



The Cell and the Sum of Its Parts: Patterns of Complexity in Biosignatures as Revealed by Deep UV Raman Spectroscopy

Haley M. Sapers^{1,2,3*}, Joseph Razzell Hollis², Rohit Bhartia^{2*}, Luther W. Beegle², Victoria J. Orphan^{1*} and Jan P. Amend^{3,4}

¹ Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, United States, ² Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, ³ Department of Earth Sciences, University of Southern California, Los Angeles, CA, United States, ⁴ Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States

OPEN ACCESS

Edited by:

Nancy Hinman,
University of Montana, United States

Reviewed by:

Kai Waldemar Finster,
Aarhus University, Denmark
James F. Holden,
University of Massachusetts Amherst,
United States

*Correspondence:

Haley M. Sapers
haley.m.sapers@jpl.nasa.gov
Rohit Bhartia
rbhartia@jpl.nasa.gov
Victoria J. Orphan
vorphan@gps.caltech.edu

Specialty section:

This article was submitted to
Microbiological Chemistry
and Geomicrobiology,
a section of the journal
Frontiers in Microbiology

Received: 18 September 2018

Accepted: 18 March 2019

Published: 14 May 2019

Citation:

Sapers HM, Razzell Hollis J, Bhartia R, Beegle LW, Orphan VJ and Amend JP (2019) The Cell and the Sum of Its Parts: Patterns of Complexity in Biosignatures as Revealed by Deep UV Raman Spectroscopy. *Front. Microbiol.* 10:679. doi: 10.3389/fmicb.2019.00679

The next NASA-led Mars mission (Mars 2020) will carry a suite of instrumentation dedicated to investigating Martian history and the *in situ* detection of potential biosignatures. SHERLOC, a deep UV Raman/Fluorescence spectrometer has the ability to detect and map the distribution of many organic compounds, including the aromatic molecules that are fundamental building blocks of life on Earth, at concentrations down to 1 ppm. The mere presence of organic compounds is not a biosignature: there is widespread distribution of reduced organic molecules in the Solar System. Life utilizes a select few of these molecules creating conspicuous enrichments of specific molecules that deviate from the distribution expected from purely abiotic processes. The detection of far from equilibrium concentrations of a specific subset of organic molecules, such as those uniquely enriched by biological processes, would comprise a universal biosignature independent of specific terrestrial biochemistry. The detectability and suitability of a small subset of organic molecules to adequately describe a living system is explored using the bacterium *Escherichia coli* as a model organism. The DUV Raman spectra of *E. coli* cells are dominated by the vibrational modes of the nucleobases adenine, guanine, cytosine, and thymine, and the aromatic amino acids tyrosine, tryptophan, and phenylalanine. We demonstrate that not only does the deep ultraviolet (DUV) Raman spectrum of *E. coli* reflect a distinct concentration of specific organic molecules, but that a sufficient molecular complexity is required to deconvolute the cellular spectrum. Furthermore, a linear combination of the DUV resonant compounds is insufficient to fully describe the cellular spectrum. The residual in the cellular spectrum indicates that DUV Raman spectroscopy enables differentiating between the presence of biomolecules and the complex uniquely biological organization and arrangements of these molecules in living systems. This study demonstrates the ability of DUV Raman spectroscopy to interrogate a complex biological system represented in a living cell, and differentiate between organic detection and a series of Raman features that derive from the molecular complexity inherent to life constituting a biosignature.

Keywords: biosignatures, deep UV Raman spectroscopy, Mars 2020, SHERLOC, spectral deconvolution

INTRODUCTION

The search for life beyond Earth has motivated decades of planetary exploration from the first life detection experiments on board the Viking missions to Mars in 1975 (Klein et al., 1972; Soffen and Snyder, 1976; Levin and Straat, 1977) to current exploration by the Mars Science Laboratory (Summons et al., 2011; Grotzinger et al., 2012) to future missions to potentially habitable icy bodies in our Solar System (e.g., Phillips and Pappalardo, 2014). Mars is the closest, potentially habitable body in the Solar System (e.g., Farmer and Des Marais, 1999; Des Marais, 2010; Grotzinger et al., 2013; Filiberto et al., 2014), and near-term missions including Mars 2020 (Mustard et al., 2013; Williford et al., 2018) and ExoMars (Vago et al., 2017) seek to understand the distribution of habitability criteria, such as the spatial and temporal extent of liquid water and availability of carbon and energy sources, in addition to the detection and evaluation of putative biosignatures. Mars 2020 will carry a suite of instrumentation capable of detecting and characterizing geochemical and organic patterns in a variety of samples that will be collected from pre-characterized sites using available orbital datasets and cached for eventual sample return and further evaluation for evidence of life (Williford et al., 2018).

On the Mars 2020 payload, the Scanning Habitable Environments with Raman and Luminescence for Organic and Chemicals (SHERLOC); a list of abbreviations and acronyms is provided below, is a DUV Raman spectrometer capable of mapping organic compositions of target samples (Beegle et al., 2015). Relevant to this study, SHERLOC has the ability to detect the Raman peaks of organic molecules *in situ* at concentrations below 0.1 wt% in a 100 micron area and 10–5 wv over an observed area of 5 mm × 5 mm (Abbey et al., 2017). This includes the nucleobases and amino acids known to be essential components of terrestrial life. Furthermore, these concentrations can be mapped across a surface in order to show spatial distributions. However, the mere presence of organic compounds is not evidence of life – here we demonstrate a spectral threshold for biogenicity based on the Raman active components within a cell compared to the Raman spectra of the cell itself. These results provide a context for the interpretation of DUV Raman spectra of organic molecules collected by SHERLOC on Mars 2020 as potential biosignatures.

The formation of a biosignature requires that a biological process results in the accumulation of a biogenic ‘signal’ that differs significantly from the background abiotic ‘noise.’ Detection requires that the signal is in high enough concentration, or chemically and physically distinct enough from

the background environment that it is both distinguishable and not subject to attrition (Des Marais et al., 2008; Des Marais, 2013; Hays et al., 2017). Raman spectroscopy can be used to detect the unique biosignature of a cell as the enrichment of specific organic molecules, in the same location with sufficient structural complexity that cannot be adequately explained by known abiotic processes. In this paper we are focused on the analysis of a single DUV Raman spectrum obtained from *Escherichia coli* cells harvested during exponential growth, without further spatial or mineralogical context, to determine if it is possible to distinguish the unique chemical biosignature of those cells from their DUV resonant molecular components alone. While minor variations in Raman spectra have been used to differentiate different microbial species (Huang et al., 2004; Pahlow et al., 2015), the dominant vibrational modes are shared reflecting similar macromolecular compositions in other bacterial cells (Wu et al., 2001), viruses (Wen and Thomas, 1998), and eukaryotic cells (Kumamoto et al., 2012).

While the chemical structure of abiotically synthesized and biogenically produced organic molecules do not differ, the distribution and co-occurrence patterns of the particular compounds is significant. Life exploits boundary conditions to harness energy and as such its distribution reflects this. Life is not homogenous: complexity in distribution is a fundamental property of life (Bhartia et al., 2010). Distribution can be described by two parameters: spatial and constituent. The inventory of organic molecules is significant as it reflects the selectivity of uniquely biological processes. The spatial distribution of these molecules is also significant: the presence of nucleobases and proteinaceous amino acids within the same sample is not necessarily a biosignature, however, a conspicuous enrichment co-occurring spatially and temporally is difficult to reconcile abiotically. These reduced organic carbon molecules are the fundamental building blocks of terrestrial life; however, they are not unique to life. Nucleobases including adenine, guanine, and uracil have been found in the Murchison, Murray and Orgueil Martian meteorites (Stoks and Schwartz, 1979; Martins et al., 2008; Steele et al., 2016) and over 80 amino acids, including 55 α -amino acids, have been detected in carbonaceous chondrites (Sephton, 2002; Sephton and Botta, 2005; Pizzarello et al., 2006; Schmitt-Kopplin et al., 2010; Burton et al., 2012). Of nearly 4000 plausible α -amino acids structures (Meringer et al., 2013), only ~700 have been isolated from biological systems (Hunt, 1985). From these 700 amino acids, only 20 are translationally encoded in all lineages of terrestrial life (Wong, 1975; Hardy, 1985). Only 8 of the 20 amino acids comprising the standard genetic code have been observed in extraterrestrial samples (Pizzarello et al., 2006; Burton et al., 2012) and only about half of the translationally encoded amino acids can be accounted for experimentally via abiotic synthesis and prebiotic simulations (Miller, 1953; Muñoz Caro et al., 2002; Johnson et al., 2008; Higgs and Pudritz, 2009; Cleaves, 2010; Parker et al., 2011). These observations have divided the 20 universally translated amino acids into early and late groups (Wong, 1975, 2005; Higgs and Pudritz, 2009). The former represents simple amino acids that can be formed prebiotically through abiotic processes comprising the earliest genetic code. The latter group were incorporated into the genetic code following the evolution of biosynthetic pathways

