.*, C₁₄ or longer) monocarboxylic acids, we refer to monocarboxylic acids of any chain length (*e.g.*, C₁ and longer) by this term for simplicity's sake in this paper. However, we note that there seems to be no commonly used or universal term to describe a group of molecules that are capable of forming cell membranes. "Soluble organics" is an operational definition but lacks the specificity of the term "lipid," which includes membrane-forming molecules (*e.g.*, carboxylic acids, aliphatic hydrocarbons, polycyclic triterpenoids, alcohols, etc.), but excludes other organics (*e.g.*, amino acids, nucleic acids, and sugar derivatives) that may be partially soluble in organic solvents but fulfill different biological functions (*e.g.*, proteins, DNA, and carbohydrates, respectively).

In cells, hydrophobic lipids are linked to various polar molecules forming a hydrophilic surface, and the resulting amphiphilic structures segregate functional biological material from the external environment (Sapers *et al.*, 2019). Simple acyclic lipids (*e.g.*, monocarboxylic acids and hydrocarbons) also form abiotically and are capable of self-assembling into functional types of primitive "membranes" that may have provided some of the materials that helped facilitate the transition from prebiotic chemistry to biochemistry on Earth (Deamer, 1985; Deamer and Pashley, 1989; Dworkin *et al.*, 2001; Segré *et al.*, 2001; Apel *et al.*, 2002; Deamer *et al.*, 2002), particularly during early stages of the Solar System's formation, when meteoritic and cometary impacts were significantly more frequent (Chyba and Sagan, 1992; Pizzarello and Shock, 2017). Shorter-chain (\leq C₁₂ fatty acids, \leq C₃₀ acyclic hydrocarbons), lipid-like hydrocarbons are the most abundant soluble organics found in carbonaceous meteorites and other infalling extraterrestrial materials. Short-chain fatty acids, *n*-alkanes and alkenes, and branched alkanes make up well over half of the soluble fraction of organic carbon in carbonaceous chondrites, where soluble molecules comprise up to 25% of the total organic carbon (TOC) contained in such meteorites (Pizzarello, 2006; Sephton, 2006; Remusat, 2014). The remaining organic carbon is bound in insoluble organic macromolecules (IOM), consisting of smaller molecules and fragments bound in a complex network, structurally similar to terrestrial kerogen (*e.g.*, Remusat *et al.*, 2007; Pizzarello *et al.*, 2013; Alexander *et al.*, 2017).

In addition to biological importance, lipid-derived hydrocarbons are geologically robust and can persist in the terrestrial rock record for billions of years, which is orders of magnitude longer than other molecular biomarkers, including information-coding biopolymers (*e.g.*, DNA, RNA), amino acid enantiomeric excesses, and macromolecules (*e.g.*, proteins, sugars, pigments) (Peters and Moldowan, 1993; Brocks and Schaeffer, 2008; Lee and Brocks, 2011). Extremely arid conditions, like those on Mars over the last ~3.7 billion years (Gya), are expected to enhance preservation of diagnostic structural features of biogenic origin (Wilhelm *et al.*, 2017), especially in the absence of metamorphism and significant geothermal heating associated with burial, which is largely responsible for post-deposition alteration of organics on Earth (Peters and

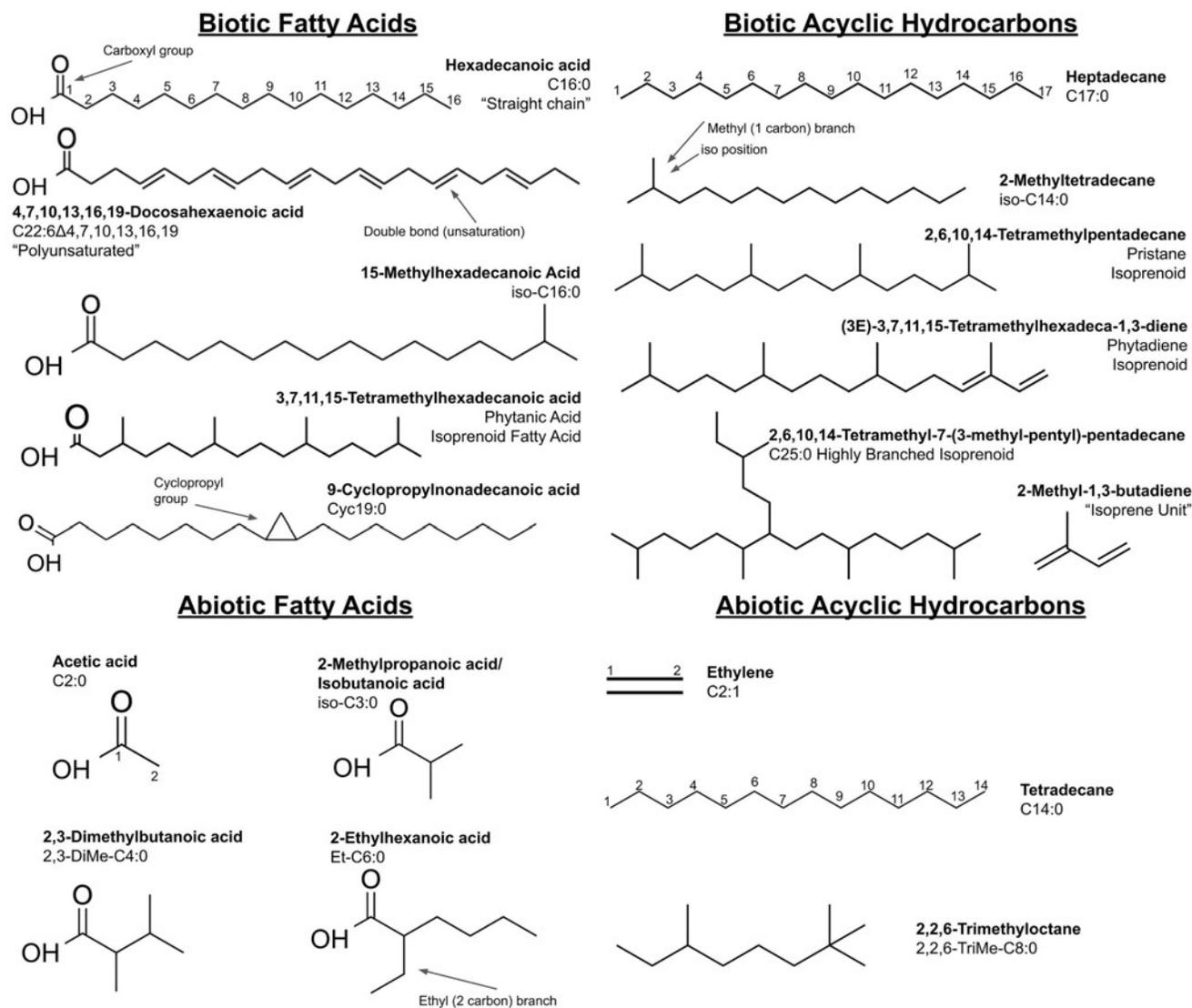


FIG. 1. Visual representation of common features and structures, along with common conformations reported in the literature, for biotic fatty acids, abiotic fatty acids, biotic acyclic hydrocarbons, and abiotic acyclic hydrocarbons. A straight chain fatty acid contains no branches or double bonds; a polyunsaturated fatty acid contains more than one double bond; a branched fatty acid contains one or more side chains made of linked carbon and hydrogen atoms that extend off the main chain of the molecule; an isoprenoid fatty acid is made up of linked isoprene (*i.e.*, 2-methyl-1,3-butadiene) units with a carboxyl group at one end of the molecule; a cyclopropyl fatty acid contains a cyclopropyl group within the main chain of the molecule. Molecular feature descriptors and nomenclature for acyclic hydrocarbons are similar to fatty acid schemes, although these molecules lack a carboxyl head group. An *n*-alkane is straight-chained with no branches or unsaturations, an alkane is straight-chained with no unsaturations (but may contain one or more branches), and an alkene is straight-chained with one or more unsaturations. Acyclic hydrocarbons with a single methyl branch at the 2-carbon position within the main chain are colloquially termed isoalkanes (*e.g.*, iso-C_{14:0}), since these molecules are thought to originate from -iso branched fatty acids that have undergone decarboxylation. Isoprenoids are made up of repeating isoprene (*i.e.*, 2-methyl-1,3-butadiene) units linked in various conformations. Highly branched isoprenoids display complex structures, which are often unresolved or resolved with low confidence via GC-MS analyses.

Moldowan, 1993). The geologic longevity of lipid hydrocarbon cores is on the same order as the age of the >3 Gy sediments laid down during the most habitable surface conditions on Mars (Summons and Walter, 1990; Brocks *et al.*, 2003; Brocks and Schaeffer, 2008; Lee and Brocks, 2011; Summons *et al.*, 2022), making these structures some of the most probable preserved and accessible indicators of past life (Dorn *et al.*, 2011; Georgiou and Deamer, 2014).

Biotic lipid structures and patterns linked to synthesis are well known within terrestrial context, and lipid-like hydrocarbons in meteorites are also well characterized. For example, terrestrial biology synthesizes fatty acids via the addition of 2-carbon groups (*e.g.*, Brindley *et al.*, 1969; McCarthy and Hardie 1984). In contrast, abiotic mechanisms add single carbons at a time through Fischer Tropsch-type (FTT) synthesis (McCullom *et al.*, 1999; Rushdi and

Simoneit, 2001) or molecule-radical-ion reactions during energetic processing of ices (Bernstein *et al.*, 1995, 2001; Dworkin *et al.*, 2001; Sandford *et al.*, 2020). A predominance for “even over odd” chain lengths is an oft-cited lipid distribution of astrobiological interest (Lovelock, 1965; Dorn *et al.*, 2011; Mißbach *et al.*, 2018). However, acyclic lipids possess additional molecular features, including the number of carbon atoms, the presence or absence of double bonds, cyclization, and branching structures and configurations, which also provide insights into their origin. Critically, each of these individual structural elements can be found in both biotic and abiotic molecules, but patterns in presence, frequency, and positions differ (Georgiou and Deamer, 2014; Summons *et al.*, 2022). Consequently, a single lipid sample can potentially hold multiple lines of evidence for biogenicity or abiogenicity, in the case that multiple origin-diagnostic patterns in distributions of these features are observed within the same sample.

To build upon Lovelock’s pattern-based approach to life detection, we have conducted an analytical review of data reported in the literature to better constrain distributions of naturally synthesized biotic and abiotic acyclic lipids, expand the breadth of known molecular patterns that may indicate the presence of life or its absence, and build a quantitative framework to support lipid-based analyses on astrobiology missions. The goal of this study is not to identify new biomarkers or quantify every molecule present within natural lipid samples, but rather to characterize broad trends, predominances, and endmembers to constrain the types of features, molecular ranges, and patterns in distributions most likely to indicate biogenicity within samples on Earth. To do this, we collected published data on naturally synthesized lipids found in natural samples of varying age, input sources, and diagenetic histories from terrestrial (*i.e.*, biotic synthesis) and meteoritic (*i.e.*, abiotic synthesis) sources, collected data from 220 of these studies (references listed in the Supplementary Information), and performed numerical analyses on the datasets we created to identify potential biomarker signatures.

We enumerate multiple origin-diagnostic patterns in the distributions of certain molecular features (*i.e.*, carbon chain length, double bonds, branching) that can constrain their origin as biotic or abiotic for two major types of lipids: fatty acids and acyclic hydrocarbons. Our results expand the utility of Lovelock’s physical approach to biomarker analyses, revealing a greater number of origin-diagnostic molecular features and patterns in distributions of fatty acids and acyclic hydrocarbons that are broadly observed on Earth and in meteorites and can be exploited in the search for life beyond.

2. Methods

2.1. Data collection

Molecular data on the identities and distributions of fatty acids and/or acyclic hydrocarbons as reported in each of the 220 studies were manually documented, used to populate four separate datasets (biogenic fatty acids, abiogenic fatty acids, biogenic hydrocarbons, abiogenic hydrocarbons), and then analyzed with supervised numerical and unsupervised statistical methods to elucidate patterns indicating biogenicity or abiogenicity. Each reviewed study included data

from one or more unique lipid samples or sets of lipids extracted from an individual geological specimen or water sample. Although some papers reported abundances of both fatty acids and acyclic hydrocarbons, each class was analyzed independently. For each lipid sample, data were cataloged pertaining to three categories of molecular features: chain length, unsaturations, and branching. Experimental techniques used to extract and analyze lipids were also collated (Fig. S1).

For terrestrial samples, we selected an astrobiologically-relevant subset of studies from the literature on organic geochemical analyses of fatty acids and acyclic hydrocarbons, with a focus on Mars analog environments primarily dominated by prokaryotic (*i.e.*, single-celled) input. Several culture studies (22 studies with 208 individual samples) were also included, as they are free of input from higher plants and other multicellular organisms. Culture and geologic samples were first analyzed independently, then together, and we identified no significant differences in distributions between the two types of samples, except that older samples generally contain fewer unsaturated species, a known consequence of diagenesis (Peters and Moldowan, 1993; Canuel and Martens, 1996; Colombo *et al.*, 1997; Eigenbrode, 2008). The selected biotic studies are a representative cross section of the literature. For each terrestrial study, a subset with a maximum of 57 samples from any single study was included so as not to bias the dataset. Subsets were either selected at random if data presented in a paper was numerically similar and from similar sample substrate, based on endmembers to show spread of the data presented if there was greater diversity substrate/soil horizon/sample type (*e.g.*, top, middle, and bottom of geologic core sections), or chosen to encompass greatest diversity in environments (*e.g.*, individual samples from representative samples within different sites).

Meteorite sample data was collected from studies of carbonaceous chondrite that quantitatively report on molecular abundances and distributions of indigenous fatty acids (*i.e.*, monocarboxylic acids) and/or acyclic hydrocarbons and molecular features (*e.g.*, chain length, unsaturations, branching), using only studies where the researchers could exclude terrestrial contamination based on analytical measurements, such as isotopic analyses, blanks and witness plates, extraction of exterior portions and protection of interior portions, and so on (Table 1). There exist fewer studies on meteoritic lipids compared to terrestrial, and to the best of our knowledge our dataset encompasses all peer-reviewed publications on carbonaceous chondrite-sourced carboxylic acids and acyclic hydrocarbons to date that match these criteria. All samples reported in each study were included in our analysis, except for isolated cases where the study reported the presence of contamination in individual samples (*e.g.*, on exterior exposed portions of the meteorite). For every lipid sample, three categories of relevant data were collected and recorded: (1) sample parameters, (2) sample processing techniques, and (3) molecular information (Table 1).

2.1.1. Lipid nomenclature. Lipid nomenclature follows standard schemes designated by IUPAC (IUPAC, n.d.). For fatty acids, molecules are identified by the number of carbons in the main chain of the molecule, along with the

TABLE 1. NUMBER OF PAPERS AND ASSOCIATED INDIVIDUAL SAMPLES FROM PEER-REVIEWED PUBLICATIONS USED IN THIS STUDY

		<i>Number of references</i>	<i>Number of samples</i>
Fatty acids/Monocarboxylic acids	Terrestrial (biotic)	114	893
	Meteoritic (abiotic)	22	58
Acyclic hydrocarbons	Terrestrial (biotic)	100	592
	Meteoritic (abiotic)	14	31

number of double bonds between carbon atoms within that chain, branch position and length, and cyclopropyl groups (Fig. 1). Carbons are counted from the carboxyl group. A straight-chain fatty acid contains no branches or double bonds; a polyunsaturated fatty acid contains more than one double bond; a branched fatty acid contains one or more side chains made of linked carbon and hydrogen atoms that extend off the main chain of the molecule; an isoprenoid fatty acid is made up of linked isoprene units with a carboxyl group at one end of the molecule; a cyclopropyl fatty acid contains a cyclopropyl group within the main chain of the molecule.

Fatty acid nomenclature is as follows. $C_{x,y}$ denotes a fatty acid by chain length and bonding, where x =the number of carbons in the main chain of the molecule and y =the number of double bonds within that chain; the positions of those double bonds may be indicated by either a delta (Δ) or omega (ω) counting scheme (*e.g.*, $C_{22:6\Delta 4,7,10,13,16,19}$), where delta indicates the positions of unsaturations as counted from the carboxyl end, while omega indicates the positions of unsaturations as counted from the opposite terminal end (IUPAC, n.d.). Branch length and position is also noted, where Me indicates methyl branching (*i.e.*, one $-CH_3$ group extending off the main chain), DiMe indicates dimethyl branching (*i.e.*, two $-CH_3$ groups that extend off the main chain), and Et indicates ethyl branching (*i.e.*, a $-C_2H_5$ group extending off the main chain). An *iso* prefix (*e.g.*, *iso-C*_{16:0}) denotes a single methyl branch positioned at the penultimate carbon opposite the carboxyl end of the molecule, and an *anteiso* prefix (*e.g.*, *anteiso-C*_{18:1}) denotes a single methyl branch positioned at the antepenultimate carbon opposite the carboxyl.

Molecular feature descriptors and nomenclature for acyclic hydrocarbons are similar to fatty acid schemes, where molecules are named and grouped based on the number of carbons in the main chain, unsaturations between carbon bonding, and branching (*e.g.*, a $C_{17:0}$ acyclic hydrocarbon contains 17 carbons in the main chain and no unsaturations) (Fig. 1). An *n*-alkane is straight-chained with no branches or unsaturations, an alkane is straight-chained with no unsaturations (but may contain one or more branches), and an alkene is straight-chained with one or more unsaturations. Acyclic hydrocarbons with a single methyl branch at the 2-carbon position within the main chain are colloquially termed isoalkanes (*e.g.*, *iso-C*_{14:0}), since these molecules are thought to originate from *-iso* branched fatty acids that have undergone decarboxylation. Isoprenoids are an important subclass of acyclic hydrocarbons and are made up of repeating isoprene (*i.e.*,

2-methyl-1,3-butadiene) units linked in various conformations. Highly branched isoprenoids (HBIs) display complex structures, which are often unresolved or resolved with low confidence via gas chromatography–mass spectrometry (GC-MS) analyses.

2.2. Sample parameters

Biotic samples span a range of terrestrial settings across the globe (Fig. S2) and are binned into broad categories based on environment, general lithology, and major input source (Figs. S3a, S4a). Samples of various ages were included, with a focus on older specimens. Fatty acid sample ages span the Devonian period (~ 420 Mya) to present, while acyclic hydrocarbons span the Precambrian supereon (~ 4 Gya) to present (Figs. S3b, S4b), recognizing that some reported ages may refer to the geologic unit and that microbial overprint may contribute younger organics. Abiotic lipid sample data was collected from studies on carbonaceous chondrites that report the detection of fatty acids and/or acyclic hydrocarbons (Fig. S5). Since meteorites represent materials formed during the early stages of Solar System formation, around 4.6 Gya (Pizzarello, 2006; Sephton, 2006), ages for these samples were not included, although group, petrographic type, and specific meteorite were cataloged.

2.2.1. Terrestrial sample parameters.

- Geographical location (latitude/longitude);
- Age (if reported);
- Broad category pertaining to environment. See Fig. S4 for detailed category descriptions.

2.2.2. Meteorite sample parameters.

- Group and petrographic type;
- Individual specimen.

2.3. Sample processing techniques

Experimental protocols used to process samples were recorded, including the extraction technique used to isolate lipids and the analytical method used to characterize molecular structures. These methods are characterized according to:

- Lipid extraction technique (*e.g.*, modified Bligh and Dyer, Soxhlet, pyrolysis, solid-phase micro-extraction [SPME], etc.) (Appendix 2);

- Solvent cocktail (*i.e.*, composition and ratio) when applicable;
- Analytical method (*e.g.*, GC-MS, GC-FID, etc.) used to identify and quantify lipids within a given sample (Appendix 3).

2.4. Molecular information

For each sample, data were collected pertaining to (1) carbon chain length, (2) unsaturations, and (3) branching, then used to populate spreadsheets for analyses. Information was recorded on the presence, relative abundance, frequency, and position of each feature within the lipid sample, as explicitly reported by authors. Not all features or parameters are present or reported in all samples; when features are not present or parameters are not explicitly reported, they were excluded from the spreadsheets we generated (marked with “0” or “n.a.” as appropriate) and calculations we performed. Since our review focuses on monocarboxylic acids and acyclic hydrocarbons only, any other lipid classes (*e.g.*, hydroxy acids, dicarboxylic acids, polycyclic hydrocarbons) reported in the studies reviewed were excluded from our analysis. Conformations only observed in biogenic samples (cyclopropyl groups and isoprenoids for fatty acids, isoprenoids for acyclic hydrocarbons) were also identified. From a physical perspective, isoprenoids are branched molecules, so these molecules and their structural features (*e.g.*, chain length, number and position of branches, etc.) were additionally cataloged under the “branching” category.

2.4.1. Chain length parameters.

- Minimum and maximum chain lengths (*i.e.*, the shortest and longest molecules identified within a set of lipids found within a natural or environmental sample);
- Statistical distribution of the chain lengths (if explicitly reported by the researchers, *e.g.*, Poisson, unimodal, bimodal);
- Even or odd predominance of carbon atom number;
- Identity of the most abundant and second most abundant molecules.

2.4.2. Unsaturation parameters provide details regarding multiple carbon-carbon bonds.

- Presence or absence of unsaturated molecules;
- Number of unique unsaturated molecules (each isomer counted as a unique molecule);
- Maximum number of double bonds found in a single molecule;
- Identity of the most abundant unsaturated molecule present within the sample (*e.g.*, C_{18:1}, C_{16:2}, C_{20:6}, etc.). Double bond position was not recorded, as determining this information requires additional chemical processing or derivatization steps during sample preparation and was not conducted or reported for many of the studies reviewed.

2.4.3. Branching parameters summarize the extent and nature of structural branches.

- Number of unique branched molecules (each isomer counted as a unique molecule);

- Minimum and maximum chain lengths (*i.e.*, number of carbon atoms in the main chain) of molecules that contain branches;
- Minimum and maximum number of branches present in any single molecule;
- Length (*e.g.*, number of carbon atoms) of the shortest and longest branches present in any single molecule;
- Range of positions within the main chain where branching points occur (*i.e.*, the first and last carbon atoms with a branching point for any of the molecules in the sample);
- Identity of the most abundant branched molecule present within a sample (main chain length and branch length/position, *e.g.*, *iso*, *anteiso*, 9-methyl, isoprenoid, etc.).

2.4.4. Acyclic hydrocarbon parameters specify.

- Presence of homologous series (*e.g.*, *iso*, *anteiso*, monomethyl alkanes);
- Presence of unresolved complex mixtures (UCMs), *i.e.*, mixtures of hydrocarbons that cannot not be identified on an individual basis because structural similarity and numerous individual isomers precludes separation.

2.4.5. Biological acyclic hydrocarbon (isoprenoid) parameters tabulate.

- Total number of unique isoprenoids;
- Minimum and maximum number of carbon atoms (*i.e.*, fewest and greatest number of carbon atoms present in any isoprenoid);
- Identity of the single most abundant isoprenoid.