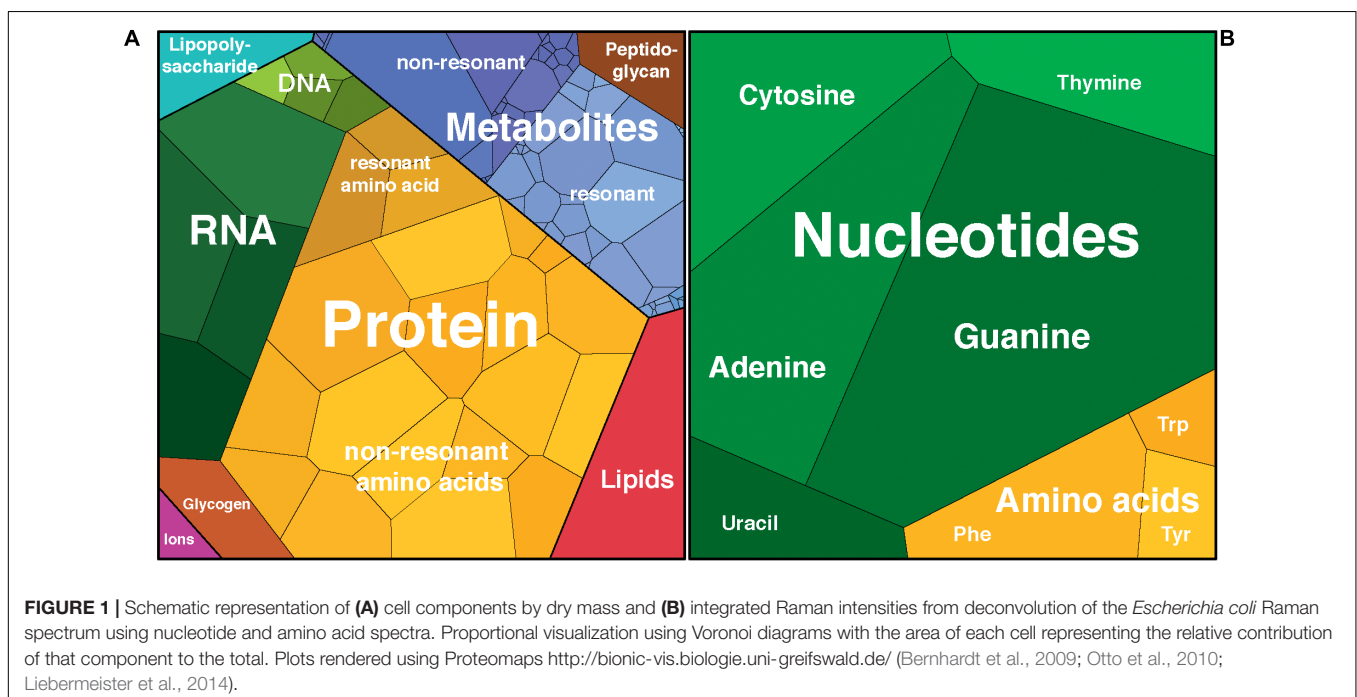
Abbreviations: AAA, aromatic amino acid. In this paper, refers to phenylalanine, tryptophan, and tyrosine; dATP, deoxyribose adenosine triphosphate, a nucleotide of adenine; dCTP, deoxyribose cytidine triphosphate, a nucleotide of cytosine; dGTP, deoxyribose guanosine triphosphate, a nucleotide of guanine; dTTP, deoxyribose thymidine triphosphate, a nucleotide of thymine; DUV, deep ultraviolet, light with a wavelength ~100–300 nm; MOBIUS, Mineralogy and Organic Based Investigations with UV Spectroscopy; Nucleobase, molecular derivatives of purine and pyrimidine. In this paper, refers to adenine, cytosine, guanine, thymine and uracil; Nucleotide, molecules containing a nucleobase, a ribose unit and a triphosphate group; SHERLOC, Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals; UTP, ribose uridine triphosphate, a nucleotide of uracil.

modifying simpler precursors (Wong, 1975). Calculations of the Gibbs free energy of formation for each of the 20 proteinaceous amino acids indicate that the latter group requires a significantly higher energy cost (Amend and Shock, 1998) and to date these amino acids have not been observed in extraterrestrial materials implying their presence requires biosynthetic pathways. Thus an organic biosignature is not simply the enrichment of a specific subset of organic molecules, but that the molecules enriched display a structural complexity not explained or expected to be produced by purely abiotic processes. As recognized by Nelson et al. (1992), ‘the cell is more than the sum of its parts’; we leverage this to illustrate a spectral distinction between a collection of organic molecules and those that comprise a living system.

Identifying complex mixtures of similar compounds *in situ* is challenging, as many analytical techniques either consume the sample and employ chromatographic separation methods or can only probe bulk composition at length scales far beyond that of individual cells. Assessing the biogenicity of a putative biosignature requires multiple, complementary analytical techniques and contextual information including spatial distribution, destructive and bulk methods are limited. Raman spectroscopy offers the necessary sensitivity to chemical structure, at spatial resolutions comparable to the size of a cell, without destroying the sample. By using DUV excitation, we can also exploit the combined signal enhancement of both high-frequency excitation and molecular resonance with opto-electronic transitions (Nelson et al., 1992; and references therein; Tarcea et al., 2007). This enables the identification of aromatic components within cellular materials even at very low concentrations that would otherwise be undetectable using more conventional excitation wavelengths, such as the 532 and 633 nm lasers employed in Green and Red Raman

respectively (Beegle et al., 2015; and references therein). The Raman scattering intensity is related to excitation frequency such that high frequency excitation leads to a greater proportion of Raman-scattered light for a given laser power (Long, 1977). Using DUV excitation also provides resonance with the π - π^* absorption band of many aromatic molecules, including the nucleic acids and some amino acids, leading to an overall increase in scattering cross-section of up to 10,000x (Asher and Johnson, 1984; Asher and Murtaugh, 1988; Ianoul et al., 2002) vs. non-resonant, lower-frequency excitation. Resonance provides particular sensitivity to minor conformational and structural changes that involve the aromatic ring (Asher, 1993; Toyama et al., 1999), and resonant Raman has been used previously to probe molecular conformers, intermolecular packing, and photo-oxidation reactions in aromatic compounds (Razzell-Hollis et al., 2014; Wade et al., 2017; Wood et al., 2017). Identification of molecular structures by the pattern of peaks in the Raman spectrum is made more challenging when several similar molecules are present together, as the identifying peaks of one molecule may overlap with modes from others. However, by using DUV excitation to resonantly enhance signals from aromatic molecules, we can reduce the number of detectable molecules to a smaller subset that still constitute a distinctive biosignature. For terrestrial cells this subset has been established to consist of the five nucleobases and three aromatic amino acids (AAAs) (Britton et al., 1988; Nelson et al., 1992; Chadha et al., 1993). We therefore define a set of molecular standards based on these eight aromatic molecules (Figure 1).

By using *E. coli* as a model organism, we can demonstrate that not only does its DUV Raman spectrum reflect the enrichment of specific aromatic molecules, but that molecular complexity,



i.e., spectra from nucleotides rather than simple nucleobases, is required to deconvolute the cellular spectrum. We also illustrate the ability of DUV Raman spectroscopy to differentiate between the spectrum of a cell and a representative artificial mixture of its Raman resonant components, i.e., whether the cell is more than the sum of its parts and if this itself constitutes a distinctive biosignature. Here we present an illustration of the importance of structural complexity in biosignatures by systematically increasing the complexity of biomolecules to deconvolute the DUV Raman spectra of *E. coli* into its constituent DUV resonant components.

MATERIALS AND METHODS

Escherichia coli Cultures

Escherichia coli K12 was grown from frozen stocks overnight in 1 mL of defined minimal media (M9) containing 0.4% glucose, 47.6 mM Na₂HPO₄, 22.06 mM KH₂PO₄, 8.56 mM NaCl, 18.7 mM NH₄Cl, 99.12 μM CaCl₂ and 0.1 mg/L thiamine, pH adjusted to 7.0 and filtered sterilized through a 0.22 μm PES membrane filter. Cultures were incubated in a shaker at 37°C and were propagated to a sufficient volume for subsequent sampling. Cells were further transferred three times during mid-log growth as determined by measuring the absorbance at 600 nm (OD₆₀₀) using a DR-2700 spectrophotometer (Hach, Inc.). Triplicate 150 mL cultures were established and cells were harvested aseptically after 4 h during exponential growth and fixed with 4% paraformaldehyde for 1 h at room temperature. It has been noted that fixation does not influence the cellular spectra and also prevents spectral changes due to radiation-induced stress observed in live cells (Kumamoto et al., 2011). Following fixation cells were pelleted, washed twice in phosphate buffered saline (PBS), resuspended in 50% PBS, and finally re-suspended in MilliQ H₂O to an OD₆₀₀ of 0.2 (1.6 × 10⁸ cells/ml) based on the initial optical density reading. 2 μL of the washed and re-suspended sample was spotted onto a sterile aluminum wafer (Multipurpose 6061, McMaster-Carr) and allowed to air dry prior to Raman analysis. Given a laser diameter of approximately 68 μm and a dry spot with a diameter of 2 mm, each laser spot would interrogate ~370 cells, assuming a roughly equal distribution of cells. A 50 μl droplet of cell suspension was also measured with Raman immediately to assess spectral artifacts produced by drying. The DUV Raman spectrum of the aluminum wafer displayed no intrinsic vibrational modes (Supplementary Figure S1).

Molecular Standards

Samples of the nucleobases adenine (Sigma, A8751), cytosine (Sigma, C3506), guanine (Aldrich, G11950), thymine (Sigma, T0376), uracil (Alfa Aesar, A15570), and the amino acids phenylalanine (Sigma, P2126), tryptophan (Sigma, T0254), tyrosine (Sigma, T3754), were individually dissolved in MilliQ H₂O at a concentration of 10 mM. 5M NaOH was added dropwise to the guanine solution until it fully dissolved. The deoxyribonucleotides dATP, dCTP, dGTP, dTTP (Sigma-Aldrich, DNTP100) were received as PCR-grade 100 mM aqueous solutions, the ribonucleotide UTP (Sigma-Aldrich, T25) was

received as a powder that was subsequently dissolved in MilliQ H₂O to a final concentration of 100 mM. Custom DNA/RNA strands were ordered (Sigma-Aldrich, VC00021 and VC40001) with the following single-strand 10mer sequences: DNA-A: 5'-AAAAAAAAAA-3', DNA-C: 5'-CCCCCCCC-3', DNA-G: 5'-GGGGGGGG-3', DNA-T: 5'-TTTTTTTT-3', RNA-U: 5'-UUUUUUUU-3'. One 19 unit ssDNA strand, 5'-CAATTGTACTAGCCGGATC-3', was designed to incorporate every possible base-pair combination without forming secondary structures, as assessed using the NUPACK analysis on-line tool¹. All oligomers were received as 100 μM solutions. All solutions were diluted 1:1 with a 100 mM aqueous solution of Na₂SO₄, as an internal standard, and 50 μL of solution was dropped onto an Al wafer immediately prior to measurement. DUV Raman measurements were completed within 20 min of deposition to minimize the impact of evaporation.