2.5. Principal component analysis

Principal component analysis (PCA) is a statistical technique used to identify patterns in high-dimensional data and reduce its complexity by transforming it into a lower-dimensional space, by selecting certain features of the data and combining them in a specific way to create new “principal components.” These new principal components are created using the eigenvectors and eigenvalues of the covariance matrix of the original data. Eigenvectors show the directions of maximum variance in the data, and the eigenvalues show the amount of variance in those directions. PCA is often used as a preprocessing step for other machine learning algorithms, as it can help reduce the data’s dimensionality and complexity while preserving as much of the underlying structure as possible. The versatility and interpretability of PCA have been well documented and shown to be effective in various fields (Jolliffe and Cadima, 2016). However, PCA is sensitive to outliers, which can have a significant impact on the results of the analysis. Outliers can affect the direction of the principal components and the amount of variance captured by each component, leading to a disproportionate influence on the results of the PCA. To ensure reliable and meaningful results, it is generally good practice to identify and handle outliers before running PCA. In our case, we omitted samples with missing features when concatenating the biotic and abiotic fatty acid datasets. Our resulting dataset comprises 381 biotic and 31 abiotic samples with 16 features (Table S1), which is a subset of the features that were analyzed individually.

To validate that the identified molecules and their structural qualities indicate biogenicity or abiogenicity, we applied the k -means clustering algorithm (Lloyd, 1982) over the PCA-reduced data. The k -means clustering algorithm is an unsupervised learning approach that groups unlabeled data into sets of clusters, k , based on their degree of similarity. Clusters are iteratively formed by minimizing the sum of the distance of points from their respective cluster centroid. From visual inspection, we determined a k value of two and applied k -means over the subspace of the top two PCs (which account for the highest variance in the dataset).

Given that our hydrocarbon dataset contains a mixture of quantitative and qualitative values, we additionally required a metric with the ability to measure the similarities between datasets consisting of numerical, categorical, and text data. To this end, we relied on the Gower distance metric, which measures the similarities of two records that have numeric and non-numeric entries (Gower, 1971). The resulting dissimilarity matrix produced by the Gower can then be dimensionally reduced with PCA and/or clustered with k -means.

The Gower similarity matrix D_{Gower} is calculated as the average score taken over all possible comparisons of features,

$$D_{\text{Gower}}(x_1, x_2) = 1 - \left(\frac{1}{p} \sum_{j=1}^p s_j(x_1, x_2) \right)$$

where p are the features for two observations $x_i = (x_{i1}, \dots, x_{ip})$ and $x_j = (x_{j1}, \dots, x_{jp})$, and the score $s_j(x_1, x_2)$ is the partial similarity function defined as $s_{ij} \in [0, 1]$ for each descriptor. If a feature k is comparable for x_i and x_j , then a score of positive fraction or one is assigned ($s_{ijk} = 1$), and a score of zero if they are dissimilar ($s_{ijk} = 0$). Whether features are comparable depends on the type of feature k and is represented by the quantity δ_{ijk} , where δ_{ijk} is equal to 1 when a feature k can be compared for x_i and x_j , and zero otherwise.

The scores s_{ijk} can be determined for three feature types: qualitative, quantitative, and dichotomous characters. For quantitative values, s_{ijk} is determined as

$$s_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k}$$

where R_k is the range of feature k and can be the total range of the population or sample. For qualitative values, $s_{ijk} = 1$ if feature k of the two observations x_i and x_j agree and $s_{ijk} = 0$ if they differ. A third feature is dichotomous characters which refer to missing k features in the dataset; since these observations are removed during preprocessing, their calculations are omitted. Our resulting dataset comprises 14 features (Table S2), which is a subset of the features that were analyzed individually.

3. Results

Analysis of data reported for 1574 unique samples in total from 220 peer-reviewed studies reveals 15 potential origin-diagnostic patterns in distributions of molecular features for fatty acids and 12 for acyclic hydrocarbons, where patterns pertain to chain length, unsaturations, and branching.

Individual molecules that possess uniquely biogenic conformations (*e.g.*, isoprenoids, cyclopropyl fatty acids) are identified, along with a list of each individual fatty acid and acyclic hydrocarbon reported in the dataset. Trends in sample processing techniques and analytical methods used to characterize lipid molecules and distributions are also recorded.

3.1. Fatty acids

3.1.1. Trends and patterns in fatty acid distributions. Fatty acids extracted from terrestrial (893 samples) and meteoritic (58 samples) specimens (Table 1) display trends in (1) chain length range and distribution; (2) presence, frequency, and degree of unsaturations; and (3) length, frequency, and positions of branches in chains (Fig. 1) that can link to either biotic or abiotic origins (Figs. 2–6, S6–S8).

3.1.2. Fatty acid chain length distributions. The terrestrial samples in our dataset include fatty acids with main chain lengths that range from 4 to 34 carbon atoms (C_4 to C_{34}). The shortest fatty acid in each sample ranges from C_4 to C_{20} , with C_{14} being the most frequent minimum length (44% [391/893] of samples). The longest fatty acid in each of these samples ranges from C_{10} to C_{34} , with C_{18} being the most frequent maximum length (21% [188/893] of samples) (Fig. 2). Predominance of either even or odd chain lengths occurs for 279 terrestrial samples; almost all samples (278/279) exhibit predominance of even chain lengths over odd (Fig. S6), while only one sample exhibits odd chain length predominance. While the data for the remaining 614 samples do not include even/odd predominance information, it is important to note that a lack of reported information does not necessarily imply a lack of predominance.

Meteoritic samples in our dataset include fatty acids with main chain lengths that range from C_1 to C_{12} . The shortest fatty acid in each sample is either C_1 (47% [27/58] of samples), C_2 (52% [30/58] of samples), or C_3 (2% [1/58] of samples). The longest fatty acid in these samples ranges from C_3 to C_{12} , with a maximum length of C_{10} occurring most frequently (50% [29/58] of samples) (Fig. 2). No abiotic samples report a chain length predominance (Fig. S6).

3.1.3. Fatty acid chain lengths and chain configurations: Most abundant fatty acid. In studies of terrestrial samples, the most abundant fatty acid (*i.e.*, the fatty acid with the highest relative concentration compared to other fatty acids within the same sample) contains between 4 and 28 carbon atoms in the main chain; this information is reported for 863 of the 893 total samples in the dataset. The straight-chain, saturated $C_{16:0}$ fatty acid is the most abundant molecule in the majority of samples analyzed (54% [467/863] of samples) (Fig. 3a). The fatty acid with the second highest abundance in each sample contains between 4 and 28 carbon atoms in the main chain, as reported for 826 of the 893 total samples in the dataset. The molecule with the second highest abundance is $C_{16:0}$ in 21% (173/826) of the samples, $C_{18:0}$ in 19% (158/826) of the samples, and $C_{18:1}$ in 16% (135/826) of the samples (Fig. 3b).

The configuration of the dominant fatty acid in terrestrial studies is unbranched and saturated (with variable chain

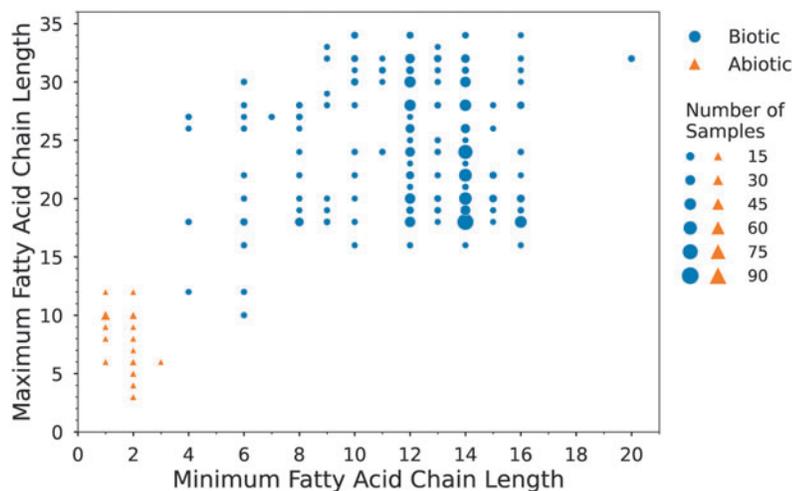


FIG. 2. Minimum and maximum chain length of fatty acids found in each biotic (blue circles) and abiotic (orange triangles) lipid sample analyzed in this study. Minimum fatty acid chain length refers to the number of carbons in the shortest fatty acid chain, and the maximum fatty acid chain length enumerates the number of carbons in the longest fatty acid chain. Symbol size scales with the number of samples analyzed.

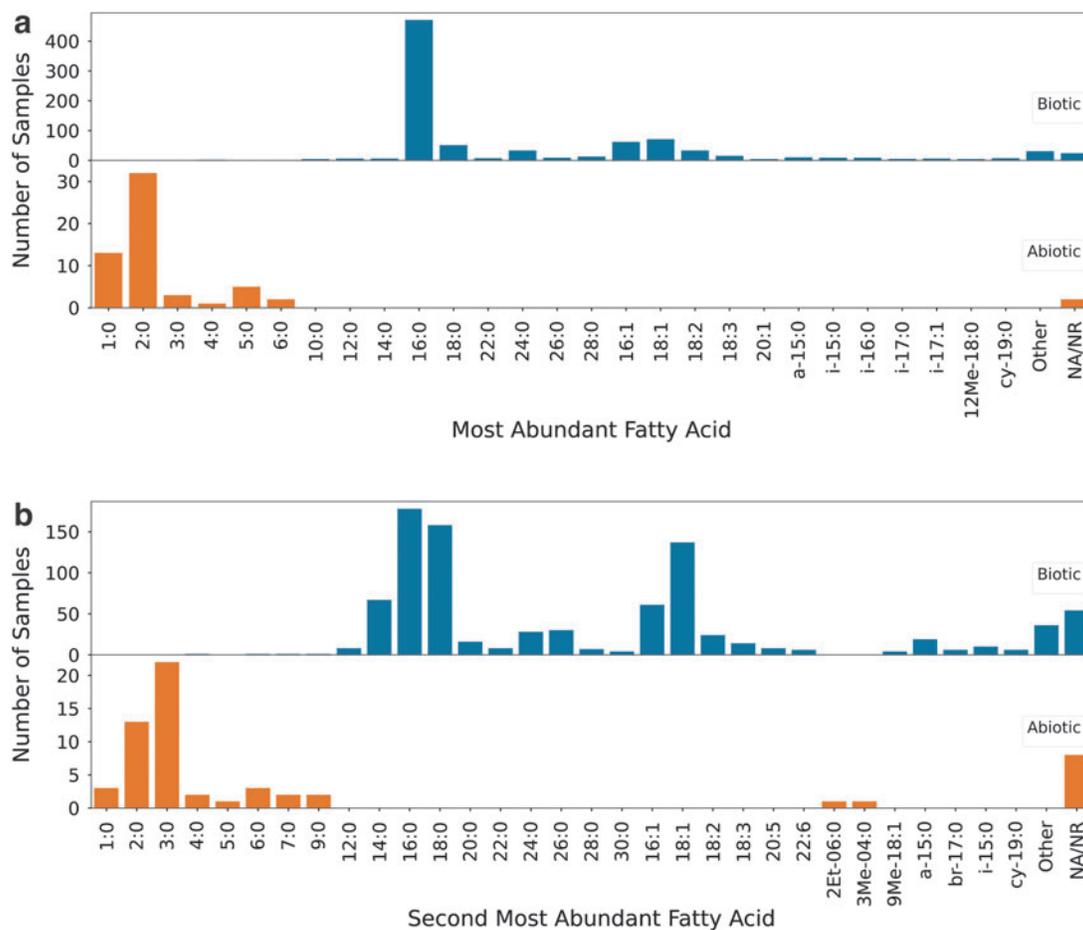


FIG. 3. Most abundant (a) and second most abundant (b) fatty acids reported for each biotic (blue) and abiotic (orange) sample, where X:Y (horizontal axis) refers to (X) the number of carbons in the main chain and (Y) the number of double bonds between carbons within that main chain. Vertical axis is the number of samples. a- indicates *anteiso* branching, i- indicates *iso* branching, Me denotes a methyl branch (preceding number refers to the position of that branch within the main chain of the molecule), Et denotes an ethyl branch (preceding number is the position of the branch), br- refers to a branched molecule with the branch length and position not specified, and cy- describes a cyclopropyl fatty acid. For clarity, for biotic samples, the most abundant and second most abundant fatty acids having less than 3 occurrences were grouped and displayed as “Other.” “NA/NR” refers to samples where the most abundant or second most abundant fatty acid is not reported. Due to axis scales, some of the lower abundance (*i.e.*, 4–10 individual samples or counts) members of the displayed dataset may be difficult to see and counted as “Other.”

lengths) in 70% (608/863) of the samples, monounsaturated in 17% (143/863), polyunsaturated with 2–6 double bonds in 7% (58/863), both *iso* branched and saturated in 3% (23/863), *anteiso* branched in 1% (10/863), monomethyl branched in 1% (9/863), both *iso* branched and monounsaturated in 0.7% (6/863), and cyclopropyl in 0.7% (6/863) (Fig. 3a).

For meteoritic studies, the most abundant fatty acid reported in each sample contains between 1 and 6 carbon atoms in the main chain and is most frequently C_{2:0} in 57% (32/56) of the samples, as studies for 2 of the 58 total samples do not identify the dominant fatty acid. The second most abundant fatty acid for these samples contains between 1 and 9 carbon atoms in the main chain and is most frequently the C_{3:0} fatty acid in 39% (22/56) of the samples (Fig. 3b).

The configuration of the dominant fatty acid in all abiotic samples is straight-chained (*i.e.*, non-branched) and saturated (Fig. 3a). The second most abundant fatty acid in these samples is straight-chained and saturated in 96% (54/56) of the samples and branched in the remaining 4% (2/56) of samples (Fig. 3b).

3.1.4. Distribution of unsaturated fatty acids. In biotic studies, the presence of unsaturated fatty acids is reported in 77% (692/893) of the samples, while the remaining samples contain only saturated species. Quantitative information on the number, frequency, and degree of unsaturations are detailed for 670 samples (Fig. 4). Approximately 37% (248/670) of these samples contain monounsaturated fatty acids only, while the remaining 63% (422/670) contain polyunsaturated fatty acids with up to 6 double bonds in a single chain (Fig. 4). The most abundant unsaturated fatty acid (*i.e.*, with the highest relative concentration compared to other unsaturated fatty acids within the sample) is a monounsaturated C_{18:1} in 47% (318/670) of the samples and C_{16:1} in 28% (185/670) (Fig. S7).

Unsaturated fatty acids are infrequently identified in meteoritic studies; within the dataset, only 2 of 58 samples are reported to contain fatty acids with double bonds. Both these samples are from the Tagish Lake meteorite (Herd *et al.*, 2011; Hilts *et al.*, 2014), and both samples contain the same two monounsaturated C_{4:1} fatty acid isomers (*cis*-C_{4:1} and *trans*-C_{4:1}) each (Figs. 4, S7).

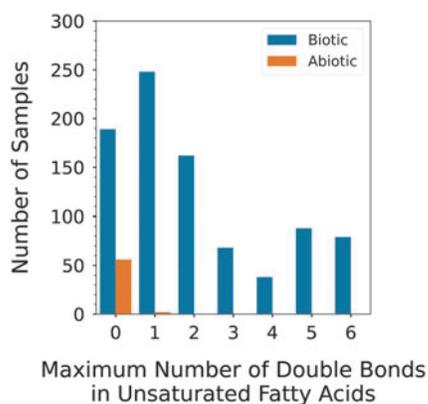


FIG. 4. Maximum number of unsaturations (*i.e.*, C=C bonds) in a single fatty acid, for the biotic (blue) and abiotic (orange) lipid samples.

3.1.5. Distributions of branched fatty acids. Branched fatty acids are reported in 63% (562/893) of terrestrial samples. Of the studies reviewed, branched fatty acids consist of between 6 and 32 carbon atoms in the main chain, each molecule contains between 1 and 5 branches that extend off this main chain, and individual branches contain one carbon atom, that is, methyl (Me) branches only, except for one study reporting three samples that contain a single ethyl (Et) branched fatty acid each (Malherbe *et al.*, 2017). The position of the branches can vary but most frequently occurs between the middle of the main chain and the terminal carbon atom, that is, opposite from the carboxyl group. *Iso* and *anteiso* configurations are favored. The most abundant branched fatty acid (*i.e.*, molecule with the highest concentration relative to other branched fatty acids within the same sample) in biotic samples is most frequently *anteiso*-C_{15:0} (in 34%, or 151/446 samples that report this information), followed by *iso*-C_{15:0} (in 25%, or 110/446 samples that report this information) (Fig. 5).

The main chain lengths of branched fatty acids are reported for 551 of the 562 terrestrial samples that report branching. In 51% (281/551) of these samples, the shortest branched fatty acid has a main chain length of 15 carbon atoms. In 41% (224/551) of the samples, the longest branched fatty acid has a main chain length of 17 carbon atoms (Fig. 6a).

All biotic samples with branched fatty acids explicitly report that the maximum branch length is 1 carbon atom long (*i.e.*, methyl branching only), except for one study (Malherbe *et al.*, 2017) reporting the presence of one ethyl branched fatty acid in each of three samples (Fig. S8). Among samples with branched fatty acids, 92% (504/549) of the samples that report this information contain monomethyl-branched fatty acids only. The remaining 45 samples report fatty acids with 2, 3, 4, or 5 methyl branches in a single molecule, and 31 of these samples contain one or more fatty acids with an isoprenoid configuration (Fig. S8).

For the majority of terrestrial samples with branched fatty acids, the range of branch positions, counted from the carboxyl group, tends to fall between the mid-chain and terminal end (*iso* and *anteiso*), although branching at the second or third carbon atom is occasionally reported (Fig. 6b). The first branching point within the main chain for any branched fatty acid in a sample is most frequently located at the 13th (38%, 203/529 samples that report this positional information) or the 10th carbon atom (31%, 162/529 samples). The first branching point occurs at the 2nd carbon atom in 4.7% (25/529) of the samples and at the 3rd carbon atom in 3.6% (19/529) of the samples. Finally, the last branching point within the main chain for any branched fatty acid in a sample is most frequently located at the 16th (in 58%, or 314 of 545 samples that report this positional information), 14th (15% [81/545] of samples), or 15th carbon atom (12% [67/545] of samples) (Fig. 6b).

The configuration of the most abundant branched fatty acid is reported for 449 terrestrial samples; it is *iso* branched in 48% (217/449) and *anteiso* branched in 43% (193/449) of the samples. The most abundant branched fatty acid displays mid-chain branching with a methyl group located at the 9th, 10th, or 12th carbon atom within the main chain in 5.6% (25/449) of the samples, 2-Me branching in one sample, tetramethyl branching with an isoprenoid configuration for 2.2% (10/449) of the samples, and 2-Et branching in 0.7% (3/449) (Fig. 5).

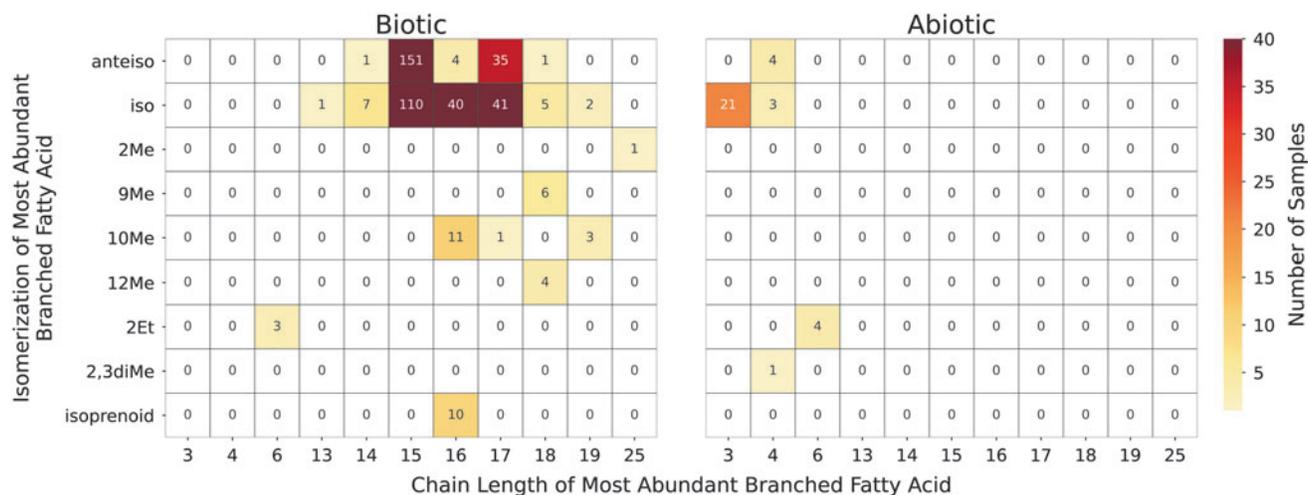


FIG. 5. Isomerization of the most abundant branched fatty acids within the biotic (left) and abiotic (right) samples. The box shade corresponds to the number of samples, with red indicating more samples and yellow indicating fewer. Abbreviations: *iso* = methyl branch at penultimate position relative the carboxyl end; *anteiso* = methyl branch at the antepenultimate position relative the carboxyl end; Me = single methyl branch (preceding number refers to position of the branch relative the carboxyl end); diMe = dimethyl branches (preceding numbers are positions relative the carboxyl); Et = ethyl branch (preceding number refers to position relative the carboxyl); isoprenoid = isoprenoid configuration.

Branched fatty acids are reported in 79% (46/58) of the meteoritic samples. For these samples, branched fatty acids have 3–10 carbon atoms in the main chain, 1–2 branches that extend off this main chain, and individual branches are 1–3 carbon atoms long. The most abundant branched fatty acid is 2-Me-C_{3:0} (*i.e.*, *iso*-C_{3:0}) in 64% (21/33) of the samples that contain branched fatty acids and for which the identity of the most abundant branched fatty acid is re-

ported; reports for 13 samples that contain branched fatty acids provide information on the branching positions but do not provide relative abundances for individual molecules.

The shortest branched fatty acid in each meteoritic sample has 3–6 carbon atoms in the main chain, with a minimum length of 3 carbon atoms in 80% (37/46) of the samples. The longest branched fatty acid in meteorite samples contains 3–10 carbon atoms in the main chain, with

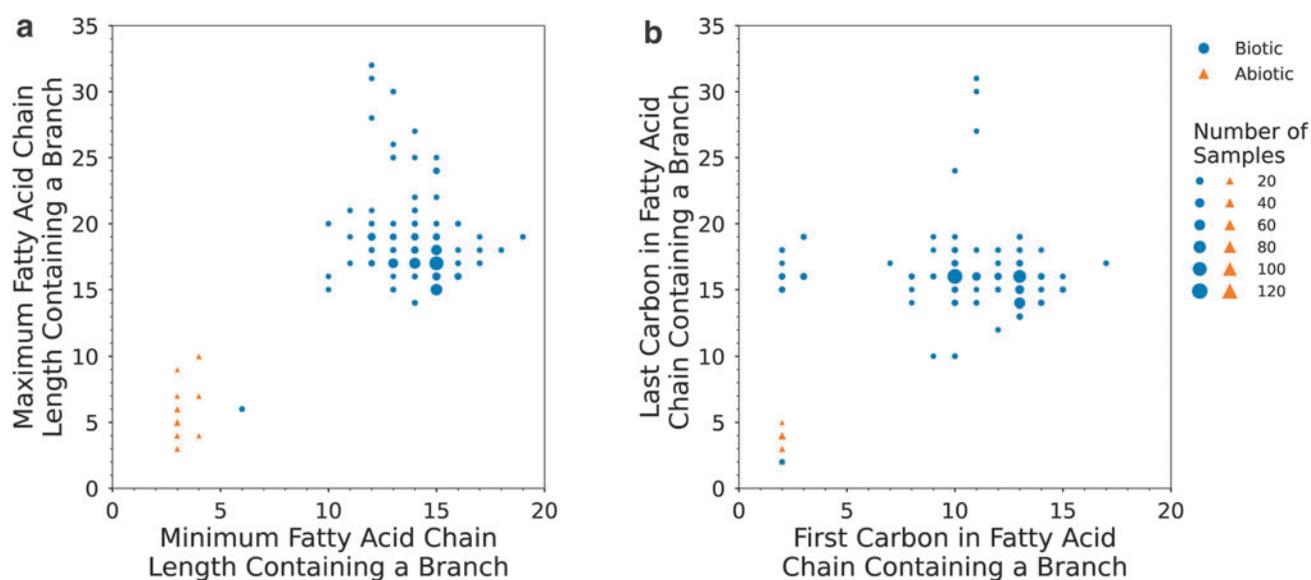


FIG. 6. (a) Minimum (*i.e.*, shortest) and maximum (*i.e.*, longest) chain lengths containing a branch for each reported fatty acid sample from biotic (blue circles) and abiotic (orange triangles) studies. (b) First and last carbon atom in the chain that contains a branch refers to the position of those branches within the main chain of the molecule, for fatty acids found in biotic (blue circles) and abiotic (orange triangles) lipid samples. The position is counted by the number of carbon atoms within the main chain, relative to the carboxyl group of the molecule. Size of circles and triangles indicates the number of samples.

a maximum length of 5 carbon atoms in 30% (14/46) of the samples and a maximum length of 6 carbon atoms in 24% (11/46) of the samples (Fig. 6a).