Artificial Mixture

A mixture of molecular standards was prepared according to the relative concentrations of the various major aromatic residues in *E. coli* undergoing rapid division with a doubling time of 40 min (see Table 3). The numbers of residues per cell were calculated from macromolecular composition data adapted by Milo et al. (2010) from the reports of Nierlich (1972), Neidhardt et al. (1990), and Neidhardt (1996), the proteome database from Kozłowski (2017), and the metabolite pool reported by Bennett et al. (2009). Because macromolecular nucleic acids represent such a large proportion of nucleobase residues, in order to accurately represent the composition of the cell, the DNA/RNA standards were used for the A, C, G, T, and U residues of nucleic acid, while dATP, dCTP, dGTP, dTTP, and UTP were used for the A, C, G, T, and U-containing free nucleotides. Phe, Trp, and Tyr were used for their equivalent residues in both macromolecular protein and the free metabolite pool. The total concentration of the mixture was 1.00 mM, equating to 0.26 fg per 0.9 μm³. Although this is a significantly lower total concentration of aromatic residues compared to that of the cell, a spectrum with a signal-to-noise ratio of 186:1 was still obtained. The spectrum exhibits the same major peaks as the cell spectrum, with the notable exception of a much more intense peak around ~1600 cm⁻¹ and a generally lower intensity for the minor peak regions (<1200 cm⁻¹ and in between the major peaks).

DUV Raman

MOBIUS (Mineralogy and Organic Based Investigations with UV Spectroscopy), A custom DUV resonance Raman spectrometer at the NASA Jet Propulsion Laboratory, was used for all measurements. MOBIUS uses a 248.6 nm NeCu pulsed laser (Photon Systems, Inc.) reflected off of a 248 nm RazorEdge ultra-steep long-pass edge filter (Semrock, Inc.) and focused onto the sample through a DUV chromatically corrected objective lens with a numerical aperture of 0.13 (ThorLabs LMU-5x-UVB). Raman-scattered photons were collected using 180° backscatter geometry, a Horiba 550i spectrometer, and a Horiba Symphony e2v 42-10 CCD liquid nitrogen cooled (-140°C) detector. Based on a 550 mm focal length, a slit width of 250 μm, and a

¹nupack.org

grating groove density of 1800 lines/mm, the spectral accuracy was 3.8 cm^{-1} and the true spectral resolution (minimum peak-to-peak separation for distinguishing overlapping peaks) was $\sim 25 \text{ cm}^{-1}$. A laser spot diameter of $\sim 75 \mu\text{m}$ had an energy at the sample of 0.8–1.2 $\mu\text{J}/\text{pulse}$ and a pulse width of 40 μs for cell measurements integrated over a total of 1200 pulses per point in order to minimize photodamage to the cell (Taguchi et al., 2011). The laser was replaced prior to measurements of the molecular standards, and adjusted to maintain a consistent output energy of 0.8–1.2 μJ per pulse. A total of 25 points in a 5×5 array were acquired for each sample. Prior to sample data collection, calibration was achieved by validating the position of the secondary laser line at 252.93 nm (McNeil et al., 1978) and at zero-order reflection. Resulting spectra were corrected for laser intensity variability using a normalized laser intensity correction factor, which represents the relative laser intensity during data acquisition. Cosmic rays were identified as outliers in the distribution of intensity values in each wavelength channel (Uckert and Bhartia, 2019) and replaced by the value of adjacent points. Downstream processing was completed using a combination of the R package HyperSpec and in-house python scripts utilizing SciPy (Jones et al., 2001) and LMFIT (Newville et al., 2014). Raman shifts were recalibrated using the atmospheric N_2 peak to a standard peak position of 2330 cm^{-1} (Burris et al., 1992). The spectra were then sectioned to focus on the ‘organic fingerprint range’ between ~ 800 and 1800 cm^{-1} (Zhu et al., 2011) and least-squares regression used to subtract a linear background. All steps are visualized in **Supplementary Figure S2**. Spectra in each sample were averaged and all cellular samples were intensity normalized relative to each other to the mean intensity of the guanine peak at $\sim 1460 \text{ cm}^{-1}$. Cell spectra were fitted by scalable linear combination of individual molecular standard spectra, based on non-linear least-squares regression of all points between 800 and 1800 cm^{-1} , which was done using the built-in functionality of python and the LMFIT package (Newville et al., 2014).

RESULTS

Molecular Standards

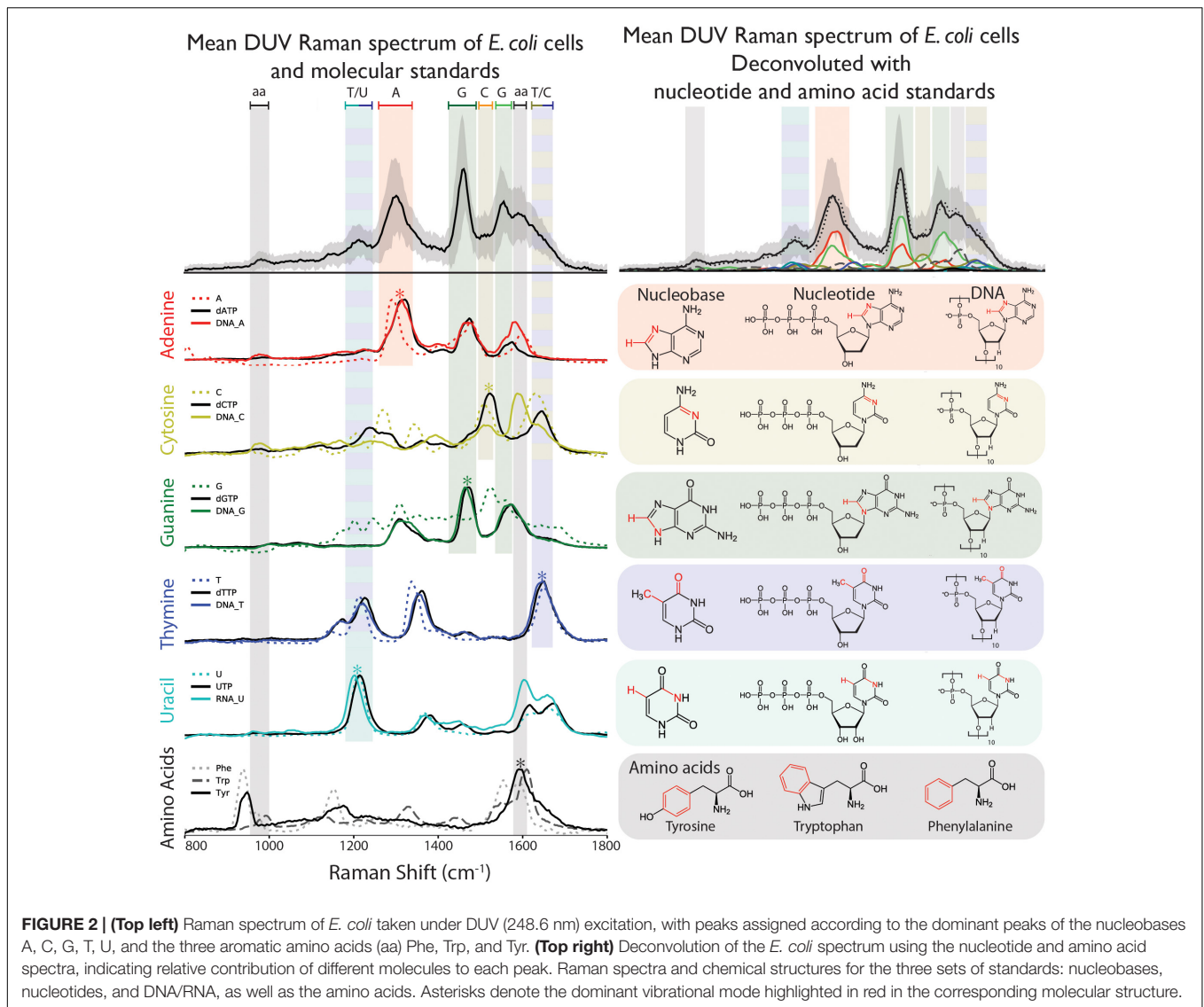
Eight aromatic molecules: five nucleobases (adenine, A, cytosine, C, guanine, G, thymine, T, and uracil, U) and three amino acids (phenylalanine, Phe, tryptophan, Trp, and tyrosine, Tyr) that are known to contribute to the observed vibrational modes in the cellular spectrum were measured separately in solution. The nucleobases, which contribute the majority of Raman scattering at our chosen excitation wavelength, were measured in multiple forms of increasing structural complexity: simple nucleobases, deoxyribose/ribose nucleotide triphosphates, and single-stranded DNA/RNA 10-base oligomers containing mono-nucleotides of A, C, G, or T. As shown in **Figure 2**, each component exhibits a unique DUV Raman spectrum dominated by the resonant vibrational modes of its aromatic rings, with major peak positions and mode assignments presented in **Table 1**. For the simple nucleobases A, C, G, T and U, the dominant modes were 1291, 1512, 1440, 1647, and 1210 cm^{-1} , respectively, that have been

previously assigned to coupled vibrations of multiple bonds on each aromatic moiety (see **Table 1**) along with additional strong modes and several minor modes across the $800\text{--}1800 \text{ cm}^{-1}$ range (see **Supplementary Table S1** for assignments). The amino acids Phe, Trp, and Tyr exhibit similar spectra to one another dominated by the ring-stretching mode of the aromatic moiety at $\sim 1600 \text{ cm}^{-1}$, in all three amino acids (Jenkins et al., 2005).