Information on the number of branches is reported for 41 of the 46 abiotic samples with branched molecules. A maximum of either one (59% [24/41] of samples) or two (41% [17/41] of samples) branches is identified in any single fatty acid chain (Fig. S8). The length of those branches is reported for 40 samples; 55% (22/40) of samples contain fatty acids with methyl branches only, but ethyl-(Et) and propyl-branched fatty acids are identified in 43% (17/40) and 2.5% (1/40) of the samples, respectively (Fig. S8).

Of the 46 meteoritic samples that contain branched fatty acids, analyses for 40 of them provide information on the range of branching positions. For all 40 of these samples, the first branching point within the main chain for any fatty acid within the sample is located at the 2nd carbon atom. The last branching point within the main chain for any fatty acid within the main chain is located between the 2nd and 5th carbon atom and is most frequently at the 4th carbon atom in 53% (21/40) of the samples, followed by the 3rd, 2nd, and 5th carbon atom in 23% (9/40), 18% (7/40), and 2.5% (1/40) of the samples, respectively (Fig. 6b).

The identity of the most abundant branched fatty acid is reported for 33 of the 46 meteoritic samples that contain branched molecules. The most abundant branched fatty acid is monomethyl branched in the majority (29/33) of the samples and is 2-Me-C_{3:0} (*i.e.*, *iso*-C_{3:0}) in 64% (21/33) of the samples, 2-Me-C_{4:0} (*i.e.*, *anteiso*-C_{4:0}) in 12% (4/33), and 3-Me-C_{4:0} (*i.e.*, *iso*-C_{4:0}) in 9% (3/33). In the remaining samples, the most abundant branched fatty acid is either ethyl branched (2-Et-C_{6:0} in 12% [4/33] of samples) or dimethyl branched (2,3-dimethyl-C_{4:0} in 3% [1/33] of samples) (Fig. 5).

3.1.6. Biogenic fatty acids. Certain subgroups of fatty acids are only found in samples of biotic origin. These fatty acids bear diagnostic configurations or repeating structural elements that are uniquely biotic, as they are inextricably linked to biotic synthesis or modification. These include cyclopropyl fatty acids (Grogan and Cronan, 1997) and isoprenoid fatty acids (Summons *et al.*, 2022) (Fig. 1), which have not been reported as indigenous in meteorites to our knowledge.

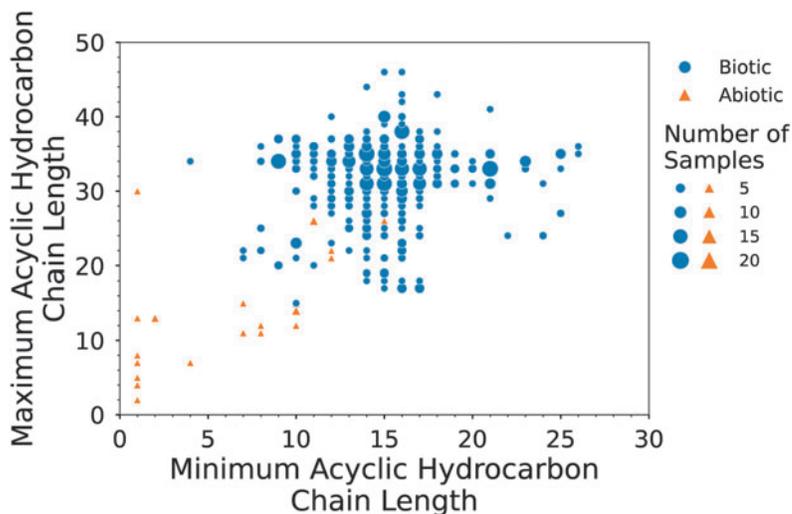
3.2. Trends and patterns in acyclic hydrocarbon distributions

Acyclic hydrocarbons in 592 terrestrial samples (Table 1) display trends in chain length ranges and distributions, saturations, and branching (Figs. 7–10, S9–S15). Because of the low concentrations (compared to fatty acids), presence of UCMs, widely reported contamination issues (Cronin and Pizzarello, 1990; Sephton *et al.*, 2001a, 2001b), and comparatively fewer studies published on acyclic hydrocarbons extracted from meteorites, trends from the 31 meteoritic samples reported in 14 studies only allowed for discernment of clear patterns in chain length distribution, even/odd chain-length predominance, and the presence of branches or unsaturations in the chain (Figs. 7–10, S9–S15). However, numerous trends were identified within the biotic samples.

3.2.1. Acyclic hydrocarbon chain length distributions. The terrestrial samples in our dataset include acyclic hydrocarbons with main chain lengths that range from C₄ to C₄₆. The shortest acyclic hydrocarbon in each biotic sample ranges from C₄ to C₂₆, with a minimum length of C₁₅ occurring in 17% (100/592) of the samples, C₁₆ occurring in 16% (97/592), and C₁₄ in 16% (93/592). The longest acyclic hydrocarbon in these samples ranges from C₁₅ to C₄₆, with a maximum chain length of C₃₃ occurring in 15% (88/592) of the samples, C₃₅ in 13% (74/592), C₃₄ in 12% (73/592), and C₃₁ in 12% (71/592) (Fig. 7). For biotic samples, unimodal chain length distributions are reported in 15% (88/592) of samples, bimodal distributions are reported in 12% (72/592) of samples, a trimodal distribution is reported in 1 sample, a uniform distribution is reported in 3 samples, and reports for the remaining 428 samples do not include this information (Fig. S9a). A predominance of odd (over even) chain lengths is reported for 51% (302/592) of the samples, while a predominance of even chain lengths is reported for 10% (60/592) of the samples. Approximately 9% (55/592) of the samples exhibit no predominance of either even or odd, and the remaining 30% (175/592) of the samples do not report this information (Fig. S9b).

Meteoritic samples in our dataset include acyclic hydrocarbons with main chain lengths that range from C₁ to C₃₁. The shortest acyclic hydrocarbon in each sample ranges

FIG. 7. Minimum and maximum chain length of acyclic hydrocarbons found in each biotic (blue circles) and abiotic (orange triangles) lipid sample analyzed in this study. Minimum acyclic hydrocarbon chain length refers to the number of carbons in the shortest acyclic hydrocarbon chain, and the maximum acyclic hydrocarbon chain length enumerates the number of carbons in the longest acyclic hydrocarbon chain. Symbol size scales with the number of samples analyzed.



from C_1 to C_{16} , with a minimum length of C_1 occurring in 23% (7/31) of the samples and a minimum length of C_{10} occurring in 19% (6/31) of the samples. The longest acyclic hydrocarbon in each sample ranges from C_7 to C_{31} , with a maximum length of C_{26} reported in 16% (5/31) of the samples and a maximum length of either C_{13} , C_{14} , or C_{15} each reported in 10% (3/31) of the samples (Fig. 7). Unimodal chain length distributions are reported in 16% (5/31) of abiotic samples, a uniform distribution is reported in 1 sample, and the remaining 81% (25/31) of samples do not report this information (Fig. S9a). None of the meteoritic acyclic hydrocarbon samples were reported to display a preference for either even or odd chain lengths (Fig. S9b).

3.2.2. Acyclic hydrocarbon chain lengths and chain configurations: Most abundant acyclic hydrocarbon. The most abundant acyclic hydrocarbon reported in terrestrial samples contains between 10 and 33 carbon atoms in the main chain; this information is reported for 487 of the 592 total samples in the dataset. The single most abundant acyclic hydrocarbon is a $C_{27:0}$ *n*-alkane in 12% (57/487) of the samples and a $C_{17:0}$ *n*-alkane in 11% (53/487) of the samples. The acyclic hydrocarbon with the second highest abundance in each sample contains between 10 and 31 carbon atoms in the main chain; this information is reported for 347 of the 592 terrestrial samples. The molecule with the second highest abundance is a $C_{29:0}$ *n*-alkane in 17% (58/347) of the samples and a $C_{27:0}$ *n*-alkane in 13% (45/347) of the samples (Fig. 8a).

The configuration of the most abundant acyclic hydrocarbon in terrestrial samples is an *n*-alkane (*i.e.*, straight-chain and saturated, lacking branches) in 83% (404/487) of the samples, a monounsaturated alkene in 1.8% (9/487), polyunsaturated with 2–7 double bonds (including some isoprenoids) in 2.5% (12/487) of the samples, and monomethyl branched in 1.4% (7/487) of the samples. An isoprenoid species is reported to be the most abundant acyclic hydrocarbon in 13% (64/487) of the samples, and this configuration includes both saturated and unsaturated species (Fig. 8a).

In meteoritic studies, the chain length of the most abundant acyclic hydrocarbon is only identified and reported in 19 out of 31 samples, and the chain length of the second most abundant acyclic hydrocarbon is reported for 18 of these samples. For those samples, the most abundant acyclic hydrocarbon contains between 1 and 26 carbon atoms in the main chain and is most frequently C_1 (in 26% [5/19] of samples) or $C_{14:0}$ (in 16% [3/19] of samples). In the other 11 samples, the most abundant acyclic hydrocarbon is a straight chain, unsaturated *n*-alkane with variable chain lengths (Fig. 8a). The second most abundant acyclic hydrocarbon in these samples contains between 1 and 25 carbon atoms in the main chain, often with branches, and there is no clear result for the most frequent chain length. The second most abundant acyclic hydrocarbon is a monounsaturated $C_{2:1}$ alkene in one sample (Levy *et al.*, 1973), but in every other instance, the reported configuration of the second most abundant acyclic hydrocarbon is straight-chained, unsaturated, and unbranched (Fig. 8b).

3.2.3. Distribution of unsaturated acyclic hydrocarbons. Unsaturated acyclic hydrocarbons (*i.e.*, alkenes)

with 1–7 double bonds are reported in 16% (95/592) of the biotic samples (Figs. S10, S11). The most abundant alkene in biotic samples is most frequently an isoprenoid possessing one or more double bonds (in 55%, or 52/95 of samples) (Fig. S11).

Alkenes are reported in 9 out of 31 abiotic samples; among these, the majority (7/9) derive from studies on acyclic hydrocarbons extracted from the IOM, as opposed to free compounds extracted from the soluble fractions of the meteorites. Liberation of these alkene fragments from the larger IOM structure requires additional processing steps that often employ mineral dissolution with HF or HCl and/or high temperatures (*i.e.*, pyrolysis) to break the oxygen and alkyl bridges that bind fragments into the larger macromolecular matrix (*e.g.*, Levy *et al.*, 1973; Shimoyama, 1997; Wang *et al.*, 2005; Remusat *et al.*, 2007; Okumura and Mimura, 2011). The identity of the most abundant unsaturated acyclic hydrocarbon is only reported in two of these samples and is $C_{2:1}$ in both cases (Levy *et al.*, 1973; Yuen *et al.*, 1984) (Fig. S11).

3.2.4. Distributions of branched acyclic hydrocarbons. Branched acyclic hydrocarbons are common in both terrestrial and meteoritic samples, but the structures of all isomers within a sample are not always characterized, especially in meteorites. UCMs are usually reported in meteorites and sometimes in terrestrial samples, but for biotic samples, certain branched molecules are present in higher relative abundances. These resolved structures and distributions are included in our analysis of branching patterns, and UCMs are addressed separately. Meteoritic samples rarely identify individual branched species above background UCMs.

In studies of terrestrial samples, 70% (414/592) of the samples contain branched acyclic hydrocarbons with between 4 and 41 carbon atoms in the main chain; each molecule contains between 1 and 8 branches that extend off this main chain, and individual branches contain between 1 and 6 carbon atoms. Branches are positioned between the 2nd and 33rd carbon atom within the main chain, but branching most frequently begins at the 2nd carbon atom. Isoprenoids are the most common branched configuration, and the majority of the samples with branched acyclic hydrocarbons contain one or more isoprenoids. Pristane or phytane is often the most abundant branched molecule within a sample (Fig. S11).

The main chain lengths of branched acyclic hydrocarbons are reported for 381 of the 415 samples that report branching. Among these samples, the shortest branched acyclic hydrocarbon contains between 4 and 24 carbon atoms in the main chain, with a minimum length of 15 carbon atoms occurring in 35% (135/381) of the samples. The longest branched acyclic hydrocarbon in these samples contains between 12 and 41 carbon atoms in the main chain, with a maximum length of 16 carbon atoms occurring in 32% (122/381) of the samples (Fig. 9a).

In biotic samples, branched acyclic hydrocarbons contain between 1 and 8 branches in a single molecule, with a maximum of 4 individual branches in 63% (258/412) of the samples that report this information (Fig. S12a). It is reported that 87% (348/400) of the samples contain molecules with methyl branches only, and the remaining 13% (52/400)

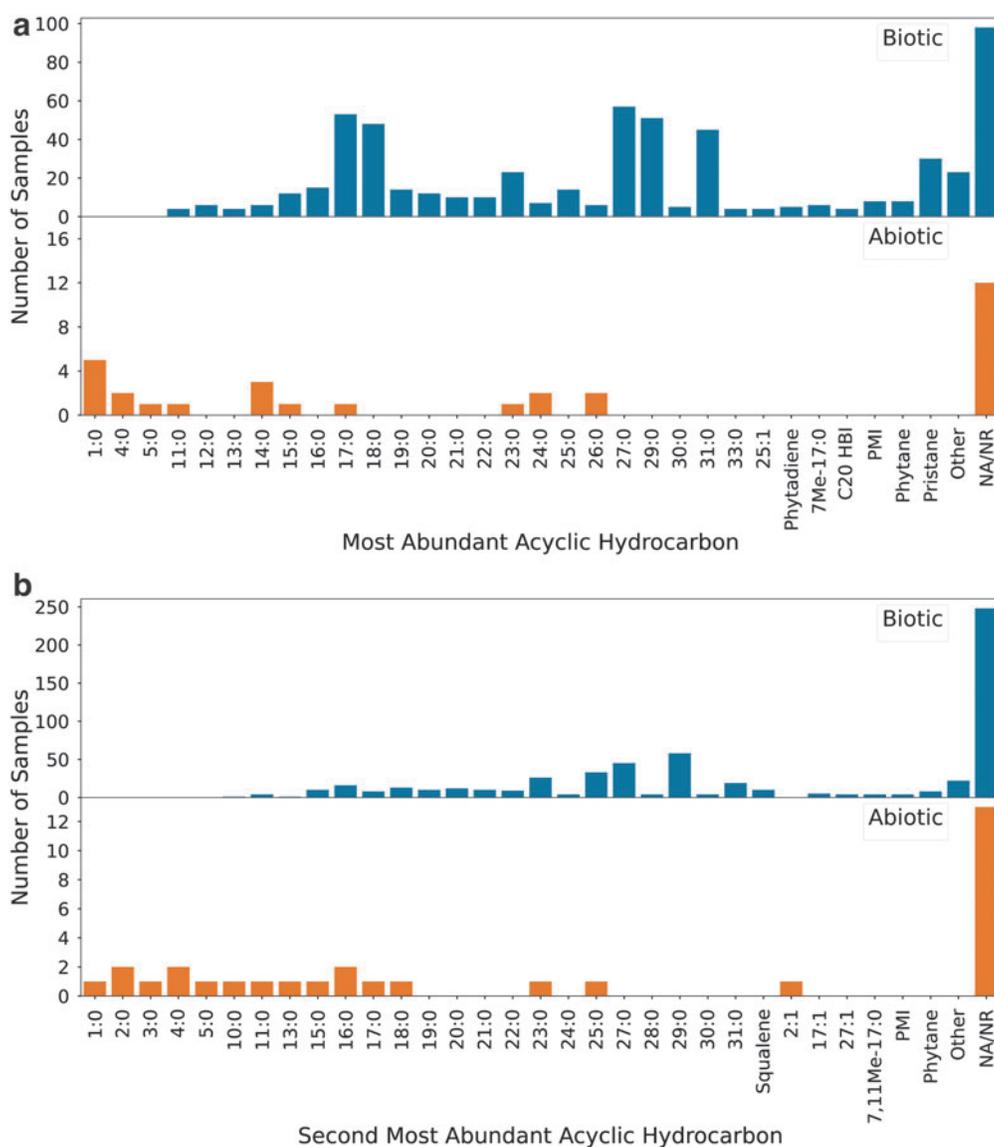


FIG. 8. Most abundant (a) and second most abundant (b) acyclic hydrocarbons reported for each biotic (blue) and abiotic (orange) sample, where X:Y (horizontal axis) refers to (X) the number of carbons in the main chain and (Y) the number of double bonds between carbons within that main chain. Vertical axis is the number of samples. Isoprenoid configurations include phytadiene, highly branched isoprenoids containing 20 carbon atoms (C20 HBI), pentamethyleicosane (PMI), phytane, pristane, and squalene. Me denotes a methyl branch (preceding number or numbers refer to the position of that branch within the main chain of the molecule). For clarity, for biotic samples, the most abundant and second most abundant acyclic hydrocarbons having less than 3 occurrences were grouped and displayed as “Other.” “NA/NR” (*i.e.*, not available or not reported) refers to samples where the most abundant or second most abundant acyclic hydrocarbon is not reported. Due to axis scales, some of the lower abundance (*i.e.*, 4–10 individual samples or counts) members of the displayed dataset may be difficult to see and counted as “Other.”

of the samples contain individual branches up to 6 carbon atoms long (Fig. S12b). Complex, highly branched isoprenoids are often present, but details of these structures (*i.e.*, number and length of branches, configuration) are typically either not reported or reported as “low confidence.”

For the majority of the terrestrial samples with branched acyclic hydrocarbons, the range of branching positions usually begins at the 2nd carbon atom in the main chain then extends to the mid-chain or terminal end (Fig. 9b). The first branching point within the main chain for any branched acyclic hydrocarbon in a sample is most frequently at the 2nd

carbon atom (93% [383/412] of samples). For the remaining 7.0% (29/412) of the samples, the first branching point occurs at the 3rd, 4th, 5th, 6th, or 7th carbon atom. The last branching point within the main chain for any branched acyclic hydrocarbon in a sample falls between the 2nd and 32nd carbon atom and is most frequently located at the 14th carbon atom, in 63% (259/412) of the samples (Fig. 9b).

The configuration of the dominant branched acyclic hydrocarbon is reported for 405 terrestrial samples and is an isoprenoid in 77% (310/405) of these samples (Fig. 10). These molecules contain multiple (between 3 and 8) methyl

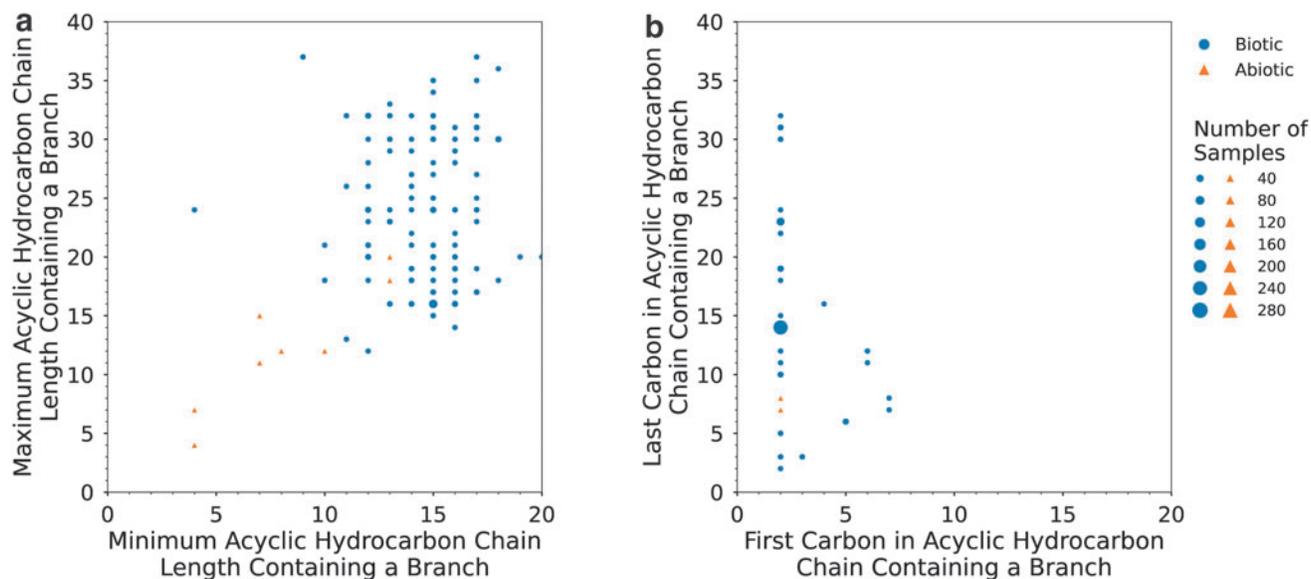


FIG. 9. (a) Minimum (*i.e.*, shortest) and maximum (*i.e.*, longest) chain lengths containing a branch for each reported acyclic sample from biotic (blue circles) and abiotic (orange triangles) studies. (b) First and last carbon atom in the chain that contains a branch refers to the position of those branches within the main chain of the molecule, for acyclic hydrocarbons found in biotic (blue circles) and abiotic (orange triangles) lipid samples. The position of the branch is counted from the end of the molecule that is nearest to a branch. Size of circles and triangles indicates the number of samples.

branches spaced evenly throughout the length of the main chain (Figs. 1, S13–S15). Occasionally, these isoprenoids also possess one or more double bonds (*e.g.*, squalene, phytadiene).

In meteorites, branched acyclic hydrocarbons are reported in 52% (16/31) of the samples, but the identities, positions, and configurations of these molecules typically are not

comprehensively reported (only detailed in 5–9 samples). This is likely due to the structural complexity and poor chromatographic resolution of low-abundance branched acyclic hydrocarbons. Therefore, trends in meteoritic branched acyclic hydrocarbons cannot be determined from so few data points; however, information on the structural features, positions, and isomerization is cataloged.

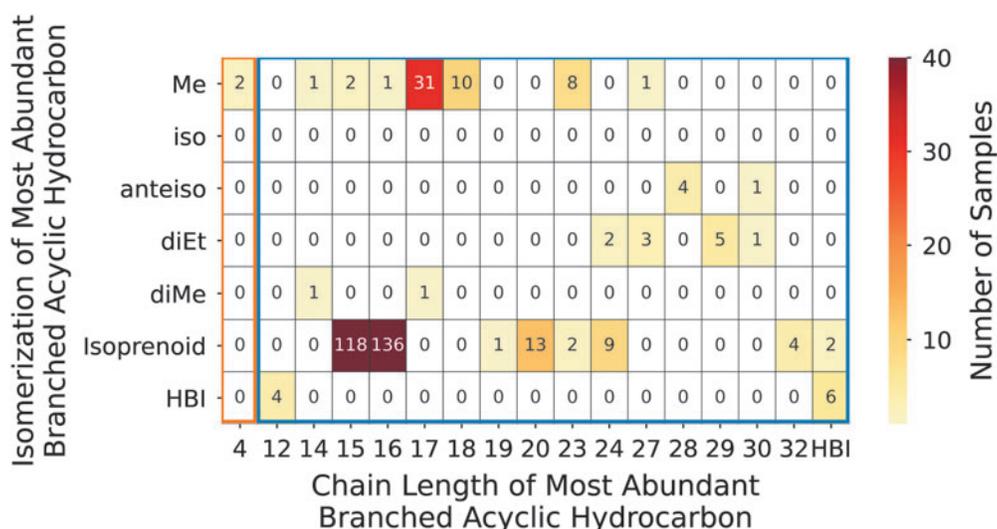


FIG. 10. Isomerization of the most abundant branched acyclic hydrocarbons within the biotic (blue outline) and abiotic (orange outline) samples, where biotic and abiotic samples are displayed in one plot but binned within respective blue and orange outlines. The box shade corresponds to the number of samples, with red indicating more samples and yellow indicating fewer. Abbreviations: *iso*=methyl branch at the 2-Me position (counted as penultimate position, as these molecules are thought to derive from *iso* branched fatty acids); *anteiso*=methyl branch at the 3-Me position (counted as antepenultimate position, as these molecules are thought to derive from *anteiso* branched fatty acids); Me=single methyl branch (position not specified); diMe=dimethyl branches (positions not specified); diEt=diethyl branches (position not specified); isoprenoid=isoprenoid configuration; HBI=highly branched isoprenoid.

Chain length information is reported in 9 samples, for which the shortest branched hydrocarbon contains between 4 and 13 carbon atoms in the main chain and the longest branched hydrocarbon contains between 4 and 20 carbon atoms in the main chain. Information on branch number and length is reported in 5 samples that contain molecules with between 1 and 5 individual branches (Fig. S12a); individual branches contain between 1 and 3 carbon atoms each (methyl, ethyl, or propyl) (Fig. S12b). Branching position information is reported in 6 samples, for which branches fall between the 2nd and 10th carbon atom within the main chain. The identity of the dominant branched acyclic hydrocarbon is reported in 2 samples; in both cases, the most abundant molecule is monomethyl branched, with 4 carbon atoms in the main chain (Levy *et al.*, 1973; Yuen *et al.*, 1984).

3.2.5. Unresolved complex mixtures and homologous series. Unresolved complex mixtures (UCMs) are reported in 25% of 592 terrestrial samples and 71% of 31 meteoritic samples (Fig. S16a). In biotic systems, certain molecules and homologous series are discernable, with the balance of the mixture unresolved (Figs. S16b–S16d). Series of monomethylalkanes are reported in 29% of 592 samples, defined as a sequence of molecules of varying chain lengths, each possessing a single methyl branch located at variable positions on the main hydrocarbon chain. Series of isoalkanes (*i.e.*, 2-methylalkanes), which are thought to derive from decarboxylated *iso*-fatty acids (Peters and Moldowan, 1993), are reported in 13% of 592 samples (Fig. S16c), and series of *anteiso*-alkanes (*i.e.*, 3-methylalkanes) deriving from decarboxylated *anteiso*-fatty acids are reported in 12% of 592 samples (Fig. S16d).

3.2.6. Biogenic acyclic hydrocarbons. Aliphatic isoprenoids are diagnostic biosignatures and reported in 62% of 592 terrestrial samples. In 13% of the 487 of these samples for

which the information is reported, the most abundant acyclic hydrocarbon (including all acyclic hydrocarbons in the sample) is an isoprenoid. Up to 26 unique isoprenoids are reported in a single sample, but most frequently there are 2 unique isoprenoids present in a given sample, reported in 39% of these 335 samples (Fig. S14). Pristane or phytane are usually the dominant isoprenoid present in a sample (81% of 331 samples that identify the most abundant isoprenoid) (Fig. S13), but many other isoprenoids occur, possessing between 13 and 40 carbon atoms in any given molecule (Figs. S14, S15); the isoprenoids are arranged in either straight-chain, head–head, head–tail, tail–tail, or highly branched configurations.

3.3. Principal component analysis of fatty acid and acyclic hydrocarbon features

3.3.1. Fatty acid PCA. The concatenated biotic and abiotic fatty acid datasets consist of 381 terrestrial and 31 meteoritic samples with a total of 16 of the features we individually analyzed (Table S1). Approximately 80% of the variance in the dataset was contained within the first three principal components. The *k*-means clustering algorithm reveals two clearly distinguishable clusters that correctly differentiate samples as biotic or abiotic (Figs. 11, S17). Our results suggest that the two leading principal components derived from lipid features can be used to distinguish a sample's origin as biogenic or abiogenic. This confirms the utility of lipids for life-detection applications and supports the results of our supervised learning analyses. Of the 16 features (Table S1) included in our PCA (Figs. 11, S18), the parameters with the greatest influence on separation between the biotic and abiotic clusters include minimum and maximum chain length (Fig. 2), most abundant unique fatty acid (Fig. 3a), and second most abundant unique fatty acid (Fig. 3b).

3.3.2. Acyclic hydrocarbon PCA. After calculating the Gower metric and applying PCA and *k*-means to the

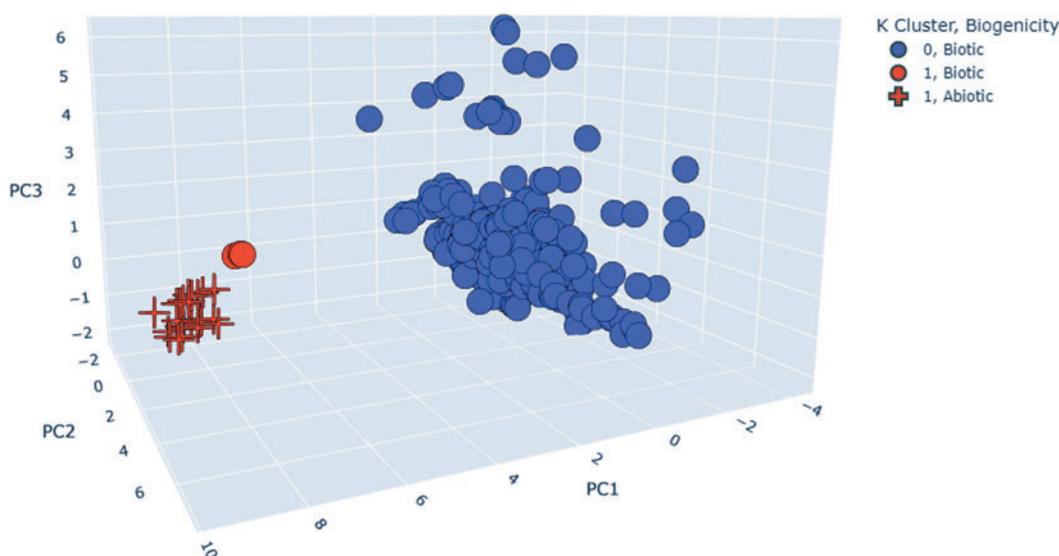


FIG. 11. Samples from the terrestrial and meteoritic fatty acid datasets are differentiated as biotic (circles) or abiotic (crosses), using a 16-parameter PCA (listed in Fig. S1). Separation is visualized in 3-D by the first three principal components. Three distinct clusters are identified, where red crosses are abiotic, blue circles are biotic, and red circles are a distinct biotic subgroup characterized by shorter chain lengths and distinct branching patterns.

resulting dissimilarity matrix, we identify three distinct clusters (red: #0, blue: #1, green: #2) within the 14-parameter analysis (Table S2). Most abiotic samples (cross) are well clustered in blue while the remaining biotic samples (circle) are evenly distributed between three cluster sets (Table S2; Figs. 12, S18). We find samples that fall within cluster #1 are samples without MMA and homologous series features. We observe an abiotic outlier in cluster #0 (red), and this outlier is the only abiotic sample that reports the presence of MMA and homologous series; presence of MMA and homologous series are common among samples within this cluster #0. In addition, the presence of indigenous isoprenoids is an important feature in differentiating between biotic and abiotic samples in the biotic sample type determined by clustering (blue: cluster #1). We also observe that cluster #2 (green) comprises samples that have the presence of *iso/anteiso*-alkanes suggesting their importance in differentiating between different types of biotic samples. Taken together, these results suggest the possibility of several types of distinct profiles that each may indicate biogenicity.

3.4. Trends in lipid extraction and analytical techniques

3.4.1. Solvent-based techniques for extracting lipids. In the studies reviewed, solvent-based techniques are most commonly used to extract fatty acids and acyclic hydrocarbons from natural, geologic, or environmental samples, reported in 83% of the 1574 samples included in our study. These techniques are further binned below (Figs. 13a, 13b). Water extraction (without addition of organic solvents) was used in 74% of the 58 meteoritic fatty acid samples and 1 of 31 meteoritic acyclic hydrocarbon samples, but none of the terrestrial samples. Pyrolysis (thermal extraction) techniques were used for

7% of 1574 samples, and the remaining 7% of reported analyses used a variety of chemical extraction techniques that rely on neither solvents nor pyrolysis as the primary method for liberating lipids. Solvent refluxed through sample in a closed vessel with use of a commercially available apparatus (*i.e.*, Soxhlet, ASE) was reported in 22% of all samples, including 17% (165/951) of fatty acid samples and 29% (182/623) of acyclic hydrocarbon samples (Figs. 13a, 13b).

The five most common extraction techniques for fatty acids and acyclic hydrocarbons leverage organic solvents (occasionally with added water, or buffer to adjust pH) in tandem with other sample processing steps that can vary with the apparatus (*e.g.*, Soxhlet, accelerated solvent extractor), pressure, temperature, sonic energy applied, and/or solvent type and ratio. A post-extraction, pre-analysis derivatization or methylation step is typically applied for fatty acids to increase volatility for GC-MS, but these procedures are not included in our review.

For fatty acids, the five most common techniques are, in decreasing order of frequency, (1) Modified Bligh and Dyer (2:2:1.8 [v/v/v] ratio of methanol, water, and chloroform or dichloromethane—this hallmark ratio of solvents defines both modified and traditional Bligh and Dyer methods) (Bligh and Dyer, 1959); (2) solvent extraction (an individual, sequence, or cocktail of organic solvent without reporting any use of commercial instrumentation, refluxing, or ultrasonication); (3) ultrasonic extraction (organic solvents with the addition of ultrasonic energy) (Keris-Sen *et al.*, 2014); (4) Soxhlet (organic solvent and sample are refluxed) (Luque de Castro and Priego-Capote, 2010); and (5) accelerated solvent extraction (ASE) (organic solvent is introduced under high temperature and pressure via a commercially available instrument) (Richter *et al.*, 1996) (Fig. 13). For acyclic hydrocarbons, the five most common extraction techniques are, in decreasing order of frequency,

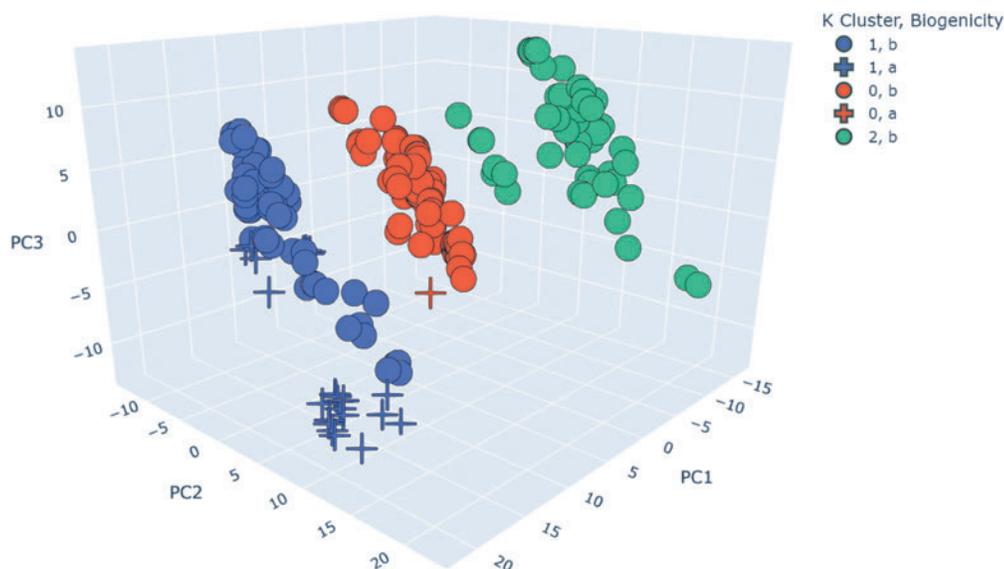


FIG. 12. Three distinct hydrocarbon profiles are identified from a 14-parameter PCA (listed in Table S2). Separation is visualized in 3-D by the first three principal components. Three distinct clusters (red: #0, blue: #1, green: #2). Samples that fall within cluster #1 (blue) are samples without MMA and homologous series features while cluster #2 (green) consists of samples that have the presence of *iso/anteiso*-alkanes. For abiotic samples (crosses), the presence of indigenous isoprenoids is apparently an important feature in differentiating between biotic and abiotic samples in the biotic sample type determined by clustering (blue: cluster #1).

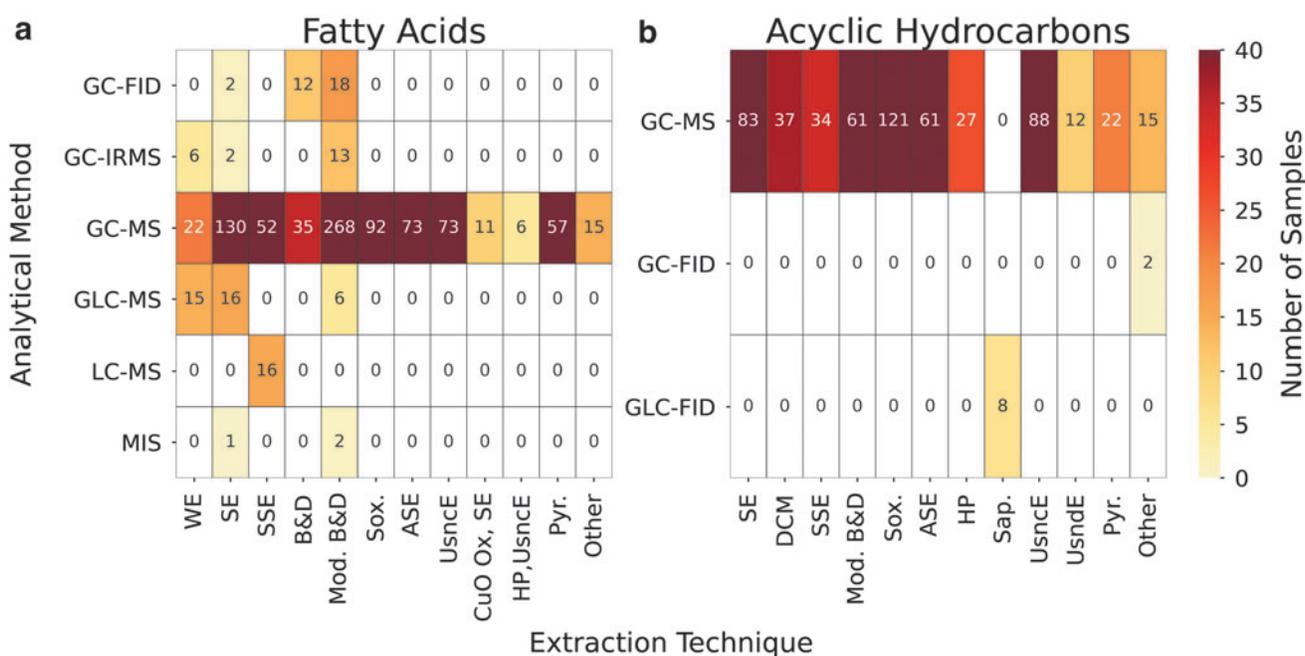


FIG. 13. Extraction technique and analytical method in samples containing fatty acids (a) and acyclic hydrocarbons (b) show that gas chromatography–mass spectrometry (GC-MS) is the most commonly used analytical technique, and solvent extraction techniques (top five most frequent for both fatty acids and acyclic hydrocarbons) are the most commonly used. The box shade corresponds to the number of samples, with red indicating more samples and yellow indicating fewer. Analytical method abbreviations: GC-FID=gas chromatography–flame ionization detection; GC-IRMS=gas chromatography–isotope ratio mass spectrometry; GC-MS=gas chromatography–mass spectrometry; GLC-MS=gas liquid chromatography–mass spectrometry; LC-MS=liquid chromatography–mass spectrometry; MIS=microbial identification system; GLC-FID=gas liquid chromatography–flame ionization detection. Extraction technique abbreviations: WE=water extraction; SE=solvent extraction (without additional defined processing steps); SSE=sequential solvent extraction (use of differing organic solvents in sequence applied to the same sample); B&D=Modified Bligh and Dyer; Mod. B&D=Modified Bligh and Dyer; Sox.=Soxhlet; ASE=accelerated solvent extraction; UsncE=ultrasonic assisted extraction; CuO Ox, SE=cupric oxide oxidation followed by solvent extraction; HP, UsncE=hydrous pyrolysis followed by ultrasonic assisted extraction; Pyr.=pyrolysis (*i.e.*, thermal extraction); DCM=dichloromethane (*i.e.*, where dichloromethane was the solvent used for extraction, without use of additional techniques or apparatuses previously described, such as ASE, Soxhlet, or ultrasonic agitation); HP=hydrous pyrolysis; Sap.=saponification; UsndE=ultrasound assisted extraction. Extraction technique reports for fewer than 3 individual samples were combined and displayed as “Other.” The maximum value shown on the color scale bar is 40 occurrences to highlight categories with lower frequencies.

(1) Soxhlet, (2) ultrasonic extraction, (3) solvent extraction, (4) ASE, and (5) Modified Bligh and Dyer. Due to the wide range of complex, multistep extraction techniques utilized in the cited studies, some overlap may exist between the categories we delineate (*e.g.*, “solvent extraction” encompasses numerous sequences, sonic energy may be added during some Modified Bligh and Dyer methodologies); however, all are solvent-based (Fig. 13).

Two of the five most frequent extraction techniques for both fatty acids and acyclic hydrocarbons utilize a commercially available sample processing unit to extract organics from pre-ground samples. These include Soxhlet and ASE, and both work by refluxing organic solvent through samples in a closed vessel with variable times, temperatures, and solvent cocktails (Fig. 13). Following extraction, the analyte is separated from any residual minerals via filtration, producing a purified lipid extract for downstream analysis.

3.4.2. Analytical techniques for characterizing lipids. The analytical method most frequently used for molecule

identification was gas chromatography–mass spectrometry (GC-MS), leveraged in 90% (1397/1574) of the samples (Fig. 13). Identification using mass spectrometry (MS) was by far the most common (97%, 1532/1574 samples).

4. Discussion

4.1. Origin-diagnostic distributions of fatty acids and acyclic lipids can indicate biotic or abiotic origin

In the search for signs of life beyond Earth, an ideal molecular biosignature should (i) be fundamental to life as we know it or can imagine it based on carbon chemistry in water (Dorn *et al.*, 2011; Georgiou and Deamer, 2014; Neveu *et al.*, 2018), (ii) possess one or more structural features and distributions that are distinct from abiotically produced counterparts (*e.g.*, meteoritic or hydrothermal organics within the same molecular classes) (McCollom *et al.*, 1999; Mißbach *et al.*, 2018), (iii) display forms and conformations that reflect evolution and indicate function within a cell (*e.g.*, metabolite, structural component, information storage, etc.) (Boucher

et al., 2004; Summons *et al.*, 2022), (iv) exhibit preservation potential over geologically relevant timescales (Peters and Moldowan, 1993; Brocks and Schaeffer, 2008; Lee and Brocks, 2011; Grotzinger *et al.*, 2014), and (v) be analyzable with techniques that can be adapted to spaceflight (Lovell, 1965; Mahaffy *et al.*, 2012). The acyclic lipid groups we cataloged—fatty acids and acyclic hydrocarbons—fulfill these criteria, demonstrating their utility as ideal astrobiological targets.

For the terrestrial and meteoritic lipid data we analyzed, each examined molecular structure (*e.g.*, chain length, branching, double bonds) and its distribution (*e.g.*, frequency, range, position, predominance) within a sample falls on a spectrum based on both presence and diagnosticity and can be impacted by how well the structure is preserved throughout geologic timescales. Furthermore, a set of fatty acids or acyclic hydrocarbons present in a single terrestrial or extraterrestrial sample typically contains multiple indicators of biogenicity or abiogenicity, including (1) patterns in chain length ranges and distribution; (2) presence, frequency, and degree of unsaturations within a chain; or (3) presence, frequency, number, length, and position of branches.

In total, we identify 15 potential origin-diagnostic distributions for fatty acids (Table 2) found in natural samples and 12 potential origin-diagnostic distributions for acyclic hydrocarbons (Table 3), which are representative of the acyclic lipid data we analyzed. Deeper analysis is likely to reveal additional trends not reported here. In addition to these distributions, the presence of unique, individual molecules that themselves contain repeating patterns in structure (*i.e.*, isoprenoids), and/or preference for one or more specific conformers (*e.g.*, *iso*-C_{15:0} fatty acid, C_{18:1} fatty acid, pristane, etc.) can constitute a potential biosignature, if preferential synthesis of that conformer is kinetically unlikely or thermodynamically unfavorable in the context of known abiotic reactions (*e.g.*, observed in the natural environment and/or in laboratory synthesis experiments) or expected abiotic scenarios (Bernstein *et al.*, 1995; Mißbach *et al.*, 2018; Nuevo *et al.*, 2018; Sandford *et al.*, 2020). These findings reiterate the astrobiological utility of acyclic lipids as a ubiquitous class of organics that can provide a uniquely rich and well-preserved range of origin-diagnostic information displayed by physical parameters in molecular structures (Georgiou and Deamer, 2014).