The nucleotides exhibit similar spectra to their respective nucleobases, though the frequencies and relative intensities of many peaks are altered by the addition of the ribose triphosphate. Specifically, the dominant modes in the nucleotide spectra are shifted to higher frequencies: adenine by 31 cm^{-1} , cytosine by 12 cm^{-1} , guanine by 34 cm^{-1} , thymine and uracil by 4 cm^{-1} . In most cases the number of peaks is unchanged, with the exception of dCTP, which exhibits fewer peaks than cytosine in the $1200\text{--}1400 \text{ cm}^{-1}$ region. The spectra of the single-stranded DNA/RNA standards, each containing 10 units of a particular base, exhibit relatively slight spectral changes over the nucleotides: dominant peaks were consistently shifted to lower frequencies by $2\text{--}10 \text{ cm}^{-1}$, with some alterations in relative intensities with respect to other modes. The most significant exception was the DNA-C 10-mer, which exhibits a strong mode (at $\sim 1574 \text{ cm}^{-1}$) not previously observed directly in either cytosine or dCTP.

Cellular Spectra

The spectrum of fixed dry cells was functionally identical to that of a fixed cell suspension (see **Supplementary Figure S3**) but provided better signal and a detectable N_2 peak for reliable calibration. There were no differences between the spectra of each replicate (**Supplementary Figure S4**) and replicate A was used for further analysis. The DUV Raman spectrum of *E. coli* cells is dominated by peaks that are attributed to specific molecules based on comparison to their dominant vibrational modes: the nucleobases adenine, guanine, cytosine, and thymine, and the AAAs tyrosine, tryptophan, and phenylalanine (**Figure 2**). The predominance of these molecules in the DUV Raman spectrum can be explained by the resonant enhancement of their aromatic ring vibrations, which occurs when the excitation energy is comparable to the $\pi\text{-}\pi^*$ electronic transition localized on the aromatic ring. The molecular vibration of atmospheric N_2 and O_2 at ~ 2331 and $\sim 1550 \text{ cm}^{-1}$, respectively, were quantified and subtracted from the spectrum prior to further analysis. The cellular spectrum is characterized by three major peaks centered at ~ 1310 , ~ 1470 , and $\sim 1600 \text{ cm}^{-1}$, derived from vibrational modes in adenine, guanine, and a combined peak from the overlapping vibrational modes of guanine and the AAAs. Peaks were assigned to specific vibrations of each molecule according to the literature (Wen et al., 1997): the dominant marker peak for adenine is assigned to a coupled vibration of C8N7 stretching, and C8H bending is apparent between 1220 and 1350 cm^{-1} , with an observed peak center at 1305 cm^{-1} . The characteristic $\sim 1450 \text{ cm}^{-1}$ guanine peak is assigned to C8H bending/N7C8 stretching and is the sharpest, best-defined peak in the majority of spectra, centered at 1457 cm^{-1} . The $1490\text{--}1650 \text{ cm}^{-1}$ region is characterized by two peaks: a composite peak at $1540\text{--}1555 \text{ cm}^{-1}$ due to overlapping modes of N3C4 stretching and C5C4/C4N3



stretching in guanine, and a $\sim 1600\text{ cm}^{-1}$ peak due to in-plane ring stretching of the AAAs. The broad, asymmetric tail of the 1600 cm^{-1} peak toward 1700 cm^{-1} indicates the presence of further vibrational modes that cannot be clearly defined due to their overlap, but are reported to include contributions from thymine, cytosine and guanine (Wen and Thomas, 1998). Finally, there is a small peak at $\sim 1750\text{ cm}^{-1}$ that is not assigned to any molecular vibration and is attributed to a secondary NeCu laser emission line reflected off the Al wafer used as a sample substrate. Secondary peaks apparent in the spectrum can be ascribed to vibrational modes in cytosine, thymine, and uracil. A peak centered at $\sim 1175\text{ cm}^{-1}$ is due to an undefined vibrational mode in thymine, $\sim 1200\text{ cm}^{-1}$ to a coupled C5H bending and N3C4 stretching vibration in uracil and $\sim 1510\text{ cm}^{-1}$ to N3C4 stretching in cytosine. A shoulder around 1350 cm^{-1} on the main adenine peak is due to C2H and N9C8 stretching in cytosine. A minor peak just below 1000 cm^{-1} is due to symmetric ring stretching in phenylalanine and tyrosine.

Deconvolution

The deconvolution of the cell spectrum and of the artificial mixture spectrum were done with three different sets of

TABLE 1 | Dominant peak positions (in cm^{-1}) of the molecular standards and their assignments to vibrations of specific atomic bonds, categorized as either stretching (s), or bending (b) modes (Wen et al., 1997; Wen and Thomas, 1998).

Nucleobase	Simple NB	dNTP	DNA/RNA	Assignment
A	1291	1322	1314	C8H (b), C8N7 (s)
C	1512	1523	1508	N3C4 (s)
G	1440	1474	1462	C8H (b), N9C8 (s)
T	1647	1651	1634	C4 = O (s), C5C7 (s)
U	1210	1214	1203	C5H (b), N3C4 (s)
Amino acids	Phe	Trp	Tyr	
	1595	1610	1591	Benzene (s)

standards, reflecting the increasing structural complexity of the nucleobases, with the AAAs represented by the same Phe, Trp, and Tyr spectra in all cases. The relative integrated intensities of each component in each fit were obtained based on their respective fitting coefficients, with uncertainties derived from the coefficient's standard error and the standard deviation of the spectrum. In all cases, the five nucleic acids represented the majority of Raman intensity across the cellular spectrum. The overall goodness of each fit was expressed numerically by its chi-squared (χ^2) value, constructed from the sum of the square of the fit's residuals. The absolute values of χ^2 obtained were large, due to the fit having 253 degrees of freedom, with the poorest fit having an χ^2 of 320 and the best fit having an χ^2 of 49.

DISCUSSION

Biological Patterns as Revealed by DUV Raman Spectroscopy

The three AAAs Phe, Trp, and Tyr exhibit DUV Raman spectra with unique patterns of minor vibrational modes across the 800–1800 cm^{-1} range (see **Figure 2**), but share a common dominant mode at 1600 cm^{-1} that makes them difficult to distinguish in mixtures but can be treated as indicative of the presence of AAAs. It is of importance to note that AAAs have not been identified in extraterrestrial material. Tryptophan, specifically, has the highest free energy of formation (Amend and Shock, 1998), as such the observation of AAAs is potentially indicative of biosynthetic pathways. In contrast to the AAAs, the five nucleobases A, C, G, T, and U exhibit unique Raman spectra dominated by different vibrational modes at distinct frequencies (**Figure 2**). The pattern of peak intensities was not consistent between molecules, beyond being assigned to aromatic modes, due to the dependence of Raman scattering cross-sections and vibrational frequencies on the exact structure of each molecule. In general, the most intense peaks of the five nucleobases are sufficiently separated in frequency that they can be considered specific markers for each base. Furthermore, individual spectra are sensitive to increasing structural complexity; there are significant spectral changes between the nucleobases and their respective nucleotides, consisting primarily of shifts in peak position and the suppression of various minor peaks. Changes in vibrational frequency, such as the $\sim 31 \text{ cm}^{-1}$ shift in the major C8H/C8N7 mode between adenine and dATP, can be attributed to a redistribution of π -electron density on the adenine's aromatic purine moiety upon addition of the deoxyribose triphosphate unit. Changes in the relative intensity of various modes, to the point of peaks disappearing for cytosine and guanine, is likely due to vibrational coupling between in-plane vibrations of the aromatic and ribose moieties, suppressing some Raman peaks while enhancing others. This coupling has been reported previously based on the sensitivity of DUV resonant aromatic vibrational frequencies to selective deuteration of the ribose unit (Toyama et al., 1993). The precise impact varies from nucleobase to nucleobase, due to the different structures of their aromatic moieties leading to different degrees of coupling, with

some exhibiting more significant shifts in frequency or fewer suppressed modes, e.g., adenine and cytosine respectively.

Further increase in structural complexity from nucleotides to DNA is expected to have a small but definite impact on the Raman spectra of the various nucleobases, due to π - π stacking between neighboring nucleobases altering the electron density on their aromatic moieties and therefore changing their vibrational properties. By using relatively short oligomers of single-stranded DNA, each nucleobase can be assessed individually, without hydrogen bonding interactions caused by Watson–Crick pairing or larger scale structure. When compared in **Figure 2**, we observe that the majority of each nucleotide's Raman peaks appear in the DNA spectrum as well. Minor changes were observed, e.g., a slight broadening of the bimodal $\sim 1300 \text{ cm}^{-1}$ guanine peak, or a small ($\sim 8 \text{ cm}^{-1}$) down-shift of the thymine peaks at ~ 1200 and 1300 cm^{-1} . There were also more significant effects: the relative intensity of the $\sim 1550 \text{ cm}^{-1}$ adenine peak increases between dATP and the DNA-A 10-mer; the spectrum of the DNA-C 10-mer is dominated by a mode at $\sim 1574 \text{ cm}^{-1}$ that was a hidden peak in cytosine and dCTP; and the ~ 1600 and $\sim 1650 \text{ cm}^{-1}$ modes of uracil both increase in relative intensity. We attribute all of these spectral changes to the π - π stacking of neighboring nucleobases within DNA, which are extremely unlikely to occur between free nucleotides in solution at the concentrations being considered. The effect of close interactions between aromatic π systems is well-known, producing easily measurable, if difficult to predict, changes in vibrational frequency and relative peak intensities under resonant excitation (Milani et al., 2007). It is not immediately clear if these minor changes make the DNA/RNA standards more representative of the cell spectrum, and a more thorough method of comparison is required.