4.1.1. Origin-diagnostic distributions of fatty acids. Key origin-diagnostic distributions of biotic fatty acids revealed by our study include (i) chain lengths that can range from C₄ to C₃₄ but more frequently fall between C₁₄ and C₁₈, with (ii) a C_{max} that usually peaks at C₁₆ or C₁₈, and (iii) a predominance of even-numbered fatty acids (Figs. 2, 3, S6); (iv) frequent mono- and polyunsaturated molecules with up to 6 double bonds in a single chain (Figs. 4, S7); (v) branched fatty acids with main chain lengths that can range between 6 and 32 carbon atoms long but are most frequently restricted to ranges between 15 and 17 carbon atoms long, (vi) methyl groups (vii) positioned from the second carbon atom (adjacent the carboxyl group) to the terminal end, with mid-chain (*e.g.*, 9-Me, 10-Me) and terminal (*e.g.*, *iso*, *anteiso*) positions most common (Figs. 5, 6, S8); and (viii) occasional fatty isoprenoids or cyclopropyl fatty acids.

The molecular structures of these fatty acids are reflective of well-known biochemical mechanisms and demonstrative of cellular functionality. For example, C_{16:0} and C_{18:0} fatty acids are preferentially synthesized to support membrane geometry in both prokaryotes and eukaryotes, indicating that these traits emerged early and have persisted throughout the history of life on Earth (Coskun and Simons, 2011; Koga, 2012). The incorporation of (poly)unsaturations, branching, and/or cyclopropyl groups into fatty acid tails serves as an adaptation to regulate fluidity in cold environments (*i.e.*, ≤ 40°C) by creating space between molecules that make up lipid bilayers (Grogan and Cronan, 1997; Hagve, 1988). Absent these structural additions, closely packed saturated/unbranched molecules with these chain lengths would otherwise exist in a gel state at near- to subfreezing temperatures, leading to membrane stiffening and loss of cell function by preventing passage of solutes in and out of the cell (Hazel and Eugene Williams, 1990; Mansy, 2009).

Hallmark fatty acid distributions observed in abiotic meteorite samples include (i) chain length ranges from C₁ to C₁₂, with (ii) a C_{max} that peaks at lower molecular weights (*e.g.*, C₁, C₂, C₃), and (iii) no predominance of even versus odd carbon atom number (Figs. 2, 3, S6); (iv) rare unsaturations (Figs. 4, S7); (v) branched fatty acids with main chain lengths that range between 3 and 10 carbon atoms long, (vi) branching positions that always begin at the second carbon atom but can extend throughout the length of the main chain, (vii) individual branch lengths that range between 1 and 3 carbon atoms long with (viii) randomized isomerization (Figs. 5, 6, S8), and (ix) no isoprenoid or cyclopropyl fatty acids.

Laboratory experiments simulating the formation of organic compounds via energetic processing of ices at low temperature (<80 K) have shown that the chemistry taking place is one of opportunity, in which molecules, radicals, and ions react with their closest neighbor, rather than a chemistry driven by thermodynamics (Sandford *et al.*, 2020). The resulting products typically display distributions in which smaller compounds are the most abundant and the abundances of larger compounds decrease exponentially with increasing carbon-chain length, as has also been observed for amino acids and sugar derivatives (Nuevo *et al.*, 2008, 2018; Meinert *et al.*, 2016). Other experiments simulating FTT reactions also suggest that while chain length distributions display Poisson distributions, there is no preference for specific isomers or nonrandom positioning of molecular features within fatty acids. This is illustrated by the shorter chain lengths and highly branched molecules that characterize the abiotic fatty acid distributions in our dataset (McCullom *et al.*, 1999; Rushdi and Simoneit, 2001; Mißbach *et al.*, 2018).

4.1.2. Origin-diagnostic acyclic hydrocarbon distributions. Biotically synthesized acyclic hydrocarbons are characterized by (i) chain lengths that can range from C₄ to C₄₆ but more frequently fall between C₁₅ and C₃₄, with (ii) a C_{max} that often peaks at C₁₇ or C₂₇, and (iii) unimodal, bimodal, or trimodal chain length distributions with (iv) an occasional preference for either odd or even carbon number (Figs. 7, 8, S9), (v) frequent mono- and polyunsaturated molecules with up to 7 double bonds in a single chain (Figs. S10, S11), (vi) branched acyclic hydrocarbons with

TABLE 2. ORIGIN-DIAGNOSTIC PATTERNS AND DISTRIBUTIONS FOR BIOTIC AND ABIOTIC FATTY ACIDS

<i>Origin-Diagnostic Patterns and Distributions: Fatty Acids</i>							
<i>Always present; always origin-diagnostic</i>							
<i>Feature</i>	<i>Pattern</i>	<i>Category</i>	<i>Fig. #</i>	<i>Feature</i>	<i>Pattern</i>	<i>Category</i>	<i>Fig. #</i>
Chain length	Minimum chain length	Distribution	2	Chain length	Maximum chain length (>12=biotic; <10=abiotic)	Distribution	2
				Chain length	Even or odd preference (even=biotic; odd=biotic)	Distribution	S6
				Chain length	Most abundant molecule	Indv. conformer	3
				Chain length	2 nd most abundant molecule	Indv. conformer	3
<i>Sometimes present; always origin-diagnostic</i>							
<i>Sometimes present; sometimes origin-diagnostic</i>							
<i>Feature</i>	<i>Pattern</i>	<i>Category</i>	<i>Fig. #</i>	<i>Feature</i>	<i>Pattern</i>	<i>Category</i>	<i>Fig. #</i>
Unsaturations	Most abundant unsaturated molecule	Indv. conformer	S7	Unsaturations	# of unsaturated molecules (>2=biotic)	Distribution	
Branching	Min-max chain length with a branch	Range	5	Unsaturations	Max 3 unsaturations in a single molecule (>1=biotic)	Distribution	4
Biogenic compounds	Isoprenoids (biotic only)	Indv. conformer		Branching	Most abundant branched molecule	Indv. conformer	6
Biogenic compounds	Cyclopropyl fatty acids (biotic only)	Indv. conformer		Branching	First-last carbon in chain with a branch (mid-chain or terminal=biotic)	Range	5
				Branching	Max # branches in a single molecule (>1=biotic)	Distribution	S8
				Branching	Max carbons in a single branch (>2=abiotic)	Distribution	

TABLE 3. ORIGIN-DIAGNOSTIC PATTERNS AND DISTRIBUTIONS FOR BIOTIC AND ABIOTIC ACYCLIC HYDROCARBONS

<i>Origin-Diagnostic Patterns and Distributions: Acyclic Hydrocarbons</i>				
<i>Always present; always origin-diagnostic</i>				
<i>Feature</i>	<i>Pattern</i>	<i>Category</i>	<i>Fig. #</i>	<i>Fig. #</i>
				<i>Always present; sometimes origin-diagnostic</i>
Chain length	Min chain length	Distribution		7
Chain length	Even or odd preference (even = biotic; odd = biotic)	Distribution		S9
Chain length	Distribution of chain lengths (bimodal = biotic; trimodal = biotic)	Distribution		S9
Chain length	Most abundant molecule	Indv. conformer		8a
Chain length	2 nd most abundant molecule	Indv. conformer		8b
<i>Sometimes present; always origin-diagnostic</i>				
<i>Feature</i>	<i>Pattern</i>	<i>Category</i>	<i>Fig. #</i>	<i>Fig. #</i>
				<i>Sometimes present; sometimes origin-diagnostic</i>
Biogenic compounds	Isoprenoids (biotic only)	Indv. conformer	S15	Distribution
Branching	Homologous series (isoalkanes, anteisoalkanes = biotic)	Distribution	S16c-S16d	Range
		Branching		Distribution
		Branching		Distribution
		Branching		9

main chain lengths that can range between 4 and 41 carbon atoms long, (vii) branching positions that typically begin at the second carbon atom and extend to the mid-chain or terminal end, (viii) frequent methyl branching but occasionally long and complex individual branches, and (ix) a clear predominance of isoprenoids with variable carbon atom numbers and configurations (Figs. 9, 10, S13–S15).

Monomethyl branched alkanes are often synthesized directly by various organisms, but can also represent diagenetic products of membrane fatty acids that have undergone decarboxylation. *N*-alkanes with odd or even chain length preferences can be similarly sourced from fatty acids or biosynthesized via head-to-head condensation and decarboxylation of fatty acids (Peters and Moldowan, 1993; Ladygina *et al.*, 2006; Georgiou and Deamer, 2014). UCMs are sometimes present in older, degraded, or thermally processed samples, but resolvable acyclic hydrocarbons with diagnostic structures and distributions typically rise well above this background. Other branching patterns are due to the presence of isoprenoids, which are not known to form abiotically. While chlorophyll is the source of the geologically ubiquitous isoprenoids pristane and phytane, which are common to terrestrial samples, many other types of acyclic isoprenoids are sourced from archaeal membranes, and the numerous branches contained within these hydrocarbon chains reinforce stability in high-temperature or extreme-pH environments (*e.g.*, Peters and Moldowan 1993; Summons *et al.*, 2022). Incorporation of double bonds within these lipid chains regulates fluidity at lower temperatures, as with eukaryotic and prokaryotic fatty acids (Kaneda, 1991; Summons *et al.*, 2022). Membrane-stabilizing isoprenoids are sometimes incorporated into bacterial membranes as well (Jordan *et al.*, 2019). Isoprene units also serve as bioessential metabolites and are subcomponents or precursors of pigments, hormones, vitamins, membrane-stabilizing polycyclic hydrocarbons, and other life-enabling molecules in all branches of the tree of life (Zeng and Dehesh, 2021).

Meteoritic acyclic hydrocarbon distributions are not as well-constrained as terrestrial acyclic hydrocarbons (or meteoritic fatty acids) but typically include molecules that are characterized by (i) chain lengths that range from C_1 to no longer than C_{30} , with (ii) a C_{max} that may peak at C_1 or C_{14} , and (iii) random or unimodal chain length distributions with (iv) no predominance of even or odd carbon number (Figs. 7, 8, and S9), (v) occasional unsaturations (*i.e.*, most typically in IOM-sourced components, where IOM is a kerogen-like macromolecule containing smaller organic fragments bound in a complex organic matrix via alkyl and oxygen bridges) (Figs. S10, S11), (vi) the presence of a complex mixture of highly branched molecules that exhibit wide structural diversity and contain countless low-abundance isomers with no clear preference for one configuration (as the analytical methods used in the studies reviewed did not have the resolution to obtain this information), and (vii) no indigenous isoprenoids or similar molecules.

Total hydrocarbons in meteorites are usually only present in concentrations at part-per-million to part-per-billion levels, that is, typically an order of magnitude less abundant than total fatty acids in the same specimens, and are dominated by UCMs of cyclic and acyclic molecules (Sephton, 2006; Pizzarello and Shock, 2010). These random structures

that lack preference for specific isomers, chain lengths, and patterns in unsaturations or branching are reflective of formation, which is thought to proceed via many of the same astrochemical and geochemical processes (*e.g.*, molecule-ion-radical reactions during energetic processing of ices) responsible for fatty acid synthesis in these primitive extraterrestrial materials (*e.g.*, Nuevo *et al.*, 2008, 2018; Meinert *et al.*, 2016; Sandford *et al.*, 2020). However, while UCMs comprise a significant majority of meteoritic acyclic hydrocarbons, some “resolvable” acyclic hydrocarbons are present in high individual abundances relative to the background UCM (*e.g.*, *n*-alkanes with unimodal chain lengths), possibly explained by FTT synthesis (Levy *et al.*, 1973; Yuen *et al.*, 1984; Shimoyama, 1997; Wang *et al.*, 2005; Hiltz *et al.*, 2014; Simkus *et al.*, 2019). Alternatively, a few studies have suggested that the ^{13}C signatures of free *n*-alkanes in carbonaceous meteorites point to terrestrial contamination source, as opposed to an extraterrestrial origin (Cronin and Pizzarello, 1990; Sephton *et al.*, 2001a, 2001b). Other studies of meteoritic IOM find that *n*-alkanes released at high temperatures via pyrolysis bear isotopic signatures consistent with an extraterrestrial origin (Wang *et al.*, 2005; Okumura and Mimura, 2011). Indigenous isoprenoids are notably absent in meteorites.

4.2. Acyclic lipid distributions and the physical approach to life detection

The numerous origin-diagnostic distributions revealed by our study reaffirm the applicability of Lovelock’s physical approach to life detection, while expanding the number of potential biosignatures that may be used to discern origin within extraterrestrial acyclic lipid samples, reinforcing the utility of this class of organics in the search for life (Dorn *et al.*, 2011; Georgiou and Deamer, 2014). Critically, our analysis quantitatively demonstrates that, on a global scale, distributions of acyclic lipid biomarkers fall within specific ranges and display unique and nonrandom distributions, even when samples contain potentially mixed signals from numerous organisms and geologic epochs. From an evolutionary perspective, molecular form begets biological function, and each of the structural features we examined plays a life-enabling role. The genes coding for these biosynthetic pathways emerged very early in the tree of life, and while this basic lipid template remained static, it was decorated in response to changing environments and to infer greater fitness (*e.g.*, incorporation of double bonds, branching, and specific chain lengths to reinforce membrane stability and/or modulate membrane fluidity) (Hazel and Eugene Williams, 1990; Segré *et al.*, 2001; Boucher *et al.*, 2004). Although individual conformations and distributions of acyclic membrane lipids can be unique to specific organisms, multiple distinct patterns are both broadly observed and preserved in the rock record across Earth, as illustrated by the distributions we delineate in this study (*e.g.*, Peters and Moldowan 1993; Summons *et al.*, 2008).

Our results indicate that the determination of life does not require the presence of specific biomolecules that may be unique to terrestrial organisms (*e.g.*, DNA, proteins, hopanoids), but rather can be demonstrated by analysis of generalized physical parameters within monomers. Finally, identification of a sample’s origin does not hinge on

detection of only one diagnostic molecule, pattern, or structure, and confidence in origin is strengthened by the identification of multiple, additive origin-diagnostic distributions of independent structural features that can occur within a single sample (Figs. 11, 12, S17, S18).

4.3. The spectrum of origin-diagnosticity

Results from our analyses indicate that, if sets of fatty acids or acyclic hydrocarbons are detected in a geologic sample, distributions of structural elements within that set of lipids can fall into one of the following four categories along what we term the “spectrum of origin-diagnosticity”: (1) always present and always origin-diagnostic, (2) sometimes present and always origin-diagnostic, (3) sometimes present and sometimes origin-diagnostic, (4) always present and sometimes origin-diagnostic (Table 2). “Presence” refers to whether a given structural feature is observed in a sample in some or all cases (*e.g.*, chain length is always an observable parameter, but molecules with branches or unsaturations are only present in some samples). “Origin-diagnostic” refers to whether the distribution of that feature is different for biotic versus abiotic lipids in all cases analyzed within our dataset, or only in some cases. A lack of certain patterns or features does not necessarily indicate a lack of biogenicity, but the presence of multiple distributions can provide evidence in favor. Further, the distributions of individual features can differ from one biotic sample to the next, depending on the class (fatty acid or acyclic hydrocarbon), source input, sample age, mineralogy, environment, or diagenetic history, but for all the terrestrial lipid samples we reviewed, the distributions of one or more features differ relative to meteoritic lipids.

The parameters most closely linked with origin for both fatty acids and acyclic hydrocarbons include (i) minimum and maximum chain length (Figs. 2, 7), (ii) chain length distribution and predominance (Figs. S6, S9), (iii) identity of the most abundant and second most abundant molecules within a sample (Figs. 3, 8), (iv) number and length of branches (Figs. S8 and S12), (v) position and conformation of branching (Figs. 5, 6, 9, 10), and (vi) degree of unsaturation (Figs. 4, S10).

4.3.1. Origin-diagnostic features and patterns in fatty acids. Our analysis reveals 15 potential origin-diagnostic distributions in fatty acids that can indicate whether a sample is biogenic or abiogenic (Table 2). Some patterns (*e.g.*, min and max chain length) are inextricably linked to (bio)synthesis and provide strong evidence on origin, while other distributions (*e.g.*, branch position and length) are less diagnostic but can still provide added information to increase confidence in assessing origin, when identified in concert with other diagnostic distributions within the same sample.

4.3.1.1. Always present and always origin-diagnostic: Fatty acids. Fatty acid distributions that are always present and always origin-diagnostic include:

- (1) Minimum chain length of the molecules within a sample (Fig. 2).

Biotic fatty acids have longer chains than abiotic fatty acids and typically range from C₄ to C₃₀ (and rarely, longer), although short-chain (C₄-C₁₂) molecules are relatively uncommon. In contrast, meteoritic fatty acids are shorter, with chain lengths ranging from C₁ to C₁₂. For all biotic samples, the minimum chain length is C₄ or longer, while for all abiotic samples, the minimum chain length is C₃ or shorter.

4.3.1.2. Always present but sometimes origin-diagnostic: Fatty acids. Fatty acid distributions that are always present but sometimes origin-diagnostic include:

- (1) Maximum chain length present (Fig. 2);
- (2) Even or odd preference (Fig. S6);
- (3) Identity of the most abundant molecule (Fig. 3a);
- (4) Identity of the second most abundant molecule (Fig. 3b).

For all terrestrial samples, minimum–maximum chain length range always differs relative to all meteoritic samples. However, while biotic fatty acids are longer, these samples can contain short-chain molecules that fall within the “abiotic” range. When the maximum chain length in a fatty acid sample contains more than 12 carbons in the main chain, this indicates biogenicity. When the maximum chain length in a fatty acid sample contains fewer than 10 carbons in the main chain, this indicates abiogenicity. However, in our study, there are studies in both the biotic and abiotic datasets that report the maximum fatty acid chain length in a sample is either a C₁₀ or C₁₂, so these distributions are non-diagnostic of origin.

A predominance of even or odd chain lengths is only observed in biotic samples and reflects synthesis via addition of 2-carbon groups (*e.g.*, Brindley *et al.*, 1969; McCarthy and Hardie, 1984). This sawtooth distribution is often cited as an important lipid biomarker for astrobiological applications (Dorn *et al.*, 2011; Aerts *et al.*, 2014). Within our dataset, a predominance of either even or odd chain lengths is reported for 31% of terrestrial samples (odd predominance for one sample, even predominance for the remainder). Chain length predominance is not explicitly reported for the remaining 69% of the biotic samples in the dataset, and some of these samples only contain fatty acids with one or two different chain lengths, so that an even/odd predominance cannot be measured. No meteorite samples display a predominance of even or odd chain lengths, and predominance is not kinetically or thermodynamically expected for abiotic synthesis (McCullom *et al.*, 1999; Rushdi and Simoneit, 2001; Mißbach *et al.*, 2018). Predominance of either even or odd chain lengths can provide added information indicating a biotic origin, but lack of predominance is non-diagnostic since this pattern is observed in all meteorite and some terrestrial samples. For example, on Earth, diagenesis and microbial degradation are processes that can lead to loss of predominance (*e.g.*, Peters and Moldowan, 1993).

Since biotic and abiotic fatty acids span different chain length ranges, the dominant molecules usually differ. For biotic samples, the most abundant molecule is usually C_{16:0} or C_{18:1}, and the second most abundant molecule is usually C_{16:0} or C_{18:0}. Preferential synthesis of these chain lengths is

enzyme-modulated and well-suited for building the membrane bilayers that support cell structure, while unsaturations enable membrane fluidity at lower temperatures (Hagve, 1988; Shivaji and Prakash, 2010). For abiotic samples, the most abundant and second most abundant molecules are usually the shortest chained species within those sets, typically C₁, C_{2:0}, or C_{3:0}. This reflects the Poisson distribution of chain lengths that characterizes molecules that form abiotically from synthesis that proceeds via addition of single carbon atoms. However, for three samples from two biotic studies (Malherbe *et al.*, 2017; Williams *et al.*, 2021), the most abundant fatty acid is either C_{4:0} or C_{6:0}; several meteorite studies also report that a C_{4:0} or a C_{6:0} fatty acid is the most abundant molecule. Similarly, the second most abundant fatty acid usually differs for biotic and abiotic studies, but a few biotic studies report a C_{4:0}, C_{6:0}, C_{7:0}, or C_{9:0} fatty acid is the second most abundant (Garcette-Lepecq *et al.*, 2004; Williams *et al.*, 2019, 2021).

When the most abundant fatty acid is a C_{10:0} or longer, this indicates biogenicity, and when the most abundant fatty acid is a C_{3:0} or shorter, this indicates abiogenicity. When the second most abundant fatty acid is a C_{12:0} or longer, this indicates biogenicity, and when it is C_{3:0} or shorter, this indicates abiogenicity.

4.3.1.3. Sometimes present but always origin-diagnostic: Fatty acids. Fatty acid distributions that are sometimes present but always origin-diagnostic include

- (1) Identity of the most abundant unsaturated molecule (Fig. S7);
- (2) Minimum–maximum chain length for molecules that contain branches (Fig. 6a);
- (3) Presence of isoprenoid fatty acids;
- (4) Presence of cyclopropyl fatty acids.

For terrestrial samples, the most abundant unsaturated fatty acid is usually C_{18:1} or C_{16:1}. Monounsaturated fatty acids (frequently in the *cis* conformation) are synthesized by a vast array of prokaryotes and eukaryotes, and polyunsaturated molecules are common to a smaller set of organisms (Kaneda, 1991; Mansy, 2009). These double bonds are incorporated at various positions within the main chain of the fatty acid during biosynthesis or via post-enzyme modification to increase fluidity (Shivaji and Prakash, 2010). In meteoritic samples, unsaturated fatty acids are rarely reported, but two of the studies of separate samples of the Tagish Lake meteorite identify the short-chained C_{4:1}, and for both of those samples, both the *cis* and *trans* isomers are present.

For samples that contain branched fatty acids, the minimum and maximum chain length of branched molecules and the identity of the most abundant branched molecule within each sample always differs for terrestrial and meteoritic cases. Biotic fatty acids have longer chains to begin with, and branching is usually restricted to a subset of those fatty acids, with unbranched fatty acids comprising a significant portion of biosynthesized molecules. Branched and unsaturated fatty acids serve similar life-enabling functions in modulating membrane fluidity (Kaneda, 1991; Jordan *et al.*, 2019). Although branched fatty acids with as few as 6 carbon atoms in the main chain are sometimes reported in

terrestrial samples, the shortest branched fatty acid in a sample in the studies we review usually has 15 carbon atoms in the main chain, while the longest branched fatty acid most often has 17 carbon atoms in the main chain. The most abundant branched fatty acid in biotic samples is usually *iso*-C_{15:0} or *anteiso*-C_{15:0}. Unlike biosynthesized fatty acids, abiotic fatty acids display random and frequent branching that appears in molecules of any length. This lack of predominance and presence of branching in molecules of any length reflects synthesis that adds single carbon atoms to any available position as chains are grown, instead of preferentially in specific positions and conformations leveraged by life (Sephton, 2002; Pizzarello, 2006).

Isoprenoid fatty acids and cyclopropyl fatty acids are identified in some terrestrial samples but no meteoritic samples in the studies we review. Isoprenoid fatty acids with repeated, nonrandom branching on Earth are thought to originate from biosynthesized isoprenoid fatty alcohols that are subsequently oxidized by microbes (van den Brink and Wanders, 2006). Cyclopropyl fatty acids derive from enzymatic modification of monounsaturated fatty acids in membrane phospholipids in certain types of bacteria (Yuan *et al.*, 1995; Grogan and Cronan, 1997) and plants (Bao *et al.*, 2002). Synthesis of these “biogenic” conformations requires multiple thermodynamically unfavorable steps and specific biochemical pathways, which can explain why molecules with these conformations have not been identified (and are not expected) in abiotic contexts.

4.3.1.4. Sometimes present and sometimes origin-diagnostic: Fatty acids. Fatty acid distributions that are sometimes present and sometimes origin-diagnostic include

- (1) Number of unique unsaturated molecules in a sample;
- (2) Maximum number of unsaturations present within a single molecule (Fig. 4);
- (3) Identity of the most abundant branched molecule (Fig. 5);
- (4) Range of branch positions within the main chain (*i.e.*, first and last carbon atom containing a branch) (Fig. 6b);
- (5) Maximum number of branches present within a single molecule (Fig. S8a);
- (6) Maximum branch length (*i.e.*, number of carbon atoms within a single branch off the main chain) (Fig. S8b).

Some terrestrial samples contain multiple unique unsaturated fatty acids, but the maximum number of unsaturated fatty acids in any meteorite sample is 2. Terrestrial samples can contain both monounsaturated and polyunsaturated species with up to six double bonds in a single chain, but polyunsaturated fatty acids have not been reported in meteorites. Double bonds are important molecular structures that enable membrane function by modulating fluidity in cold environments (Hazel and Eugene Williams, 1990; Shivaji and Prakash, 2010), and different terrestrial organisms synthesize a diverse array of monounsaturated and polyunsaturated fatty acids. However, some organisms only synthesize saturated species, while diagenetic processes and microbial uptake can oxidize and cleave double bonds, leaving only saturated fatty acids behind (Canuel and

Martens, 1996; Meyers, 1997). Consequently, some terrestrial samples contain no unsaturated fatty acids, only one or two unique fatty acids, or fatty acids with only one double bond in a single chain. This does not indicate a lack of biogenicity, but these distributions are non-diagnostic of origin since they have also been reported in meteorites.

Branch positions and length are sometimes origin-diagnostic. In some biotic samples, the most abundant branched fatty acid contains a methyl branch at the 2-Me or mid-chain (9-Me, 10-Me) position; these configurations serve similar purposes to *iso* and *anteiso* positional isomers (Hazel and Eugene Williams, 1990; Mansy, 2009). Meteoritic samples contain branched fatty acids with between 3 and 10 carbon atoms in the main chain, but a minimum length of 3 and a maximum length of 5 or 6 is most common. The most abundant branched fatty acid in abiotic samples is usually 2-Me-C_{3:0} (*i.e.*, *iso*-C_{3:0}). However, in four meteoritic samples (Shimoyama *et al.*, 1989; Pizzarello *et al.*, 2012; Aponte *et al.*, 2014) and three terrestrial samples (from the same study) (Malherbe *et al.*, 2017), the most abundant branched fatty acid is 2-Et-C_{6:0}, making this conformer non-diagnostic of origin.

Although some terrestrial samples contain fatty acids with branching at the second carbon atom, branching is usually restricted to mid-chain (*e.g.*, 9-Me, 10-Me) or terminal (*i.e.*, *iso*, *anteiso*) positions within the main chain of the molecule. This predominance of specific, nonrandom branching reflects directed synthesis (Hazel and Eugene Williams, 1990; Mansy, 2009). For all meteoritic samples with branched fatty acids, branching begins at the 2-Me position but can continue down the length of the chain. Random and frequent branching at every possible position has been widely reported as a hallmark characteristic of abiotic organics in meteorites (Sephton, 2002; Pizzarello and Shock, 2010), and our analysis reiterates that this distribution is indeed broadly observed.

Likewise, branch length is sometimes origin-diagnostic, as the terrestrial samples in our dataset only contain methyl-branched fatty acids aside from three outliers (from one study; all samples are desert varnish) with one ethyl-branched molecule each (Malherbe *et al.*, 2017), while meteoritic samples contain fatty acids with methyl, ethyl, and propyl branches. Living organisms specifically synthesize fatty acids that contain single-carbon branches to add space between membrane-packed molecules, but abiotic synthesis has no similarly specific process and can randomly add carbon atoms to both branches and main chain. For samples that contain mid- to long-chain branched fatty acids with branch positions that are restricted to the mid-chain to terminal positions, biogenicity may be inferred. However, when branching begins at the 2-Me position (regardless of main chain length) and continues throughout the main chain, branch position ranges are non-diagnostic, as this distribution is observed in both biotic and abiotic samples. When a sample contains fatty acids with branches longer than two or three carbon atoms, this may indicate an abiotic origin. However, for samples that contain methyl-branched fatty acids only, branch length is non-diagnostic because these conformations are observed in samples of either origin.

4.3.2. Origin-diagnostic features and patterns in acyclic hydrocarbons. Origin-diagnostic distributions are less

distinct for acyclic hydrocarbons compared to fatty acids, primarily because there are fewer published studies and lack of quantitative data on meteoritic acyclic hydrocarbons compared to fatty acids, along with historical issues with terrestrial contamination from atmospheric, laboratory, and biological sources, particularly for meteorites that are collected after longer residence time on Earth before their discovery (Cronin and Pizzarello, 1990; Sephton *et al.*, 2001a). Additionally, acyclic hydrocarbons make up a far smaller fraction of meteoritic organics than fatty acids and are dominated by UCM. Terrestrial acyclic hydrocarbon structures and distributions are better constrained than meteoritic, but there exist complexities that complicate elucidation of origin-diagnostic patterns compared to terrestrial fatty acids. For example, while biosynthesized terrestrial fatty acids are primarily sourced from cell membranes, acyclic hydrocarbons can derive from numerous sources and complex parent molecules (*e.g.*, plant waxes, decarboxylated fatty acids, polycyclic compounds, etc.), each with varying distributions of molecular features (Boucher *et al.*, 2004; Summons *et al.*, 2022). Furthermore, biological reprocessing and diagenesis can degrade and transform molecules through oxidation and loss of functionalization, potentially obscuring origin-diagnostic features and distributions.

Despite these challenges, origin-diagnostic information can still be obtained from acyclic hydrocarbons, especially when multiple structural elements are examined together (Figs. 7–10, 12, S9–S15). In particular, identification of the most prevalent and most universal trends that characterize terrestrial biological samples can provide an indication of the types of deviations that should be expected in a biotic extraterrestrial sample relative to a signal dominated by exogenous input (Lovelock, 1965; Georgiou and Deamer, 2014). Despite lack of a comparably extensive set of meteoritic data, the numerous nonrandom and distinct distributions displayed by terrestrial acyclic hydrocarbons in the dataset suggest that extraterrestrial acyclic lipids have the potential to hold key information on biogenicity or abiogenicity (Table 3), but analysis of multiple structural elements and their distributions within a sample is critical for predicting origin.

4.3.2.1. *Always present and always origin-diagnostic.* Of the studies reviewed, no feature for acyclic hydrocarbons is both always present and always origin-diagnostic. However, when a sample contains multiple “biotic” distributions of structural elements, molecules with nonrandom and repeated patterns in branching, both within individual molecules and across the molecules in a sample, this may indicate biogenicity, as revealed by 14-parameter PCA (Figs. 12, S18).

4.3.2.2. *Always present and sometimes origin-diagnostic.* Acyclic hydrocarbon distributions that are always present and sometimes origin-diagnostic include

- (1) Minimum and maximum chain length (Fig. 7);
- (2) Preference for either even or odd chain lengths (Fig. S9b);
- (3) Chain length distribution (*e.g.*, bimodal or trimodal) (Fig. S9a);
- (4) Most abundant molecule (Fig. 8a);
- (5) Second most abundant molecule (Fig. 8b).

Within our dataset, acyclic hydrocarbons with chain length ranges that fall between C₁₆ and C₄₈ are only observed in terrestrial samples. Ranges between C₁ and C₁₂ are only reported in meteorite samples. These distributions differ for biotic and abiotic samples. However, some terrestrial samples contain acyclic hydrocarbons as short as C₄, and some meteoritic samples contain chain lengths up to C₃₀. These intermediate ranges are non-diagnostic, as they are observed in samples of either origin.

In some terrestrial samples and no meteorite samples, the most abundant molecules are isoprenoids, mono- or polyunsaturated alkenes, or saturated *n*-alkanes with chains that are C₂₇ or longer. On the other hand, in some meteoritic samples (and no terrestrial samples), the most abundant molecules are saturated *n*-alkanes with chains between C₁ and C₅. However, there are both terrestrial and meteoritic samples for which the most abundant molecule is an *n*-alkane between C₁₁ and C₂₆. When the most abundant acyclic hydrocarbon is long-chained (>C₂₇), saturated, or an isoprenoid, this indicates biogenicity; when the most abundant molecule is short-chained (≤C₆), this most probably indicates a meteoritic origin; other cases are non-diagnostic. The same differences are observed for the second most abundant molecule within the same samples, except for one meteoritic sample in which the monounsaturated C_{2:1} alkene is the second most abundant (Levy *et al.*, 1973).

4.3.2.3. Sometimes present and always origin-diagnostic. Acyclic hydrocarbon distributions that are sometimes present and always origin-diagnostic include

- (1) Presence of isoprenoids (Figs. S13–S15);
- (2) Homologous series of *iso* or *anteiso* branched alkanes (Figs. S16c–S16d).

Isoprenoids are diagnostic branched molecules that are universal to life on Earth: they are metabolites for all life as we know it, are used to build complex polycyclic compounds in prokaryotes and eukaryotes (Zeng and Dehesh, 2021), and are acyclic building block components of archaeal cell membranes (Jordan *et al.*, 2019). In our dataset, acyclic isoprenoids are reported in many terrestrial samples but never reported as indigenous in meteorites. The presence of isoprenoids (or similar branched structures) indicates biogenicity; however, the absence of isoprenoids in the acyclic hydrocarbons of a sample is non-diagnostic, since some biotic samples lack isoprenoids.

Homologous series of *iso* or *anteiso* branched alkanes are observed in some terrestrial samples but not in meteoritic samples (Figs. S16c–S16d). These series are thought to derive from partially degraded fatty acids (*e.g.*, Peters and Moldowan, 1993); when they appear, this points to biotic input, but lack of these series is non-diagnostic.

4.3.2.4. Sometimes present and sometimes origin-diagnostic. Acyclic hydrocarbon distributions that are sometimes present and sometimes origin-diagnostic include

- (1) Number of unsaturations in a single molecule (Fig. S10);
- (2) Maximum number of branches in a single molecule (Fig. S12a);

- (3) Maximum number of carbons in a single branch (Fig. S12b);
- (4) Position of branching within the chain (Figs. 9b, 10).

Unsaturation are biologically important molecular features that allow organisms to regulate membrane fluidity when incorporated into fatty acid tails or GDGT-bound isoprenoids (Hazel and Eugene Williams, 1990; Mansy, 2009). Mono- and polyunsaturated acyclic hydrocarbons are synthesized by a variety of organisms and frequently identified in the terrestrial studies we reviewed; monounsaturated alkenes are reported in 9 of the meteoritic samples, although 7 of them were sourced from IOM rather than the organic-soluble fraction. It is well known that during diagenesis on Earth, double bonds are more susceptible to cleavage and oxidation compared to C–C bonds; older and/or thermally processed samples tend to contain fewer unsaturated species compared to fresher material (Peters and Moldowan, 1993; Canuel and Martens, 1996; Colombo *et al.*, 1997; Eigenbrode, 2008). Within our dataset, we found that when a sample contains more than one unsaturated acyclic hydrocarbon and/or any polyunsaturated molecules, biogenicity may be inferred, but when a sample contains only saturated alkanes or a single monounsaturated alkene, these distributions are non-diagnostic.

Branching distributions can indicate biogenicity in some cases. Due to the presence of isoprenoids, many biotic samples in our dataset include acyclic hydrocarbons that contain between 3 and 8 individual branches within a single molecule. Highly branched isoprenoids can contain individual branches that are more than 3 carbon atoms long (Summons *et al.*, 2022). UCM containing many highly branched species is common in meteorites (Fig. S16a), but individual molecules are rarely resolved from one another with the analytical techniques reported here. However, several of the studies we reviewed did identify individual branched hydrocarbons extracted from meteorites and contain up to 3 branches within a single chain, with branches up to 3 carbon atoms long. Acyclic hydrocarbon samples that contain individual molecules with more than 3 individual branches or branches that contain more than 3 carbon atoms can indicate biogenicity, while samples with fewer or no branches are non-diagnostic.

Finally, branching positions can sometimes indicate biogenicity. Only terrestrial samples exhibit branched acyclic hydrocarbons with the first branch position at the 3rd carbon atom or farther along the main chain of the molecule, and only terrestrial samples exhibit the last branch position as falling between the 10th and 32nd carbon atom. Both terrestrial and meteoritic samples can contain branched acyclic hydrocarbons with the first branch position located at the 2nd carbon of the main chain of the molecule, and both have been reported to have the last branch position located at the 8th carbon atom or earlier along the main chain. This indicates that when branching positions are restricted from the mid-chain to the terminal end, the sample may be of biotic origin, while earlier branching positions are non-diagnostic.

4.4. Additive origin-diagnosticity of molecular features and distributions

4.4.1. Additive trends are revealed by PCA of fatty acid distributions. Results from our 16-parameter PCA have

revealed that each independent feature is additive in assessing biotic versus abiotic origin for the lipids within our dataset (Figs. 11, S17; Table S1). While independent analyses of each individual feature can, in some cases, provide information on sample origin (*e.g.*, the identity of the most abundant fatty acid in a sample always differs for biotic and abiotic samples, while certain branch lengths and positions are only found in samples of either biotic or abiotic origin), multivariate approaches incorporate multiple features from the same samples into one analysis. The results of our PCA show that confidence in assessing origin is strengthened with this method, demonstrated by the clear separation of biotic and abiotic samples into distinct clusters (Fig. 11, Table S1). PCA additionally identified the parameters responsible for the greatest degree of diagnosticity for determining biogenicity (Fig. S18), which include minimum and maximum chain length (Fig. 2), most abundant unique fatty acid (Fig. 3a), and second most abundant unique fatty acid (Fig. 3b).

4.4.2. Additive trends are revealed by PCA of acyclic hydrocarbon distributions. Results from our 14-parameter PCA reveal that each independent feature is additive in assessing biotic versus abiotic origin for the lipids within our dataset (Table S2). Although lack of coverage within the meteoritic dataset precluded PCA with as many features as in our fatty acid analysis, and although there is crossover between the biotic and abiotic pools of acyclic hydrocarbons within this PCA, results indicate that multivariate analysis can help improve confidence in origin, even when independent analysis of individual features (*e.g.*, identity of the most abundant acyclic hydrocarbon, presence and number of unsaturations) does not definitively indicate whether a sample is biotic or abiotic in origin. Specifically, of the three distinct clusters resulting from the 14-parameter analysis, one cluster contained no abiotic samples, one contained a single abiotic outlier, and one contained both biotic and abiotic samples (Fig. 12). Additionally, these findings suggest that further analysis of meteoritic acyclic hydrocarbon structures and distributions has the potential to greatly enhance understanding of the differences between biotic and abiotic lipids and would be an important step to support astrobiological exploration of organics on other bodies.

4.5. Diagenesis and degradation: Mixed signals of lipids and expectations for Mars

An extraterrestrial sample may contain a mixture of lipids from abiotic and biotic sources, referred to as a “mixed signal,” which might include exogenous input, abiotic molecules synthesized *in situ*, biomolecules from extant or ancient life, and geologically re-processed or partially degraded compounds of biotic or abiotic origin. On Earth, fatty acids up to hundreds of millions of years old have been reported (Das and Harris, 1970), while preservation potential for acyclic hydrocarbons exceeds several billion years (Brocks *et al.*, 2005; Brocks and Schaeffer, 2008; Lee and Brocks, 2011; Vinnichenko *et al.*, 2020). Acyclic hydrocarbons are directly synthesized by both biotic and abiotic mechanisms, but these molecules can also represent partially degraded fatty acids that have undergone decarboxylation via geologi-

cal processing (*e.g.*, thermal and aqueous processes, irradiation, exposure to oxidizing agents) or biological re-uptake. The resulting hydrocarbon compounds that remain can retain some or all the structural elements (chain length, branching, unsaturations) possessed by the parent molecules (Peters and Moldowan, 1993; Summons *et al.*, 2008).

When distributions of these features in an acyclic hydrocarbon sample traces back to “biotic” fatty acid distributions, biogenicity may also be inferred, while comparison to carbonaceous chondrite fatty acids can inform predictions for distributions of abiotic acyclic hydrocarbons that are delivered exogenously to the surface of Mars (Freissinet *et al.*, 2015; Eigenbrode *et al.*, 2018). Biotic hydrocarbons in terrestrial systems can also contain UCMs, especially upon burial, heating, and metamorphism, but certain types of molecules and configurations frequently rise well above this background, especially when samples have only undergone low-grade metamorphism (*i.e.*, $<320^{\circ}\text{C}$) throughout their diagenetic history (*e.g.*, Peters and Moldowan, 1993). Although homologous series are only observed in some samples, their presence indicates biotic input. Monomethyl and dimethyl alkanes are directly synthesized by bacteria (*e.g.*, Robinson and Eglinton, 1990; Shiea *et al.*, 1990; Kenig *et al.*, 2003), while series of *iso*- and *anteiso*-alkanes are thought to derive from decarboxylated *iso*- and *anteiso*-fatty acids (Peters and Moldowan, 1993). Within the dataset we have compiled, these distributions are never observed in meteorites, and given the random isomerization that characterizes abiotically synthesized organics, homologous series of long-chained acyclic hydrocarbons with few discrete numbers of branches are not expected to occur outside of biologically directed synthesis.

For abiotic scenarios, observations and laboratory experiments provide information about the origin of the chemical composition of small, cold objects in the Solar System such as asteroids and comets, in particular the origin of their content of organic compounds. Through meteoritic impacts during the early stages of the Solar System formation, these organics are believed to have seeded with organics the surface of the telluric planets, including Earth and Mars, and may have triggered the emergence of life on Earth (Chyba and Sagan, 1992). The distributions of sample features identified in biotic samples can, therefore, serve as an inflection point for the types of nonrandom distributions that occur in terrestrial biology but are thermodynamically unlikely to occur in known abiotic systems.

Despite the lack of significant quantitative data on distributions of acyclic hydrocarbons in meteorites, data on the distributions of meteoritic and terrestrial fatty acids can provide insight into expected hydrocarbon distributions on Mars. Because asteroids and meteoroids are less geologically active than planets (aside from transient aqueous episodes), organics in meteorites are exceptionally well preserved (*e.g.*, Sephton, 2002; Pizzarello, 2006). However, when those objects fall onto a more dynamic planetary body, physical and chemical weathering processes can alter those molecules. Since short-chain ($\leq\text{C}_{12}$) fatty acids are the most abundant class of meteoritic organics, these molecules could potentially be an important source of acyclic hydrocarbons on Mars. If these molecules undergo decarboxylation after reaching the planet’s surface, as is believed to

have happened during the early diagenesis on Earth, the resulting acyclic hydrocarbon compounds that remain are likely to retain recalcitrant structural features (*e.g.*, C₁–C_{12:0} chain lengths, frequent branching, random isomerization). Therefore, these products could potentially represent an important source of acyclic hydrocarbons on Mars, and sets of short-chain (\leq C₁₂) acyclic hydrocarbons may indicate an abiotic origin from decarboxylated meteoritic fatty acids.

Given the much lower abundances of indigenous acyclic hydrocarbons, expected exogenous contributions could include a second set of longer-chain molecules containing *n*-alkanes with random chain length distributions, abundant UCMs, resolvable branched species with diverse and complex branching, random isomerization, abundances of polycyclic aromatic hydrocarbons (PAHs), rare unsaturations, and no molecules with nonrandom repeated branching (*i.e.*, isoprenoids or structurally similar compounds), as these general distributions are often cited for meteorites (Sephton, 2002; Pizzarello, 2006). On the other hand, if acyclic hydrocarbon patterns trace back to those identified, observed, or expected in hydrocarbon tails of biotically synthesized fatty acids, these distributions could provide evidence for biogenicity, especially if multiple diagnostic distributions are independently observed within one Martian lipid sample. Additional parameters, such as compound-specific stable isotope analyses, can additionally be applied to the same samples to further query origin; however, review of these techniques and distributions is outside the scope of this work (*e.g.*, Finkel *et al.*, 2023).

4.6. A lipid biomarker approach to life detection

Given the biological imperativeness of cellular membranes in aqueous environments, numerous origin-diagnostic distributions that can indicate biosynthesis, and the geological longevity of lipids postmortem for billions of years, acyclic lipids are ideal targets for organics-based life detection on astrobiological planetary bodies (Lovelock, 1965; Dorn *et al.*, 2011; Georgiou and Deamer, 2014). Martian surface conditions during early geologic epochs mimicked and coincided with those on Earth that facilitated the emergence of terrestrial life (Craddock and Howard, 2002; Grotzinger *et al.*, 2014), while Europa and Enceladus contain potentially habitable subsurface oceans that may host life today (Chyba, 2000; Hand *et al.*, 2009; Deamer and Damer, 2017).

In the absence of modern or ancient life, lipid analysis can provide information about the inventory of prebiotic organics and show how they compare or differ with exogenous lipids from primitive extraplanetary materials (*e.g.*, *in situ* synthesis) (Pizzarello and Shock, 2010; Steele *et al.*, 2018; Schmitt-Kopplin *et al.*, 2023). Since the Martian surface has consistently been seeded with exogenous organics throughout the history of the Solar System, identifying Martian lipids with distributions that deviate from the meteoritic abiotic background could provide clues to ancient life. Meteoritic lipids could represent a potentially important source of prebiotic organics to the early Earth and Mars.

Distinct pools of endogenous abiotic organics have been identified on Mars. Indeed, recent analysis of the Tissint Martian meteorite has revealed a wide diversity of abiotic organics, including C₃–C₇ branched monocarboxylic acids, polycyclic aromatic hydrocarbons (PAHs) and heterocycles,

aldehydes, and unsaturated acyclic hydrocarbons, with a signature unique from carbonaceous meteorites. Their compositional distributions (*i.e.*, elementary composition identified by Fourier transform ion cyclotron resonance mass spectrometry [FTICR-MS]) and spatial distributions indicate these molecules formed via *in situ* abiotic geochemical processes (Schmitt-Kopplin *et al.*, 2023). Other analyses of Martian meteorites have similarly identified complex organic matter with mineral associations, suggesting endogenous abiotic synthesis (*e.g.*, serpentinization, carbonation, shocks) (Steele *et al.*, 2016, 2022; Jaramillo *et al.*, 2019). On Earth, it has been hypothesized that abiotically synthesized carboxylic acids self-assembled to form the vesicles that served as the first primitive cell membranes, facilitating the transition from prebiotic to biotic chemistry (Segré *et al.*, 2001; Apel *et al.*, 2002; Deamer *et al.*, 2002). If extraterrestrial life on Mars followed a similar trajectory, synthesis of larger fatty acids and other complex lipids would follow, yielding molecules with structures and distributions distinct from abiotic precursors (Segré *et al.*, 2001; Deamer *et al.*, 2002).