The DUV Raman spectrum of *E. coli* consists of several peaks all of which can be attributed to vibrational modes in at least one of molecular standards used based on comparisons with the dominant vibrational modes of those molecules. The overall Raman spectrum of the cell may be considered a composite of the Raman spectra of all the components of the cell, weighted by the number of each molecule and their relative Raman cross-sections. Two well-defined peaks at ~ 1310 and $\sim 1470 \text{ cm}^{-1}$ are consistent with vibrations in adenine and guanine respectively, with several more minor peaks and overlapping modes assigned in **Figure 2**. The Raman spectrum can be modulated by reduction (hypochromism) or increase (hyperchromism) in Raman cross-section due to intermolecular interactions, such as π - π base stacking and Watson–Crick pairing in DNA, as has been reported for both nucleobase and amino acid peaks in DUV Raman spectra of biological matter (Wen et al., 1997). Specific hypochromism in nucleic acids observed as a decrease in Raman scattering of DNA bases when stacked in DNA has been previously reported (Bolton and Weiss, 1962). Identification of the constituent DUV resonant molecules is still possible despite these phenomena; however, as acknowledged earlier, even the most diagnostic peaks are complicated by significant degrees of overlap with the minor vibrational modes of the other components, and assigning peaks to only the apparent dominant modes without accounting for composition may lead to incorrect interpretation of spectra.

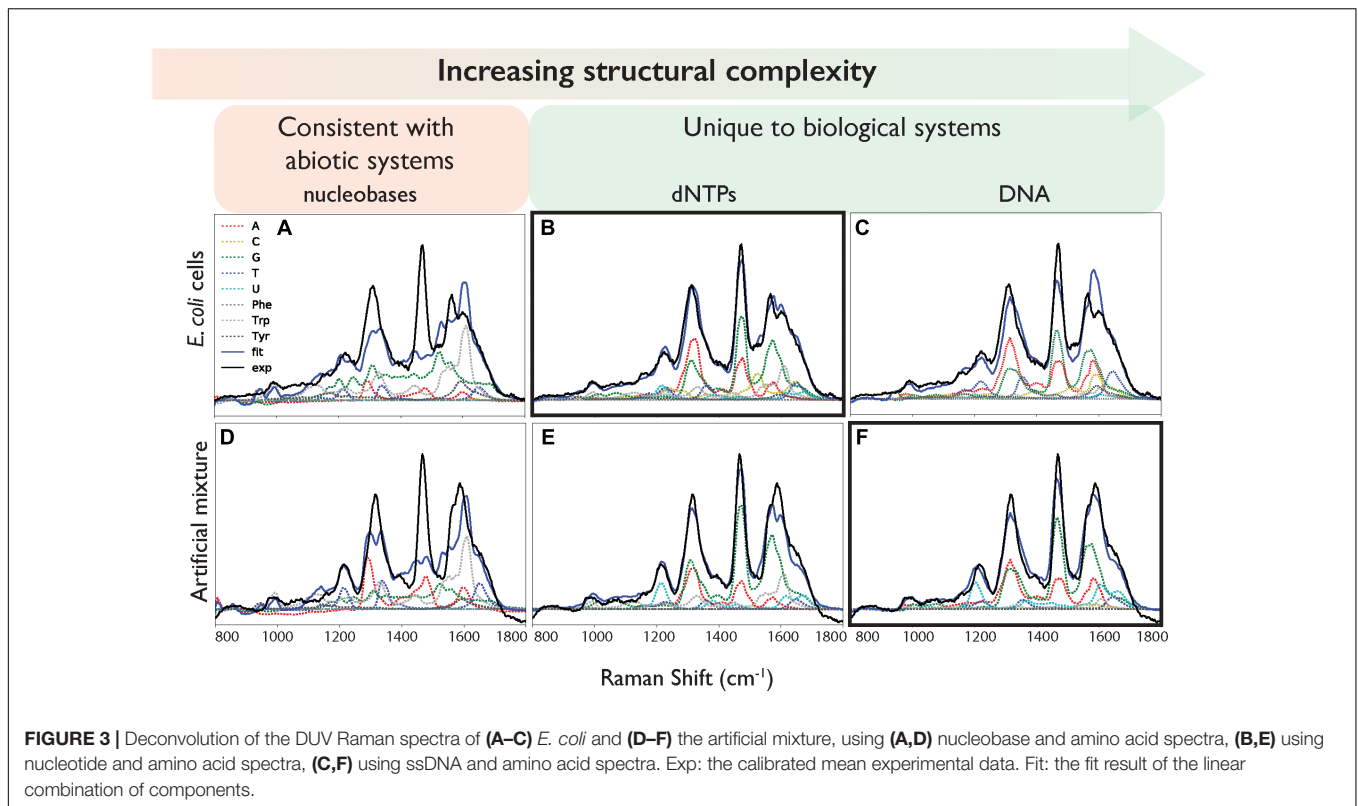


TABLE 2 | Intensities for each component of the nucleic acid + aromatic amino acid (AAA) deconvolutions of *E. coli* and the artificial mixture, fitted using standards of varying degrees of nucleic acid structural complexity.

Fit:	Adenine	Cytosine	Guanine	Thymine	Uracil	Phe	Trp	Tyr
Cell								
Nucleobase + AAAs ($\chi^2 = 320$)	0.10 ± 0.04	0.09 ± 0.06	0.44 ± 0.07	0.08 ± 0.04	0.06 ± 0.02	0.08 ± 0.03	0.00 ± 0.00	0.03 ± 0.02
Nucleotide + AAAs ($\chi^2 = 49$)	0.23 ± 0.05	0.14 ± 0.03	0.34 ± 0.05	0.06 ± 0.01	0.06 ± 0.01	0.09 ± 0.03	0.02 ± 0.01	0.02 ± 0.01
DNA + AAAs ($\chi^2 = 140$)	0.34 ± 0.12	0.03 ± 0.01	0.32 ± 0.10	0.11 ± 0.03	0.11 ± 0.04	0.02 ± 0.02	0.01 ± 0.00	0.00 ± 0.00
Artificial mixture								
Nucleobase + AAAs ($\chi^2 = 217$)	0.15 ± 0.07	0.00 ± 0.00	0.22 ± 0.04	0.12 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.34 ± 0.11	0.06 ± 0.02
Nucleotide + AAAs ($\chi^2 = 62$)	0.18 ± 0.04	0.00 ± 0.00	0.50 ± 0.08	0.06 ± 0.01	0.12 ± 0.02	0.00 ± 0.00	0.17 ± 0.06	0.00 ± 0.00
DNA + AAAs ($\chi^2 = 54$)	0.27 ± 0.10	0.03 ± 0.01	0.44 ± 0.14	0.07 ± 0.02	0.16 ± 0.05	0.00 ± 0.00	0.05 ± 0.02	0.00 ± 0.00

Intensities are integrated over 800–1800 cm^{-1} and relative to total intensity of the cell. Relative goodness of each fit is given by the value of chi-squared, i.e., the sum of the square of the residuals.

Spectral deconvolution allows for consistent identification of the components present within a complex, mixed spectrum dominated by several aromatic molecules. Deconvolution must be informed by prior measurements of standard spectra for a set of known molecules that are representative of the total system. With three sets of molecular standards of increasing structural complexity, we deconvolute the spectrum of *E. coli* in order to establish the composition of the cellular spectrum, and the minimum degree of complexity required to distinguish the cell from its most simplistic parts. Deconvolution was done by scalable linear combination of the appropriate molecular standard spectra (including the five nucleobases and three AAAs; expressed in Eq. 1), varying only their respective fitting coefficients (i.e., their relative intensities) to

minimize the residual between I_{fit} and I_{cell} , with the results presented in **Figure 3**.

$$I_{\text{fit}} = a * \text{Adenine} + c * \text{Cytosine} + g * \text{Guanine} + t * \text{Thymine} \\ + u * \text{Uracil} + f * \text{Phenylalanine} + w * \text{Tryptophan} \\ + y * \text{Tyrosine} \quad (1)$$

The “goodness of fit” defined by the sum of the square of the residual, varied depending on the structural complexity of the nucleobase components (**Figure 3** and **Table 2**). When the simple nucleobases are used as components for deconvolution, the main adenine mode at $\sim 1310 \text{ cm}^{-1}$ is adequately described, but the rest of the spectrum is fit rather poorly: failing to account

for major modes including the guanine peaks at ~ 1469 and ~ 1572 cm^{-1} ; overestimating the intensity of the amino acid peak at ~ 1600 cm^{-1} ; and underestimating the minor contribution of weak modes around 900 – 1100 cm^{-1} (see **Figure 3A**). The failure to account for the major guanine peak at ~ 1469 cm^{-1} can be attributed to the lack of any equivalent mode in the standard guanine spectrum, with the closest guanine mode some ~ 29 cm^{-1} lower in frequency and much less well-defined. The over-estimation of intensity at ~ 1600 cm^{-1} is mostly due to a contribution from adenine in that region, even though that peak is typically associated with the amino acids. It was not possible to comment on the relative contributions of the three AAAs as the fitted coefficients of Phe, Trp, and Tyr were significantly cross-correlated (**Supplementary Table S2**) such that their intensities cannot be considered separately and must be treated as a single, total 'aromatic amino acid' component that was significantly overestimated in the nucleobase fit.

When the cell deconvolution was run using the nucleotides as components, the fit immediately improved: the adenine/guanine peaks at ~ 1300 , ~ 1450 , and ~ 1550 cm^{-1} are adequately represented by the adenine and guanine components in both peak intensity and shape (**Figure 3B**). The remaining residuals are more evenly distributed and while there are still the distinctively shaped residuals produced by deviations in vibrational frequency for certain peaks (**Supplementary Figure S5**), there is clearly much better correlation than with the nucleobases. We can conclude that the nucleotides are significantly more representative of the vibrational properties of their respective equivalents within the cell itself, compared to the nucleobases. The remaining minor deviations may be the result of the slight vibrational changes observed between nucleotides and the more structurally complex DNA bases, or evidence for the collective signal from non-DUV resonant components within the cell.