4.6.1. Instrumentation to detect origin-diagnostic acyclic lipid distributions *in situ*. To detect and characterize biomarkers *in situ*, methods, instruments, and analytical techniques developed and used in terrestrial laboratories are frequently leveraged for robotic exploration of Mars (Mahaffy *et al.*, 2012; Vago *et al.*, 2017; Farley *et al.*, 2020). Knowledge gained from organic biogeochemical analyses of analog environments on Earth is another essential tool in the search for life, as an understanding of trends in biomarker distributions is key to designing and interpreting experiments for the detection of organics (*e.g.*, Finkel *et al.*, 2023). *In situ* measurements taken by the Mars Science Laboratory on board the Curiosity rover and the Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals (SHERLOC) instrument on board the Perseverance rover have detected numerous organics on Mars, including small, simple hydrocarbons, chlorinated organic fragments, and S- and N-bearing heterocycles, where these specific molecules have been identified by MSL (Ming *et al.*, 2014; Freissinet *et al.*, 2015; Eigenbrode *et al.*, 2018; Scheller *et al.*, 2022). Analyses of Martian meteorites have found a staggering diversity of indigenous abiotic organics with complex and diverse structures (*e.g.*, carboxylic acids, aromatic and aliphatic hydrocarbons, heterocycles) bound in macromolecules (Lin *et al.*, 2014; Steele *et al.*, 2016, 2022; Jaramillo *et al.*, 2019; Schmitt-Kopplin *et al.*, 2023).

Optimal sampling and processing approaches should favor the collection of these molecules while conserving structural features and distributions in molecules outlined here that can indicate biogenicity or its absence. Previous *in situ* investigations of Martian organics using thermal extraction only found evidence for the presence of simple, non-origin-diagnostic hydrocarbons (Mahaffy *et al.*, 2012; Ming *et al.*, 2014; Freissinet *et al.*, 2015; Eigenbrode *et al.*, 2018). While significantly operationally less complex than solvent extraction, thermal extraction is known to alter the origin-diagnostic structures and features of molecules (*e.g.*, oxidation, chlorination, racemization, fragmentation), particularly in the presence of oxidants such as oxychlorine anions including perchlorates which are prevalent on Mars

(Kates, 1972; Sephton, 2012; Royle *et al.*, 2022). In contrast, the use of solvent-based extraction techniques most widely used on terrestrial samples maintains the origin-diagnostic structural features identified here (Fig. 13; Tables S2, S3), especially more thermally unstable features such as double bonds (*e.g.*, Sephton, 2012). Techniques widely demonstrated on tens of thousands of terrestrial and meteorite sample analysis rely on human operators and are multistep, laborious, and require consumables (*i.e.*, organic solvents, filters), which is why their use has been historically precluded for *in situ* analysis on planetary missions.

These laboratory lipid characterization techniques that have been successfully implemented and refined over 70 years primarily use organic solvents at temperatures below 100°C to extract lipids from samples, often via mechanically refluxing organic solvent through ground sample within a closed vessel (Fig. 13) (Bligh and Dyer, 1959; Richter *et al.*, 1996; Luque de Castro and Priego-Capote, 2010). Our analysis of the literature has revealed the ubiquity of solvent-based approaches for extracting lipids from natural samples in the laboratory, and the capability of these methods to conserve key origin-diagnostic structures and distributions throughout the sample-handling process. Frequent successful utilization of Soxhlet and ASE apparatuses demonstrates the utility of automating portions of lipid extraction procedures. Additionally, our review highlights the importance of coupling organic solvent extraction to mass spectrometry-based analyses to achieve a thorough understanding of the molecular details of a given sample, including individual structural features and conformations, and overall distributions.

These techniques would be suitable for Martian samples, given their capacity to extract and concentrate low concentrations of biomass that may be heterogeneously distributed on the centimeter scale in a soil sample: multigram samples of regolith can be processed so that lipids are extracted, purified, and concentrated, improving the analytical instrumental signal by several orders of magnitude. Additionally, solvent extraction, especially with sonication, can disrupt molecular interactions between organics and the host mineral matrix, liberating and separating lipids from the associated minerals (Keil and Mayer, 2014), without significantly altering the diagnostic molecular structures and patterns reviewed here.

4.6.2. The Extractor for Chemical Analysis of Lipid Biomarkers in Regolith (ExCALiBR). Ideally, a life-detection instrument designed to measure acyclic lipids should detect and resolve molecules in the same class but with different chain lengths, number and position of branches, and double bonds. This should include C₂ chains through longer chain lengths (*e.g.*, C₃₅) for a complete window spanning known terrestrial and exogenous inputs. Leveraging the sample processing and analytical techniques used to identify lipids in the laboratory, while searching within the same molecular windows for the types of structures and distributions that set lipid biomarkers on Earth apart from abiotically synthesized exogenous organics, provides guidance for interpreting the same molecular classes from a sample of unknown origin. Because organic solvent extraction is most commonly utilized in sample analysis on Earth (Figs. 13a, 13b), and demonstrably preserves the molecular structures, features, and distributions that can help differentiate between biotic and

abiotic origin, we propose that these methodologies should be integrated into the next generation of sample processing instruments for life detection on astrobiology missions to Mars. To enable autonomous laboratory-grade lipid extraction and analysis within planetary mission mass, volume, and power constraints, we are developing an instrument optimized to extract lipids from planetary samples, utilizing low-temperature extraction techniques that preserve the molecular features and structural nuances that can provide key origin-diagnostic information.

Our instrument, ExCALiBR (Extractor for Chemical Analysis of Lipid Biomarkers in Regolith), implements common organic solvent-based laboratory sample processing techniques within a single, enclosed, autonomously operated system that extracts, concentrates, and delivers lipids from 25 cm³ of raw drilled or scooped samples, while fitting within the mass, volume, and power constraints imposed by most landed planetary mission budgets (Wilhelm *et al.*, 2021). By processing multiple 25 cm³ samples, the quantity of a given lipid available for analysis is enhanced by greater than an order of magnitude relative to previous measurements on Mars. ExCALiBR has been included as an option for the baseline payload package for the Mars Life Explorer mission concept in the 2023 Planetary Decadal Survey and prioritized by the National Academies for maturation to flight scale within the coming decade (National Academies of Sciences, Engineering, and Medicine, 2022). ExCALiBR will enable the search for ancient, trace quantities of preserved geolipids on Mars, with fidelity to standard laboratory techniques to search for the origin-diagnostic parameters outlined here. ExCALiBR specifically targets organic classes with highest preservation potential (*i.e.*, lipids) while preserving molecular structures and features that can indicate origin (*e.g.*, chain length, unsaturations, branching). ExCALiBR further overcomes common sample handling challenges with fidelity to the techniques most commonly employed on terrestrial samples (Figs. 13a, 13b) by reducing particle size and enable access to organics trapped within mineral matrixes, extracting organics with organic solvents to remove salts and inorganics that are known to interfere with analysis (*e.g.*, perchlorates can destroy organics during thermal ramps [*e.g.*, Sephton, 2012; Royle *et al.*, 2022]), applying ultrasonic energy to disaggregate particles and release molecules adsorbed onto mineral surfaces or in interlayers, filtering to separate minerals from organics, and concentrating from large sample volumes to improve signal from low-abundance or heterogeneously distributed natural samples. The design of our instrument was driven by the results of this comparative study. Materials were selected for compatibility with a range of organic solvents, and solvent-based extraction techniques and cocktails chosen based on methodologies most commonly reported in the studies reviewed (*i.e.*, solvent extraction is reported for 83% of the 1574 samples included in our study) (Figs. 13a, 13b). ExCALiBR extraction techniques are further refined to target lipids within the chain length ranges delineated (Figs. 2, 7), with the aim of preserving for analysis the molecular features (*i.e.*, chain length [Figs. 2, 7], unsaturations [Figs. 4, S7, S10, S11], branching [Figs. 5, 6a, 6b, 9a, 9b, 10, S8a, S8b, S12a, S12b, S13, S14, S15]) that provide origin-diagnostic information. The ExCALiBR sample processing instrument is capable of

coupling to a variety of novel and flight heritage analytical instruments (*e.g.*, GC-MS, Raman spectrometers, LDI-MS), delivering purified and concentrated lipid aliquots for detailed molecular characterization on flexible payload platforms and packages. ExCALiBR will advance the search for life on Mars by implementing the highest Earth-heritage lipid extraction techniques (*i.e.*, organic solvent-based) into a single sample processing unit, automating these methods for robotic implementation on Mars.

5. Conclusion

This study examines life detection as a binary question: biotic or abiotic? By cataloging published results, we better understand the bounds on the two endmember organic pools: (i) carbonaceous chondrite organics, presumably responsible for providing seed materials for life on Earth, and (ii) modern terrestrial biology that is the product of billions of years of evolution and subsequent geologic preservation. In approaching the search for life on Mars, understanding the molecular structures and distributions of these two pools of organics can frame the bounds of what we will seek and, at the very least, give us the range of solubility, molecular weight, and structural features to which a payload instrument suite should be sensitive and selective.

Previous papers have defined life detection as a departure from distributions observed in natural samples of exogenous origin (*e.g.*, meteorites) or observed in laboratory synthesis experiments. However, the reality of the problem is likely to be more complex, therefore, our approach to life detection must account for organic geochemical realities. Diagenetic processes alter both biotically and abiotically synthesized organics (Benner *et al.*, 2000; Pavlov *et al.*, 2012; Fornaro *et al.*, 2018; Fox *et al.*, 2019; Roussel *et al.*, 2022); the physicochemical conditions imposed on organics in the Martian environment such as radiation, soil oxidants, wet/dry cycling, and temperature cycling are agents that may further obscure diagnostic parameters relative to freshly synthesized organics, and *in situ* abiotic synthesis is hypothesized to impart an additional pool of organics that may display unique distributions (Steele *et al.*, 2016, 2018, 2022; Schmitt-Kopplin *et al.*, 2023). On other planets, we should not expect organics to be pristinely preserved as the abiotic compounds delivered by meteorites. Leveraging understanding of pools of organic building blocks of life on Earth can help bound the types of organics and molecular structures that will provide clues to their origins on other planetary bodies and set requirements for future life-detection payload suites.

Organics-assessment-based life-detection instruments should be sensitive to both pools, meteoritic and biogenic, as well as their mixture. An ideal approach to life detection would enable us to recognize geological changes in organic content. This includes a purely meteoritic source largely unaltered abiotic organics, followed by altered abiotic organics, followed by freshly synthesized biological material, followed by preserved ancient biomarkers heavily altered by geological processes, while keeping in mind the possibility that molecular signatures of any combination of these four stages of alteration may still be present in a sample. Thus, we should not look at the problem from the angle of a biotic/abiotic binary but rather understand how nonrandom and thermodynamically unfavorable, kinetically favored patterns in distributions of preserved,

biosynthesized lipids on Earth can inform us about how to search for physical evidence of life in extraterrestrial pools of organics. A simple deviation from observed or expected abiotic distributions is not sufficient; rather, a combination of multiple lines of evidence (*i.e.*, multiple patterns in distributions of structures and conformations) observed from within those molecules will help increase confidence in the results.

Each of the features and patterns we identified in fatty acids and acyclic hydrocarbons can hold origin-diagnostic information. While each structural nuance could independently indicate the original synthesis processes (*i.e.*, biochemical synthesis vs. abiotic synthesis), analyzing multiple such indicators in concert to assess the absence, or presence and degree, of biogenicity of a single lipid sample will diminish ambiguity and enhance the reliability of the answer of life on another body.

Acknowledgments

This work was supported by a 2019 NASA STMD Early Career Initiative award to Dr. Wilhelm. We thank Jennifer Eigenbrode, Carina Lee, George Cooper, and Andro Rios for helpful discussions.

Dedication

This paper is dedicated to Dr. James Lovelock (1919–2022), whose philosophy on biomarker detection guided NASA's first plans for the search for life on Mars with Viking. His contribution to the field of astrobiology continues to inspire and guide our thinking.

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Information
Supplementary Table S1
Supplementary Table S2
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Figure S7
Supplementary Figure S8
Supplementary Figure S9
Supplementary Figure S10
Supplementary Figure S11
Supplementary Figure S12
Supplementary Figure S13
Supplementary Figure S14
Supplementary Figure S15
Supplementary Figure S16
Supplementary Figure S17
Supplementary Figure S18

References

Aerts JW, Röling WFM, Elsaesser A, *et al.* Biota and biomolecules in extreme environments on Earth: Implications for

- life detection on Mars. *Life* 2014;4(4):535–565; doi: 10.3390/life4040535.
- Alexander CMOD, Fogel M, Yabuta H, *et al.* The origin and evolution of chondrites recorded in the elemental and isotopic compositions of their macromolecular organic matter. *Geochim Cosmochim Acta* 2007;71(17):4380–4403; doi: 10.1016/j.gca.2007.06.052.
- Alexander CMOD, Cody GD, De Gregorio BT, *et al.* The nature, origin and modification of insoluble organic matter in chondrites, the major source of Earth's C and N. *Geochemistry* 2017;77(2):227–256; doi: 10.1016/j.chemer.2017.01.007.
- Apel CL, Deamer DW, Mautner MN. Self-assembled vesicles of monocarboxylic acids and alcohols: Conditions for stability and for the encapsulation of biopolymers. *Biochim Biophys Acta Biomembranes* 2002;1559(1):1–9; doi: 10.1016/S0005-2736(01)00400-X.
- Aponte JC, Tarozo R, Alexandre MR, *et al.* Chirality of meteoritic free and IOM-derived monocarboxylic acids and implications for prebiotic organic synthesis. *Geochim Cosmochim Acta* 2014;131:1–12; doi: 10.1016/j.gca.2014.01.035.
- Bao X, Katz S, Pollard M, *et al.* Carbocyclic fatty acids in plants: Biochemical and molecular genetic characterization of cyclopropane fatty acid synthesis of *Sterculia foetida*. *Proc Natl Acad Sci USA* 2002;99(10):7172–7177; doi: 10.1073/pnas.092152999.
- Benner SA, Devine KG, Matveeva LN, *et al.* The missing organic molecules on Mars. *Proc Natl Acad Sci USA* 2000; 97(6):2425–2430; doi: 10.1073/pnas.040539497.
- Bernstein MP, Sandford SA, Allamandola LJ, *et al.* Organic compounds produced by photolysis of realistic interstellar and cometary ice analogs containing methanol. *Astrophys J* 1995; doi: 10.1086/176485.
- Bernstein MP, Dworkin JP, Sandford SA, *et al.* Ultraviolet irradiation of naphthalene in H₂O ice: Implications for meteorites and biogenesis. *Meteorit Planet Sci* 2001;36(3):351–358; doi: 10.1111/j.1945-5100.2001.tb01878.x.
- Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 1959;37(8):911–917; doi: 10.1139/o59-099.
- Boucher Y, Kamekura M, Doolittle WF. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol* 2004;52(2):515–527; doi: 10.1111/j.1365-2958.2004.03992.x.
- Brindley DN, Matsumura S, Bloch K. *Mycobacterium phlei* fatty acid synthetase—a bacterial multienzyme complex. *Nature* 1969;224(5220):666–669; doi: 10.1038/224666a0.
- Brocks JJ, Schaeffer P. Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1640Ma Barney Creek Formation. *Geochim Cosmochim Acta* 2008;72(5):1396–1414; doi: 10.1016/j.gca.2007.12.006.
- Brocks JJ, Buick R, Summons RE, *et al.* A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim Cosmochim Acta* 2003;67(22):4321–4335; doi: 10.1016/S0016-7037(03)00209-6.
- Brocks JJ, Love GD, Summons RE, *et al.* Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature* 2005;437(7060):866–870; doi: 10.1038/nature04068.
- Canuel EA, Martens CS. Reactivity of recently deposited organic matter: Degradation of lipid compounds near the sediment-water interface. *Geochim Cosmochim Acta* 1996; 60(10):1793–1806; doi: 10.1016/0016-7037(96)00045-2.
- Chyba C, Sagan C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: An inventory for the origins of life. *Nature* 1992;355(6356):125–132; doi: 10.1038/355125a0.
- Chyba CF. Energy for microbial life on Europa. *Nature* 2000; 403(6768):381–382; doi: 10.1038/35000281.
- Colombo JC, Silverberg N, Gearing JN. Lipid biogeochemistry in the Laurentian Trough—II. Changes in composition of fatty acids, sterols and aliphatic hydrocarbons during early diagenesis. *Org Geochem* 1997;26(3):257–274; doi: 10.1016/S0146-6380(96)00146-5.
- Coskun Ü, Simons K. Cell membranes: The lipid perspective. *Structure* 2011;19(11):1543–1548; doi: 10.1016/j.str.2011.10.010.
- Craddock RA, Howard AD. The case for rainfall on a warm, wet early Mars. *J Geophys Res Planets* 2002;107(E11):21-1-21–36; doi: 10.1029/2001JE001505.
- Cronin JR, Pizzarello S. Aliphatic hydrocarbons of the Murchison Meteorite. *Geochim Cosmochim Acta* 1990; 54(10):2859–2868; doi: 10.1016/0016-7037(90)90020-L.
- Das SK, Harris RS. Lipids and fatty acids in fossil teeth. *J Dent Res* 1970;49(1):126–130; doi: 10.1177/00220345700490010401.
- Deamer D, Damer B. Can life begin on Enceladus? A perspective from hydrothermal chemistry. *Astrobiology* 2017; 17(9):834–839; doi: 10.1089/ast.2016.1610.
- Deamer D, Dworkin JP, Sandford SA, *et al.* The first cell membranes. *Astrobiology* 2002;2(4):371–381; doi: 10.1089/153110702762470482.
- Deamer DW. Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* 1985;317(6040):792–794; doi: 10.1038/317792a0.
- Deamer DW, Pashley RM. Amphiphilic components of the Murchison carbonaceous chondrite: Surface properties and membrane formation. *Orig Life Evol Biosphere* 1989;19(1): 21–38; doi: 10.1007/BF01808285.
- Dorn ED, Nealson KH, Adami C. Monomer abundance distribution patterns as a universal biosignature: Examples from terrestrial and digital life. *J Mol Evol* 2011;72(3):283–295; doi: 10.1007/s00239-011-9429-4.
- Dworkin JP, Deamer DW, Sandford SA, *et al.* Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices. *Proc Natl Acad Sci USA* 2001;98(3):815–819; doi: 10.1073/pnas.98.3.815.
- Eigenbrode JL. Fossil lipids for life-detection: A case study from the early Earth record. In *Strategies of Life Detection*. (Botta O, Bada JL, Gomez-Elvira J, *et al.* eds.) Space Sciences Series of ISSI Springer US: Boston, MA; 2008; pp 161–185; doi: 10.1007/978-0-387-77516-6_12.
- Eigenbrode JL, Summons RE, Steele A, *et al.* Organic matter preserved in 3-billion-year-old mudstones at Gale Crater, Mars. *Science* 2018;360:1096–1101; doi: 10.1126/science.aas9185.
- Farley KA, Williford KH, Stack KM, *et al.* Mars 2020 mission overview. *Space Sci Rev* 2020;216(8):142; doi: 10.1007/s11214-020-00762-y.
- Finkel PL, Carrizo D, Parro V, *et al.* An overview of lipid biomarkers in terrestrial extreme environments with relevance for Mars exploration. *Astrobiology* 2023;23(5):563–604; doi: 10.1089/ast.2022.0083.
- Fornaro T, Steele A, Brucato JR. Catalytic/Protective properties of martian minerals and implications for possible origin of life on Mars. *Life* 2018;8(4):56; doi: 10.3390/life8040056.

- Fox AC, Eigenbrode JL, Freeman KH. Radiolysis of macromolecular organic material in Mars-relevant mineral matrices. *J Geophys Res Planets* 2019;124:3257–3266; doi: 10.1029/2019JE006072.
- Freissinet C, Glavin DP, Mahaffy PR, *et al.* Organic molecules in the Sheepbed Mudstone, Gale Crater, Mars. *J Geophys Res Planets* 2015;120(3):495–514; doi: 10.1002/2014JE004737.
- Garcette-Lepecq A, Largeau C, Bouloubassi I, *et al.* Lipids and their modes of occurrence in two surface sediments from the Danube Delta and Northwestern Black Sea: Implications for sources and early diagenetic alteration. *Org Geochem* 2004; 35(8):959–980; doi: 10.1016/j.orggeochem.2004.03.005.
- Georgiou CD, Deamer DW. Lipids as universal biomarkers of extraterrestrial life. *Astrobiology* 2014;14(6):541–549; doi: 10.1089/ast.2013.1134.
- Gilmour I. Structural and isotopic analysis of organic matter in carbonaceous chondrites. *Treatise on Geochemistry* 2003;1: 711; doi: 10.1016/B0-08-043751-6/01146-4.
- Gower JC. A general coefficient of similarity and some of its properties. *Biometrics* 1971;27(4):857–871; doi: 10.2307/2528823.
- Grogan DW, Cronan JE. Cyclopropane ring formation in membrane lipids of bacteria. *Microbiol Mol Biol Rev* 1997; 61(4):429–441; doi: 10.1128/mmb.61.4.429-441.1997.
- Grotzinger JP, Sumner DY, Kah LC, *et al.* A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars. *Science* 2014;343:1242777; doi: 10.1126/science.1242777.
- Hagve T-A. Effects of unsaturated fatty acids on cell membrane functions. *Scandinavian Journal of Clinical and Laboratory Investigation* 1988;48(5):381–388; doi: 10.1080/00365518809085746.
- Hand KP, Chyba CF, Priscu JC, *et al.* *Astrobiology and the Potential for Life on Europa*. NASA Jet Propulsion Laboratory/California Institute of Technology: Pasadena, CA; 2009.
- Hazel JR, Eugene Williams E. The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. *Progress in Lipid Research* 1990;29(3):167–227; doi: 10.1016/0163-7827(90)90002-3.
- Herd CD, Blinova A, Simkus DN, *et al.* Origin and evolution of prebiotic organic matter as inferred from the Tagish Lake meteorite. *Science* 2011;332(6035):1304–1307; doi: 10.1126/science.1203290.
- Hilts RW, Herd CD, Simkus DN, *et al.* Soluble organic compounds in the Tagish Lake meteorite. *Meteorit Planet Sci* 2014;49(4):526–549; doi: 10.1111/maps.12272.
- IUPAC. *Compendium of Chemical Terminology*. (the “Gold Book”); 2nd ed. Compiled by McNaught AD, Wilkinson A.). Blackwell Scientific Publications: Oxford, n.d.; doi: 10.1351/goldbook.
- Jaramillo E, Royle S, Claire M, *et al.* Indigenous organic-oxidized fluid interactions in the Tissint Mars meteorite. *Geophys Res Lett* 2019;46; doi: 10.1029/2018GL081335.
- Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philos Trans A Math Phys Eng Sci* 2016;374(2065):20150202; doi: 10.1098/rsta.2015.0202.
- Jordan SF, Nee E, Lane N. Isoprenoids enhance the stability of fatty acid membranes at the emergence of life potentially leading to an early lipid divide. *Interface Focus* 2019;9(6): 20190067; doi: 10.1098/rsfs.2019.0067.
- Kaneda T. Iso- and anteiso-fatty acids in bacteria: Biosynthesis, function, and taxonomic significance. *Microbiol Rev* 1991; 55(2):288–302; doi: 10.1128/mr.55.2.288-302.1991.
- Kates M. *Techniques of Lipidology; Isolation, Analysis and Identification of Lipids*. Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier: Amsterdam, 1972;3:347–353.
- Keil RG, Mayer LM. Mineral matrices and organic matter. In *Treatise on Geochemistry*. (2nd ed. Holland HD, Turekian KK. eds.) Elsevier: Amsterdam, 2014; pp 337–359; doi: 10.1016/B978-0-08-095975-7.01024-X.
- Kenig F, Simons D-JH, Crich D, *et al.* Branched aliphatic alkanes with quaternary substituted carbon atoms in modern and ancient geologic samples. *Proc Natl Acad Sci USA* 2003; 100(22):12554–12558; doi: 10.1073/pnas.1735581100.
- Keris-Sen UD, Sen U, Soydemir G, *et al.* An investigation of ultrasound effect on microalgal cell integrity and lipid extraction efficiency. *Bioresource Technology* 2014;152:407–413; doi: 10.1016/j.biortech.2013.11.018.
- Koga Y. Thermal adaptation of the archaeal and bacterial lipid membranes. *Archaea* 2012;2012:e789652; doi: 10.1155/2012/789652.
- Komiya M, Shimoyama A. Organic compounds from insoluble organic matter isolated from the Murchison carbonaceous chondrite by heating experiments. *Bull Chem Soc Jpn* 1996; 69(1):53–58; doi: 10.1246/bcsj.69.53.
- Kong W, Liu N, Zhang J, *et al.* Optimization of ultrasound-assisted extraction parameters of chlorophyll from *Chlorella vulgaris* residue after lipid separation using response surface methodology. *J Food Sci Technol* 2014;51(9):2006–2013; doi: 10.1007/s13197-012-0706-z.
- Ladygina N, Dedyukhina EG, Vainshtein MB. A review on microbial synthesis of hydrocarbons. *Process Biochemistry* 2006;41(5):1001–1014; doi: 10.1016/j.procbio.2005.12.007.
- Lee C, Brocks JJ. Identification of carotane breakdown products in the 1.64 billion year old Barney Creek Formation, McArthur Basin, Northern Australia. *Org Geochem* 2011; 42(4):425–430; doi: 10.1016/j.orggeochem.2011.02.006.
- Lee J-Y, Yoo C, Jun S-Y, *et al.* Comparison of several methods for effective lipid extraction from microalgae. *Bioresource Technology* 2010;101(1, Supplement):S75–S77; doi: 10.1016/j.biortech.2009.03.058.
- Levy RL, Grayson MA, Wolf CJ. The organic analysis of the Murchison meteorite. *Geochim Cosmochim Acta* 1973;37(3): 467–483; doi: 10.1016/0016-7037(73)90212-3.
- Lin Y, El Goresy A, Hu S, *et al.* NanoSIMS analysis of organic carbon from the Tissint martian meteorite: Evidence for the past existence of subsurface organic-bearing fluids on Mars. *Meteorit Planet Sci* 2014;49(12):2201–2218; doi: 10.1111/maps.12389.
- Lloyd S. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 1982;28(2):129–137; doi: 10.1109/TIT.1982.1056489.
- Lovelock JE. A physical basis for life detection experiments. *Nature* 1965;207(997):568–570; doi: 10.1038/207568a0.
- Luque de Castro MD, Priego-Capote F. Soxhlet extraction: Past and present panacea. *Journal of Chromatography A* 2010; 1217(16):2383–2389; doi: 10.1016/j.chroma.2009.11.027.
- Mahaffy P, Webster C, Cabane M, *et al.* The Sample Analysis at Mars investigation and instrument suite. *Space Sci Rev* 2012;170:1604–1604; doi: 10.1007/s11214-012-9879-z.
- Malherbe C, Hutchinson IB, Ingley R, *et al.* On the habitability of desert varnish: A combined study by micro-Raman spectroscopy, X-ray diffraction, and methylated pyrolysis–gas chromatography–mass spectrometry. *Astrobiology* 2017;17: 1123–1137; doi: 10.1089/ast.2016.1512.
- Mansy SS. Model protocells from single-chain lipids. *Int J Mol Sci* 2009;10(3):835–843; doi: 10.3390/ijms10030835.

- Martins Z. Detection of organic matter and biosignatures in space missions. *Current Issues in Molecular Biology* 2020; 38(1):53–74; doi: 10.21775/cimb.038.053.
- McCarthy AD, Hardie DG. Fatty acid synthase—an example of protein evolution by gene fusion. *Trends Biochem Sci* 1984; 9(2):60–63; doi: 10.1016/0968-0004(84)90184-1.
- McCullom TM, Ritter G, Simoneit BRT. Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Orig Life Evol Biosph* 1999;29(2):153–166; doi: 10.1023/A:1006592502746.
- Meinert C, Myrgorodska I, de Marcellus P, et al. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* 2016;352:208–212; doi: 10.1126/science.aad8137.
- Meyers PA. Organic geochemical proxies of paleoceanographic, paleolimnologic, and paleoclimatic processes. *Org Geochem* 1997;27(5):213–250; doi: 10.1016/S0146-6380(97)00049-1.
- Ming DW, Archer PD, Glavin DP, et al. Volatile and organic compositions of sedimentary rocks in Yellowknife Bay, Gale Crater, Mars. *Science* 2014;343(6169); doi: 10.1126/science.1245267.
- Mißbach H, Schmidt BC, Duda J-P, et al. Assessing the diversity of lipids formed via Fischer-Tropsch-type reactions. *Org Geochem* 2018;119:110–121; doi: 10.1016/j.orggeochem.2018.02.012.
- National Academies of Sciences, Engineering, and Medicine. *Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023–2032*. National Academies Press: Washington, DC; 2022.
- Neveu M, Hays L, Voytek M, et al. The Ladder of Life Detection. *Astrobiology* 2018;18(11):1375–1402; doi: 10.1089/ast.2017.1773.
- Nuevo M, Auger G, Blanot D, et al. a detailed study of the amino acids produced from the vacuum UV irradiation of interstellar ice analogs. *Orig Life Evol Biosph* 2008;38(1):37–56; doi: 10.1007/s11084-007-9117-y.
- Nuevo M, Cooper G, Sandford SA. Deoxyribose and deoxysugar derivatives from photoprocessed astrophysical ice analogues and comparison to meteorites. *Nat Commun* 2018; 9:5276; doi: 10.1038/s41467-018-07693-x.
- Oba Y, Naraoka H. Carbon isotopic composition of acetic acid generated by hydrous pyrolysis of macromolecular organic matter from the Murchison meteorite. *Meteorit Planet Sci* 2006;41(8):1175–1181; doi: 10.1111/j.1945-5100.2006.tb00514.x.
- Okumura F, Mimura K. Gradual and stepwise pyrolyses of insoluble organic matter from the Murchison meteorite revealing chemical structure and isotopic distribution. *Geochim Cosmochim Acta* 2011;75(22):7063–7080; doi: 10.1016/j.gca.2011.09.015.
- Pavlov AA, Vasilyev G, Ostryakov VM, et al. Degradation of the organic molecules in the shallow subsurface of Mars due to irradiation by cosmic rays. *Geophys Res Lett* 2012;39: L13202; doi: 10.1029/2012GL052166.
- Peters KE, Moldowan JM. *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments*. Prentice Hall: Englewood Cliffs, NJ; 1993.
- Pizzarello S. The chemistry of life's origin: A carbonaceous meteorite perspective. *Acc Chem Res* 2006;39(4):231–237; doi: 10.1021/ar050049f.
- Pizzarello S, Shock E. The organic composition of carbonaceous meteorites: the evolutionary story ahead of biochemistry. *Cold Spring Harb Perspect Biol* 2010;2(3):a002105–a002105; doi: 10.1101/cshperspect.a002105.
- Pizzarello S, Shock E. Carbonaceous chondrite meteorites: The chronicle of a potential evolutionary path between stars and life. *Orig Life Evol Biosph* 2017;47(3):249–260; doi: 10.1007/s11084-016-9530-1.
- Pizzarello S, Schrader DL, Monroe AA, et al. Large enantiomeric excesses in primitive meteorites and the diverse effects of water in cosmochemical evolution. *Proc Natl Acad Sci USA* 2012;109:11949–11954; doi: 10.1073/pnas.1204865109.
- Ratnayake WMN, Galli C. Fat and fatty acid terminology, methods of analysis and fat digestion and metabolism: A background review paper. *Ann Nutr Metab* 2009;55(1–3): 8–43; doi: 10.1159/000228994.
- Remusat L. Organic material in meteorites and the link to the origin of life. *BIO Web of Conferences* 2014;2:03001; doi: 10.1051/bioconf/20140203001.
- Remusat L, Robert F, Derenne S. The insoluble organic matter in carbonaceous chondrites: Chemical structure, isotopic composition and origin. *Comptes Rendus Geoscience*, 2007; 339(14–15):895–906; doi: 10.1016/j.crte.2007.10.001.
- Richter B, Jones BA, Ezzell JL, et al. Accelerated solvent extraction: A technique for sample preparation. *Anal Chem* 1996;68(6):1033–1039; doi: 10.1021/ac9508199.
- Robinson N, Eglinton G. Lipid chemistry of Icelandic hot spring microbial mats. *Org Geochem* 1990;15(3):291–298; doi: 10.1016/0146-6380(90)90007-M.
- Roussel A, McAdam AC, Graham HV, et al. Diagnostic biosignature transformation under simulated martian radiation in organic-rich sedimentary rocks. *Frontiers in Astronomy and Space Sciences* 2022;9:919828; doi: 10.3389/fspas.2022.919828.
- Royle SH, Salter TL, Watson JS, et al. Mineral matrix effects on pyrolysis products of kerogens infer difficulties in determining biological provenance of macromolecular organic matter at Mars. *Astrobiology* 2022;22(5):520–540; doi: 10.1089/ast.2021.0074.
- Rushdi AI, Simoneit BRT. Lipid formation by aqueous Fischer-Tropsch-type synthesis over a temperature range of 100 to 400 °C. *Orig Life Evol Biosph* 2001;31(1):103–118; doi: 10.1023/A:1006702503954.
- Sandford SA, Nuevo M, Bera PP, et al. Prebiotic astrochemistry and the formation of molecules of astrobiological interest in interstellar clouds and protostellar disks. *Chem Rev* 2020; 120(11):4616–4659; doi: 10.1021/acs.chemrev.9b00560.
- Sapers HM, Razzell Hollis J, Bhartia R, et al. The cell and the sum of its parts: Patterns of complexity in biosignatures as revealed by deep UV Raman spectroscopy. *Front Microbiol* 2019;10:679; doi: 10.3389/fmicb.2019.00679.
- Scheller EL, Razzell Hollis J, Cardarelli EL, et al. Aqueous alteration processes in Jezero Crater, Mars—implications for organic geochemistry. *Science* 2022;378:1105–1110; doi: 10.1126/science.abo5204.
- Schmitt-Kopplin P, Matzka M, Ruf A, et al. Complex carbonaceous matter in Tissint martian meteorites give insights into the diversity of organic geochemistry on Mars. *Sci Adv* 2023;9:eadd6439; doi: 10.1126/sciadv.add6439.
- Segré D, Ben-Eli D, Deamer DW, et al. The lipid world. *Orig Life Evol Biosph* 2001;31(1):119–145; doi: 10.1023/A:1006746807104.
- Sephton M. Organic compounds in carbonaceous meteorites. *Natural Product Reports* 2002;19(3):292–311; doi: 10.1039/B103775G.
- Sephton M. Organic matter in carbonaceous meteorites: Past, present and future research. *Philos Trans A Math Phys Eng Sci* 2006;363:2729–42; doi: 10.1098/rsta.2005.1670.

- Sephton MA. Pyrolysis and mass spectrometry studies of meteoritic organic matter. *Mass Spectrom Rev* 2012;31(5):560–569; doi: 10.1002/mas.20354.
- Sephton MA, Pillinger CT, Gilmour I. Normal alkanes in meteorites: Molecular $\Delta^{13}\text{C}$ values indicate an origin by terrestrial contamination. *Precambrian Res* 2001a;106(1):47–58; doi: 10.1016/S0301-9268(00)00124-8.
- Sephton MA, Pillinger CT, Gilmour I. Supercritical fluid extraction of the non-polar organic compounds in meteorites. *Planet Space Sci* 2001b;49(1):101–106; doi: 10.1016/S0032-0633(00)00122-7.
- Shiea J, Brassell SC, Ward DM. Mid-chain branched mono- and dimethyl alkanes in hot spring cyanobacterial mats: A direct biogenic source for branched alkanes in ancient sediments? *Org Geochem* 1990;15(3):223–231; doi: 10.1016/0146-6380(90)90001-G.
- Shimoyama A. Complex organics in meteorites. *Adv Space Res* 1997;19(7):1045–1052; doi: 10.1016/S0273-1177(97)00351-7.
- Shimoyama A, Naraoka H, Komiya M, et al. Analyses of carboxylic acids and hydrocarbons in Antarctic carbonaceous chondrites, Yamato-74662 and Yamato-793321. *Geochem J* 1989;23(4):181–193; doi: 10.2343/geochemj.23.181.
- Shivaji S, Prakash JSS. How do bacteria sense and respond to low temperature? *Arch Microbiol* 2010;192(2):85–95; doi: 10.1007/s00203-009-0539-y.
- Simkus DN, Aponte JC, Elsila JE, et al. New insights into the heterogeneity of the Tagish Lake meteorite: Soluble organic compositions of variously altered specimens. *Meteorit Planet Sci* 2019;54(6):1283–1302; doi: 10.1111/maps.13276.
- Steele A, McCubbin FM, Fries MD. The provenance, formation, and implications of reduced carbon phases in martian meteorites. *Meteorit Planet Sci* 2016;51(11):2203–2225; doi: 10.1111/maps.12670.
- Steele A, Benning LG, Wirth R, et al. Organic synthesis on Mars by electrochemical reduction of CO_2 . *Sci Adv* 2018;4(10):eaat5118; doi: 10.1126/sciadv.aat5118.
- Steele A, Benning LG, Wirth R, et al. Organic synthesis associated with serpentinization and carbonation on early Mars. *Science* 2022;375(6577):172–177; doi: 10.1126/science.abg7905.
- Summons RE, Walter MR. Molecular fossils and microfossils of prokaryotes and protists from Proterozoic sediments. *Am J Sci* 1990;290-A:212–244.
- Summons RE, Albrecht P, McDonald G, et al. Molecular biosignatures. In *Strategies of Life Detection*. (Botta O, Bada JL, Gomez-Elvira J, et al. eds.) Space Sciences Series of ISSI Springer US: Boston, MA; 2008; pp 133–159; doi: 10.1007/978-0-387-77516-6_11.
- Summons RE, Welander PV, Gold DA. Lipid biomarkers: Molecular tools for illuminating the history of microbial life. *Nat Rev Microbiol* 2022;20(3):174–185; doi: 10.1038/s41579-021-00636-2.
- Vago JL, Westall F, Coates AJ, et al. Habitability on early Mars and the search for biosignatures with the ExoMars Rover. *Astrobiology* 2017;17(6–7):471–510; doi: 10.1089/ast.2016.1533.
- van den Brink DM, Wanders RJA. Phytanic acid: Production from phytol, its breakdown and role in human disease. *Cell Mol Life Sci* 2006;63(15):1752; doi: 10.1007/s00018-005-5463-y.
- Vinnichenko G, Jarrett AJ, Hope JM, et al. Discovery of the oldest known biomarkers provides evidence for phototrophic bacteria in the 1.73 Ga Wollongorang Formation, Australia. *Geobiology* 2020;18(5):544–559; doi: 10.1111/gbi.12390.
- Wang Y, Huang Y, Alexander CMO 'D, et al. Molecular and compound-specific hydrogen isotope analyses of insoluble organic matter from different carbonaceous chondrite groups. *Geochim Cosmochim Acta* 2005;69(14):3711–3721; doi: 10.1016/j.gca.2005.03.008.
- Wilhelm MB, Davila AF, Eigenbrode JL, et al. Xeropreservation of functionalized lipid biomarkers in hyperarid soils in the Atacama Desert. *Org Geochem* 2017;103:97–104; doi: 10.1016/j.orggeochem.2016.10.015.
- Wilhelm MB, Ricco AJ, Buckner DK, et al. Extraction instruments to enable detection of origin-diagnostic lipids for life detection. In *52nd Lunar and Planetary Science Conference*. Lunar and Planetary Institute, Houston, 2021; abstract 2634.
- Williams AJ, Eigenbrode J, Floyd M, et al. Recovery of fatty acids from mineralogic Mars analogs by TMAH thermochemolysis for the Sample Analysis at Mars Wet Chemistry Experiment on the Curiosity Rover. *Astrobiology* 2019;19:522–546; doi: 10.1089/ast.2018.1819.
- Williams AJ, Craft KL, Millan M, et al. Fatty acid preservation in modern and relict hot-spring deposits in Iceland, with implications for organics detection on Mars. *Astrobiology* 2021;21(1):60–82; doi: 10.1089/ast.2019.2115.
- Yuan Y, Lee RE, Besra GS, et al. Identification of a gene involved in the biosynthesis of cyclopropanated mycolic acids in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 1995;92(14):6630–6634; doi: 10.1073/pnas.92.14.6630.
- Yuen G, Blair N, Des Marais DJ, et al. Carbon isotope composition of low molecular weight hydrocarbons and monocarboxylic acids from Murchison meteorite. *Nature* 1984;307(5948):252–254; doi: 10.1038/307252a0.
- Zeng L, Dehesh K. The eukaryotic MEP-pathway genes are evolutionarily conserved and originated from Chlamydia and Cyanobacteria. *BMC Genomics* 2021;22(1):137; doi: 10.1186/s12864-021-07448-x.

Address correspondence to:
Denise Buckner
NASA Ames Research Center
M/S n288-1
Moffett Field, CA 94035
USA

E-mail: bucknerd@ufl.edu

Submitted 26 January 2023

Accepted 8 October 2023

Associate Editor: Sherry Cady

Abbreviations Used

ASE = accelerated solvent extraction
ExCALiBR = Extractor for Chemical Analysis of Lipid Biomarkers in Regolith
FTT = Fischer Tropsch-type
GC-MS = gas chromatography–mass spectrometry
IOM = insoluble organic macromolecule
PCA = principal component analysis
UCM = unresolved complex mixtures

(Appendix follows →)

A. Appendix

A.1. Definitions

- Acyclic hydrocarbon: straight-chain lipid containing only carbon and hydrogen atoms, arranged in various conformations.
- Alkene: an acyclic hydrocarbon with one or more unsaturations.
- Biomarker: organic molecule synthesized by life and preserved in the geologic record, serving as a molecular fossil that records evidence of previous biological activity (Peters and Moldowan, 1993).
- C_{\max} : the molecule with the highest abundance within a lipid sample, where the sample contains molecules within the same lipid class (*e.g.*, fatty acids only, acyclic hydrocarbons).
- Even or odd carbon number predominance: predominance of molecules with chain lengths that contain either even numbers of carbon atoms or odd numbers of carbon atoms.
- Fatty acid: straight-chain fatty acid comprising an acyclic hydrophobic hydrocarbon “tail” bound to a hydrophilic carboxyl group ($-\text{COOH}$) “head” at the opposite end of the carbon chain relative to the tail.
- Geolipid: a lipid preserved in the rock record.
- Homologous series: series of numerous molecules of many chain lengths that each display the same non-random branching features, including length and/or position (McCollom *et al.*, 1999).
- Insoluble organic macromolecule (IOM): macromolecular materials found in carbonaceous chondrites, comprising smaller aromatic, aliphatic, and functionalized organic fragments that are bound in a complex network via alkyl and oxygen bridging (Gilmour, 2003; Alexander *et al.*, 2007).
- Isoprene: other name of 2-methyl-1,3-butadiene (C_5H_8).
- Isoprenoid: a biotically synthesized lipid made up of isoprene subunits, linked together in various conformations. Isoprenoids exist in both cyclic and acyclic conformations, but for the purposes of this paper, “isoprenoid” refers to acyclic compounds.
- Lipid sample: a set of preserved lipids (either fatty acids or acyclic hydrocarbons) extracted from a geological specimen.
- Lipid: an organic molecule that is soluble in nonpolar organic solvents. While “lipids” traditionally refer to compounds of biological origin (IUPAC, n.d.), for our purposes we refer to these classes of organics (*i.e.*, carboxylic [fatty] acids and acyclic hydrocarbons) as lipids, regardless of whether their origin was biotic or abiotic. This is both for simplicity’s sake and because within an astrobiological context an extraterrestrial organic molecule of unknown origin can fit the functional definition of a lipid (*i.e.*, soluble in organic solvents) while its synthesis is unknown.
- Molecular structure: a physical feature relating to configuration, carbon bonding, or other molecular subcomponent or feature with a discrete structure, which can include the a chain of carbon and hydrogen atoms, double or single bonding between carbon atoms, branching, or cyclization.
- Monounsaturated: said of a molecule with a single double bond.
- *n*-Alkane: an acyclic hydrocarbon with no unsaturations and no branches.
- Origin-diagnostic: a distinctive distribution of sample features observed in a set of lipids that indicates whether those molecules were synthesized biotically or abiotically.
- Polyunsaturated: said of a molecule with multiple double bonds.
- Unresolved complex mixture (UCM): complex mixture of numerous branched, cyclic, and acyclic molecules with numerous compounds such that separation and characterization of individual molecular structures is not feasible with traditional GC-MS methods.
- Unsaturation: a double bond between two carbon atoms ($\text{C}=\text{C}$).

A.2. Extraction techniques

- Accelerated solvent extraction (ASE): a technique for rapidly extracting lipids via organic solvents under elevated temperatures and pressures in a closed cell, maintaining the liquid phase while above the boiling point (Richter *et al.*, 1996).
- Bligh and Dyer: a technique for extracting lipids from geological or biological samples (*e.g.*, tissues, cultures) using a $\text{CH}_3\text{Cl}:\text{CH}_3\text{OH}:\text{H}_2\text{O}=2:2:1.8$ (v/v/v) solvent mixture (Bligh and Dyer, 1959).
- Hydrous pyrolysis: thermal-based extraction technique that uses superheated water in conjunction with pyrolysis to liberate organics from mineral matrixes. Hydrous pyrolysis simultaneously liberates both free and IOM-bound moieties (Komiya and Shimoyama, 1996; Shimoyama, 1997; Oba and Naraoka, 2006).
- Modified Bligh and Dyer: lipid extraction technique based on the methods of Bligh and Dyer that maintains the 2:2:1.8 (v/v/v) ratio but includes modifications, for example, CH_2Cl_2 substituting for CH_3Cl , a pH buffer added to the H_2O , various time spent in solvent, and so on. Many modifications exist, but for the purposes of this paper, “Modified Bligh and Dyer” is used whenever the study being referenced reports their methods with this terminology.
- Pyrolysis: thermal-based extraction technique for liberating organics from mineral matrices via heat, with no solvents utilized. Upon heating, analytes are flashed directly into the GC-MS inlet. Pyrolysis simultaneously liberates both free and IOM-bound moieties.
- Sequential solvent extraction: lipid extraction technique that uses a series of solvents (as opposed to a single mixture) to extract lipids.
- Solid-phase micro-extraction (SPME): a technique for extracting specific classes of molecules without the use of solvents. Analytes of interest are directly adsorbed onto coated fibers then desorbed into the GC-MS inlet.
- Solvent extraction: lipid extraction technique that uses a single organic solvent or a mixture of solvents (with a ratio other than 2:2:1.8), without the addition of ultrasonic energy, at variable times and temperatures.
- Soxhlet: a technique for extracting lipids via refluxing in organic solvents through a closed-cell Soxhlet ap-

(Appendix continues →)

paratus, for variable amounts of time (Luque de Castro and Priego-Capote, 2010).

- Ultrasonic extraction: lipid extraction technique that uses organic solvents with the addition of ultrasonic energy, at variable times and temperatures (Lee *et al.*, 2010; Kong *et al.*, 2014).

A.3. Analytical method

- Gas chromatography–mass spectrometry (GC-MS)
- Gas chromatography–flame ionization detection (GC-FID)
- Pyrolysis–gas chromatography–mass spectrometry (py-GC-MS)