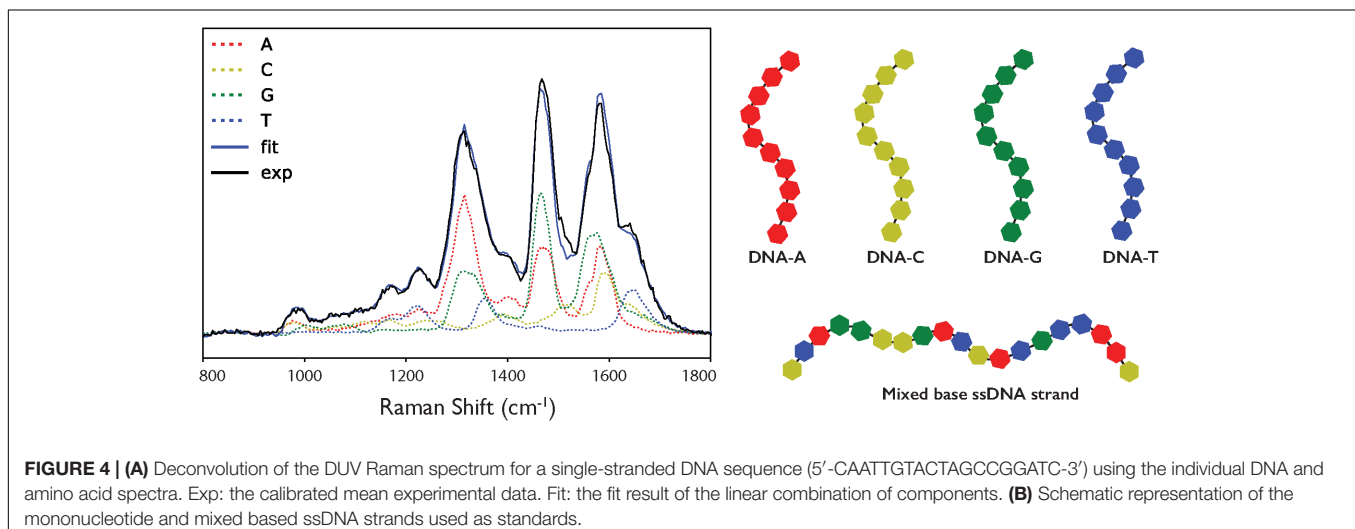
When the cell deconvolution was run using the DNA base standards, the result was a poorer fit than with the nucleotide standards ($\chi^2 = 140$ compared to 49). As shown by **Figure 3C**, there is a large error at ~ 1550 cm^{-1} that was not present in

the nucleotide fit. This corresponds to the C4N3/C5C4 mode of adenine, which exhibited a much larger relative intensity in the DNA-A 10-mer vs. dATP, along with a smaller contribution from the additional mode of the DNA-C 10-mer that was not present in dCTP. Further deviation occurs around the dominant adenine and guanine peaks of the cellular spectrum, at ~ 1300 and 1450 cm^{-1} , due to the broader modes of the DNA-A and DNA-G 10-mers spectra vs. dATP and dGTP. The deviation cannot be easily ascribed to the necessarily artificial single-base sequences of these standards: when a mixed DNA 19-mer containing all four bases (with a 5:5:4:5 ratio of A:C:G:T) was measured, it had a spectrum that was a clearly linear combination of the relevant DNA base standards (**Figure 4**). This shows that there are no spectral shifts or changes resulting from neighbor-neighbor interactions and secondary structure, though we cannot rule out the possible effects of tertiary structure as the mixed DNA sequence was specially chosen to limit pairing interactions.

Fundamentally, the cell spectrum can be adequately described by nucleotide and AAA spectra. The nucleotides must be sufficiently complex to represent the nucleic acid component, in a manner that the simpler nucleobases were not. It is of note that nucleotides have not been observed to occur in abiotic systems in contrast to nucleobases. The detection of these component molecules in combination, reflecting the complexity in chemical structure and composition of the cell, can therefore be considered a meaningful biosignature detectable by DUV Raman spectroscopy.

Macromolecular Composition of the Cell

Deconvolution of the cellular Raman spectrum may provide a first approximation of cellular composition, modulated by differences in Raman cross-section between detectable components. The fact that the dNTP standards provide a better fit than the DNA standards is surprising, considering that nucleic acids account for the majority of nucleobases in the cell (Neidhardt et al., 1990). In order to investigate this phenomenon, we first must approximate the macromolecular



composition of the cell and those components that are detectable by DUV Raman.

While it is well-established that the composition of the cell varies over time (Pahlow et al., 2015; Hlaing et al., 2016), the values given here are based on average *E. coli* cells during exponential growth and thus should be an acceptable first approximation of the composition of the cells that were measured using Raman spectroscopy in this study. The overall composition of an average *E. coli* cell was calculated based on adaptations of the values for the macromolecular composition of *E. coli* by Milo et al. (2010) from Neidhardt et al. (1990) and others, to reflect uncertainties for a cell growing exponentially at 37°C in aerobically balanced glucose minimal media with a doubling time of 40 min (Figure 1 and Supplementary Table S3). We expect that only the aromatic units will be resonantly enhanced by DUV excitation, and we approximate the aromatic-containing components with the eight molecules that are known to contribute to the DUV Raman spectrum. Bennett et al. (2009) used mass spectrometry to quantify 103 metabolites within the cell and we approximate the DUV resonant fraction as those molecules that contain at least one of the eight aromatic moieties (see Supplementary Table S3). Given a total wet mass of ~1000 fg and a volume of ~0.9 μm³ per cell (Milo et al., 2010), we calculated the number of DUV resonant residues present in each group. Protein accounts for 165 fg per cell, with Phe, Trp, and Tyr accounting for 7% of residues (Kozłowski, 2017) equating to 65 million aromatic residues per cell. Assuming a rapidly dividing cell contains on average 2 genomes accounting for nested chromosomal replication, DNA comprises ~9 fg, and RNA ~60 fg (Milo et al., 2010), and based on known A/C/G/T and A/C/G/U mole ratios, the nucleic acids contain 16.6 and 106 million nucleobases per cell respectively (Nierlich, 1972; Blattner et al., 1997). Nucleobases in the metabolome total 36 million per cell, primarily in molecules containing adenine and uracil (Table 3; Bennett et al., 2009). The prevalence of RNA over DNA is expected for a cell undergoing rapid growth (Bremer and Dennis, 2008), and we find that 77% of all nucleobases within the cell are sequestered in nucleic acids, with only 23% represented as free metabolites. The AAAs are a very minor fraction of the metabolome, totaling less than 40,000 molecules per cell, meaning that 99.4% of all AAA residues occur in proteins. In total, we estimate there are roughly 224 million DUV resonant aromatic molecules per cell, 71% of which are nucleotides and 29% are amino acids, accounting for a total of 12% of cell mass. With water accounting for ~64% of cellular mass, the remaining ~21% is expected to consist of molecules that lack a considerable DUV Raman cross-section, their collective Raman scattering and interactions with DUV resonant molecules may still contribute substantially to measured Raman intensity, and it is important to consider what effect they may have on the overall spectrum.

The Cell Approximated by the Sum of Its Parts

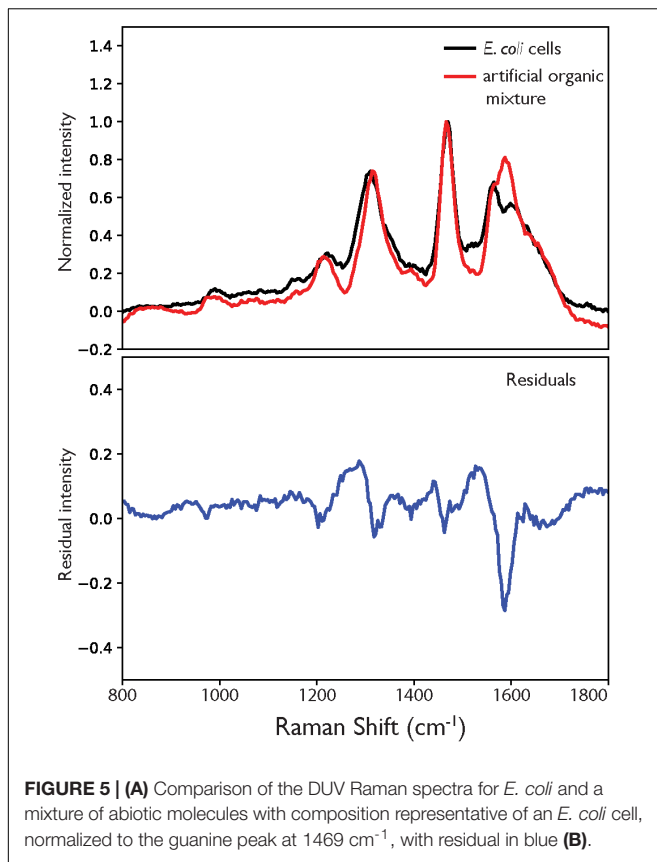
With the evidence available so far, it appears that there is clearly some effect at play that distorts Raman intensities beyond what would be expected solely from composition, leading to

TABLE 3 | Approximate composition of DUV resonant macromolecular components in an average cell of *E. coli* during exponential growth with a doubling time of 40 min, based on a cellular volume of 0.9 μm³ (Neidhardt et al., 1990).

Macromolecule	% cell mass	Weight per cell (fg)	% mol	Aromatic units per cell (x10 ⁶)
Protein¹	16.5	165		65.2
Phe			3.46	31.8
Trp			1.06	9.74
Tyr			2.58	23.6
RNA²	6	60		106
A			24.8	26.2
C			21.8	23.0
G			32.4	34.2
U			21.0	22.2
DNA³	0.9	9		16.6
A			24.5	4.07
C			24.6	4.08
G			26.2	4.35
T			24.7	4.10
Metabolites⁴	7	70	Concentration (mM)	36.4
A			16.0	10.7
C			2.75	1.84
G			5.85	3.91
T			4.98	3.33
U			22.7	15.2
Phe			2.00	0.012
Trp			1.33	0.0008
Tyr			3.22	0.019
Other	5.7	57		
Water	63.9	639		
Total	100	1000		224

(1) Calculated based on 3×10^6 proteins per cell (Neidhardt et al., 1990), cumulative weight derived from Neidhardt et al. (1990), amino acid abundance from the proteome-pl database (Kozłowski, 2017), and the molecular mass of each amino acid. (2) Calculated using the cumulative weight derived from Neidhardt et al. (1990), the RNA composition reported by Nierlich (1972) and the individual masses of the ribonucleotide monophosphates. (3) Calculated using the cumulative weight derived from Neidhardt et al. (1990), the DNA composition from Blattner et al. (1997) and the individual masses of the deoxyribonucleotide monophosphates. (4) Calculated from concentrations of 103 free metabolites given by Bennett et al. (2009), grouped according to constituent aromatic unit. Bold values indicate totals for each macromolecule class.

a spectrum dominated by a relatively minor fraction of the population of DUV resonant molecules within the cell. To assess if this variation is a function of biological structure, we can ask if the cell is truly more than the sum of its parts, i.e., is the observed spectrum unique to the cell and distinguishable from an artificial mixture containing the same components. If true, then the residual of the cell vs. the components may be treated as a biosignature in of itself, being a summation of the collective structure and complexity within the cell that, by definition, does not exist in the artificial mixture. To this end, the DUV resonant components were mixed according to their relative concentrations within the cell, derived from Table 3 (see Supplementary Table S4 for a detailed breakdown of approximations), and a Raman spectrum obtained of the mixture.



As shown in **Figure 5**, the artificial mixture exhibits a similar spectrum to that of the cell, recreating the positions and relative intensities of the major peaks with reasonable accuracy, demonstrating that the mixture has effectively recreated the relative composition (and spectral contributions) of the cell in terms of its most DUV resonant components. The largest single deviation is the additional peak at $\sim 1590 \text{ cm}^{-1}$, which at first appears to relate to the AAA component but does not perfectly align with the dominant amino acid mode at 1600 cm^{-1} . When the spectrum from the artificial mixture was deconvoluted, the best fit was obtained using DNA standards (see **Figures 3D–F** and **Supplementary Figure S6**) with the extra peak described not by any of the amino acids but by the DNA-A 10-mer, namely the bimodal vibration at $\sim 1583 \text{ cm}^{-1}$. Aside from the erroneous extra peak, the difference between cellular and abiotic spectra consisted mainly of additional background signal across the organic fingerprint region ($800\text{--}1800 \text{ cm}^{-1}$) that was apparent in the cell spectrum but not in the mixture, and accounts for 16% of total intensity across the range in question. This background cannot be attributed to molecular fluorescence, as the frequencies of Raman-scattered light under DUV excitation are significantly higher than that of photo-luminescence, nor is it an artifact of sample configuration as both spectra were measured of samples in the same conditions on the same substrate material, which does not contribute any signal in this range.

It is clear that there are distinctive and measurable spectral features that distinguish a cell from a simple mixture of its

most DUV resonant components. There are three possible explanations for why the artificial mixture deviates from the cell: (1) the cumulative contribution of all the non-DUV resonant components of the cell that were not included, (2) the lack of tertiary structure for the nucleic acid components, and (3) the free metabolites are not easily represented by their equivalent dNTP/NTP nucleotide.

There is low intensity Raman scattering across the $800\text{--}1800 \text{ cm}^{-1}$ range observed for the cell that is not apparent in the artificial mixture. This could not be attributed to fluorescence or other background effects, and may instead represent the total contribution from all non-resonant components that were not included in the mixture, but comprise approximately two thirds of the non-water mass of the cell. Considering the variety of species that group includes, such as non-AAAs, lipids and sugars, among others, the cumulative Raman scattering from their diverse vibrational modes should extend across the entire organic fingerprint region, with few distinguishable peaks. This is consistent with what we observe, as the residual (**Figure 5B**) exhibits no clearly defined peaks that are not assigned to a vibrational mode present in the DNA standards.

The artificial mixture was best fitted with the DNA standards (see **Supplementary Figure S6** for residuals and residual distributions), while the cell was best fitted using the nucleotide standards. In the artificial mixture, nucleic acids were represented by a representative proportional mixture of 10-unit oligomers of each base while in the cell these molecules are often present in complex three-dimensional conformations. We suspect that this is due to differences in the relative Raman cross-sections of the nucleobases in the free molecule vs. the macromolecule: that either the free nucleotides produce stronger Raman scattering per aromatic unit than the same nucleotides in DNA/RNA, or that tertiary structure diminishes the Raman cross-section of the aromatic unit in the nucleic acid, reducing its effective intensity consistent with previous studies (**Supplementary Figure S7**; Bolton and Weiss, 1962). This may in part be due to chromosomal and RNA packing: over 80% of total RNA is tightly folded into ribosomes (Bremer and Dennis, 2008). We have noted that differences in Raman cross-section can lead to two standards giving different apparent intensities even at the same concentration: this is illustrated by a DNA-mix 19-mer, which has a known A, C, G, T molar composition of 26, 26, 21, and 26% but integrated intensities from fitting were 37, 17, 33, and 12% respectively, indicating that per molecule the purines produce greater Raman scattering than the pyrimidines. It is probable that the introduction of tertiary structure, where every nucleobase is surrounded by other aromatic molecules and proteins, diminishes the Raman cross-section of the aromatic ring such that the nucleic acids contribute less intensity than expected given their proportion within the cell. However, it does empirically demonstrate that the DUV Raman spectrum of the cell is sensitive to this larger-scale structure that may distinguish it from its mere components.

With further work, deconvoluting the cellular spectrum into its components could be a potentially useful tool for studying terrestrial cellular activity as well as detecting biosignatures. Such analysis would require a thorough understanding of the

Raman activities of the component molecules, based on the collection of calibration curves to correlate Raman intensities to concentrations. With that information, it should be possible to derive the Voronoi plot of cellular composition in **Figure 1** from that of the Raman deconvolution. Providing the ability to spectroscopically measure changes in the composition of the cell, based on changes in the deconvolution of the Raman spectrum, would allow investigation into RNA expression and protein production as a function of cell growth rate and species differentiation based on comparisons of genome GC content and differential protein expression. However, obtaining the relevant calibration curves is not a trivial process for such a complex system as an entire cell: additional work must be done to establish the obfuscating factors that may further modulate intensities for these components in this environment, including component-component interactions, before we can employ quantitative DUV Raman spectroscopy as a tool for studying microbiology at the cellular level. While the propinquitous detection of complex aromatic molecules not expected to exist together at the observed concentrations spontaneously, constitutes a potential DUV Raman biosignature, the apparent sensitivity of the cellular spectrum to tertiary structure provides direct evidence of larger-scale structure and complexity that cannot exist in abiotic systems, strengthening the interpretation of biogenicity. It seems that, from a spectral perspective, the cell is indeed more than the sum of its parts.

Deep UV Raman spectroscopy has been selected as an instrument on the Mars 2020 rover in part due to its sensitivity and specificity for the detection of aromatic organic molecules (Beegle et al., 2015). We do not predict here the specific aromatic organics that may be detected on Mars nor do we attempt to approximate the mineral matrix in which they may be preserved. Rather we demonstrate the importance of molecular complexity to the interpretation of DUV Raman spectra of aromatic organic molecules fundamental to terrestrial life. The mechanisms that led to the structural organization of pre-biotic organic compounds into complex assemblages conferring the functions of energy transduction, replication, and information storage are currently unknown. It can be argued that the emergence of specifically structured functional complexity gave rise to molecular assemblages capable of performing the functions that we associate with life. At a basic level, these processes harness free energy to predictably and systematically produce specific outcomes that without facilitation or catalysis by living systems would not be predicted to occur. A defining characteristic of life is to produce low probability outcomes reflected in characteristic enrichments of specific organic molecules (e.g., McKay, 2004; Des Marais et al., 2008; Des Marais, 2013; Mustard et al., 2013). For example Fischer-Tropsch-Type synthesis is hypothesized to account for α amino acids in carbonaceous chondrites leading to a thermodynamically driven distribution characterized by a decrease in abundance with increasing carbon chain length (Donnelly and Satterfield, 1989) as recorded in the organic inventory of amino acids in extraterrestrial samples (e.g., Pizzarello et al., 2006) whereas biogenic processes enrich thermodynamically costly, structurally complex, molecules such as \sim C17–C31 alkanes and aromatic rings (Lovelock, 1965; Scalan

and Smith, 1970; Amend and Shock, 1998; Kuhn et al., 2010). Biosignatures reflect the persistence of these low probability outcomes, recording the mechanisms of energy capture and transduction into the unlikely emergence of complexity. It is established that simply identifying aromatic molecules does not constitute evidence of life. We show that molecular complexity is significant and the DUV Raman spectra of those molecules can be used to define a threshold for aromatic organic molecules uniquely associated with life. The premise of astrobiology relies on the assumption that the activity of living organisms will result in the formation of geochemical, molecular, and/or structural patterns that are both recognizable and distinguishable from the environment in which they formed and that their presence is statistically unachievable within a purely abiotic system (Cady et al., 2003; Des Marais et al., 2008; Summons et al., 2008; Mustard et al., 2013). While it is not expected that life beyond Earth would necessarily be comprised of the same subset of organic molecules, specificity and patterns indicative of biological complexity is thought to be a universal attribute of life (e.g., Summons et al., 2008). When searching for biosignatures beyond Earth it is imperative that these universal traits are interrogated (Des Marais, 2013). Here, we use DUV Raman spectroscopy to evaluate the increasing complexity of biomolecules and the ability of these individual components to deconvolve cellular spectra to illustrate the role of emergent molecular complexity in a cell as a fundamental component in biosignature detection.

CONCLUSION

The deconvolution of the cellular *E. coli* Raman spectrum using molecular standards of increasing complexity has provided several valuable insights into the detection of biosignatures using DUV Raman spectroscopy. Firstly, this technique is capable of distinguishing between a mixture of aromatic molecules and a complex cell built from structured components, as demonstrated by the difference between the 'best' fit spectrum using simple nucleobases vs. nucleotides. This is important because although the simple nucleobases have been detected in abiotic environments such as meteorites and molecular nebulae, they do not constitute a biosignature in of themselves. Secondly, we have confirmed that we can differentiate a cell from DNA based on its spectra and that the resulting spectra cannot be explained simply by the spectral contribution of AAAs, but rather is primarily due to the intracellular pool of free nucleotides combined with the hypochromatism of nucleobases when stacked in nucleic acids. Third and finally, we have shown that nucleotides are of sufficient structural complexity to adequately describe cellular spectra, and that obtaining standard spectra of more complex molecules may not be necessary to identify biosignatures using Raman.

It is evident that an *E. coli* cell as described by its DUV Raman spectrum is more than the sum of its DUV resonant components. While the characteristic peaks in the cellular spectrum may be assigned by the dominant molecular vibrations of the DUV resonant components as a first approximation, it is clear that a specific combination of these components at a sufficient

level of molecular complexity is required to adequately describe the cellular spectra by means of deconvolution. The observed cellular spectrum is a function of (1) the combined relative Raman cross-section of each component and; (2) the expression of that component within the cell. The former enables the selective investigation of a smaller, but still representative, subset of aromatic molecules by using DUV excitation. The cellular expression of these components is a function of billions of years of evolution selectively accumulating organic molecules, transferring a level of functional complexity reflected in a unique association of specific molecules not expected to have occurred by chance in a purely abiotic system. This study demonstrates the ability of DUV Raman spectroscopy to interrogate the nature of biological complexity and differentiate an organic signal from a definitively biological one.

AUTHOR CONTRIBUTIONS

HS designed the study with significant input from JH, LB, and RB. HS cultured the cells, prepared the cell samples, and collected the cellular Raman spectra. HS and JH designed the analytical pipeline and wrote the processing code. JH made the standard solutions, collected the standard Raman spectra, and finalized the Raman analyses. HS and JH wrote the manuscript with significant input from RB and LB. JA and VO contributed to discussion, feasibility, and text.

REFERENCES

- Abbey, W. J., Bhartia, R., Beegle, L. W., DeFlores, L., Paez, V., Sijapati, K., et al. (2017). Deep UV raman spectroscopy for planetary exploration: the search for in situ organics. *Icarus* 290, 201–214. doi: 10.1016/j.icarus.2017.01.039
- Amend, J. P., and Shock, E. L. (1998). Energetics of amino acid synthesis in hydrothermal ecosystems. *Science* 281, 1659–1662. doi: 10.1126/science.281.5383.1659
- Asher, S. A. (1993). UV resonance raman spectroscopy for analytical, physical, and biophysical chemistry. *Anal. Chem.* 65, 201A–210A. doi: 10.1021/ac00052a715
- Asher, S. A., and Johnson, C. R. (1984). Raman-spectroscopy of a coal liquid shows that fluorescence interference is minimized with ultraviolet excitation. *Science* 224:1984. doi: 10.1126/science.6740313
- Asher, S. A., and Murtaugh, J. L. (1988). Uv raman excitation profiles of imidazole, imidazolium, and water. *Appl. Spectrosc.* 42, 83–90. doi: 10.1366/0003702884428653
- Beegle, L., Bhartia, R., and White, M. (2015). “SHERLOC: scanning habitable environments with Raman & luminescence for organics & chemicals,” in *Proceedings of the Conference on 2015 IEEE Aerospace*, (Big Sky, MT: IEEE). doi: 10.1109/AERO.2015.7119105
- Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5, 593–599. doi: 10.1038/nchembio.186
- Bernhardt, J., Funke, S., Hecker, M., and Siebourg, J. (2009). “Visualizing gene expression data via Voronoi treemaps,” in *Proceedings of the Sixth International Symposium on Voronoi Diagrams*, (Copenhagen), 233–241. doi: 10.1109/ISVD.2009.33
- Bhartia, R., Salas, E. C., Hug, W. F., Reid, R. D., Lane, A. L., Edwards, K. J., et al. (2010). Label-Free bacterial imaging with Deep-UV-laser-induced native fluorescence. *Appl. Environ. Biol.* 76, 7231–7237. doi: 10.1128/AEM.00943-10
- Blattner, F. R., Plunkett, G. III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 5, 1453–1462. doi: 10.1126/science.277.5331.1453

FUNDING

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This work was funded by a NASA Astrobiology Institute–Life Underground (NAI-LU, NNA13AA92A) grant to JA, VO, and RB. Further support was provided by a Human Frontier Science Program postdoctoral fellowship to HS and a NASA Postdoctoral Program fellowship to JH.

ACKNOWLEDGMENTS

We thank Nicholas Laughton and Kyle Uckert for valuable assistance with Python programming and William Abbey, Evan Eshelman, Greg Wanger, Michael Malaska, Ranjani Murali, Daan Speth, and Sean Mullin for insightful conversations and discussions regarding data analysis and data visualization.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00679/full#supplementary-material>

- Bolton, H. C., and Weiss, J. J. (1962). Hypochromism in the ultra-violet absorption of nucleic acids and related structures. *Nature* 195, 666–668. doi: 10.1038/195666a0
- Bremer, H., and Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* 3, 1–49. doi: 10.1128/ecosal.5.2.3
- Britton, K. A., Dalterio, R. A., Nelson, W. H., Britt, D., and Sperry, J. F. (1988). Ultraviolet resonance raman-spectra of *Escherichia-Coli* with 222.5–251.0 Nm pulsed laser excitation. *Appl. Spectrosc.* 42, 782–788. doi: 10.1366/0003702884428923
- Burris, J., McGee, T. J., and Heaps, W. (1992). UV raman cross sections in nitrogen. *Appl. Spectrosc.* 46, 1076–1076. doi: 10.1366/0003702924124448
- Burton, A. S., Stern, J. C., Elsil, J. E., Glavin, D. P., and Dworkin, J. P. (2012). Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* 41, 5459–5472. doi: 10.1039/c2cs35109a
- Cady, S. L., Farmer, J., Grotzinger, J. P., Schopf, J. W., and Steele, A. (2003). Biosignatures and the search for life on mars. *Astrobiology* 3, 351–368. doi: 10.1089/153110703769016442
- Chadha, S., Manoharan, R., Moennelocoz, P., Nelson, W. H., Peticolas, W. L., and Sperry, J. F. (1993). Comparison of the UV resonance raman-spectra of bacteria, bacterial-cell walls, and ribosomes excited in the deep UV. *Appl. Spectrosc.* 47, 38–43.
- Cleaves, H. J. (2010). The origin of the biologically coded amino acids. *J. Theor. Biol.* 263, 490–498. doi: 10.1016/j.jtbi.2009.12.014
- Des Marais, D. J. (2010). Exploring mars for evidence of habitable environments and life. *Proc. Am. Philos. Soc.* 154, 402–421.
- Des Marais, D. J. (2013). “Planetary climate and the search for life,” in *Comparative Climatology of Terrestrial Planets*, eds S. J. Mackwell, A. A. Simon-Miller, J. W. Harder, and M. A. Bullock (Tucson, AZ: University of Arizona Press), 583–601.
- Des Marais, D. J., Nuth, J. A. III, Allamandola, L. J., Boss, A. P., Farmer, J. D., Hoehler, T. M., et al. (2008). The NASA astrobiology roadmap. *Astrobiology* 8, 715–730. doi: 10.1089/ast.2008.0819

- Donnelly, T. J., and Satterfield, C. N. (1989). Product distributions of the fischer-tropsch synthesis on precipitated iron catalysts. *Appl. Catal.* 52, 93–114. doi: 10.1016/S0166-9834(00)83375-8
- Farmer, J. D., and Des Marais, D. J. (1999). Exploring for a record of ancient Martian life. *J. Geophys. Res.* 104, 26977–26995. doi: 10.1029/1998JE000540
- Filiberto, J., Treiman, A. H., Giesting, P. A., Goodrich, C. A., and Gross, J. (2014). High-temperature chlorine-rich fluid in the martian crust: a precursor to habitability. *Earth Planet. Sci. Lett.* 401, 110–115. doi: 10.1016/j.epsl.2014.06.003
- Grotzinger, J. P., Crisp, J., Vasavada, A. R., Anderson, R. C., Baker, C. J., Barry, R., et al. (2012). Mars science laboratory mission and science investigation. *Space Sci. Rev.* 170, 5–56. doi: 10.1007/s11214-012-9892-2
- Grotzinger, J. P., Sumner, D. Y., Kah, L. C., Stack, K., Gupta, S., Edgar, L., et al. (2013). A habitable fluvio-lacustrine environment at yellowknife bay, gale crater, mars. *Science* 343:1242777. doi: 10.1126/science.1242777
- Hardy, P. M. (1985). “The protein amino acids,” in *Chemistry and Biochemistry of Amino Acids*, ed. G. Barrett (Amsterdam: Springer).
- Hays, L. E., Graham, H. V., Des Marais, D. J., Hausrath, E. M., Horgan, B., McCollom, T. M., et al. (2017). Biosignature preservation and detection in mars analog environments. *Astrobiology* 17, 363–400. doi: 10.1089/ast.2016.1627
- Higgs, P. G., and Pudritz, R. E. (2009). A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* 9, 483–490. doi: 10.1089/ast.2008.0280
- Hlaing, M. M., Dunn, M., Stoddart, P. R., and McArthur, S. L. (2016). Raman spectroscopic identification of single bacteria cells at different stages of their lifecycle. *Vibrat. Spectrosc.* 86, 81–89. doi: 10.1016/j.vibspec.2016.06.008
- Huang, W. E., Griffiths, R. I., Thompson, I. P., Bailey, M. J., and Whiteley, A. S. (2004). Raman microscopic analysis of single microbial cells. *Anal. Chem.* 76, 4452–4458. doi: 10.1021/ac049753k
- Hunt, S. (1985). “The non-protein amino acids,” in *Chemistry and Biochemistry of Amino Acids*, ed. G. Barrett (Amsterdam: Springer).
- Ianou, A. T., Coleman, T., and Asher, S. A. (2002). UV resonance raman spectroscopic detection of nitrate and nitrite in wastewater treatment processes. *Anal. Chem.* 74, 1458–1461. doi: 10.1021/ac10863q
- Jenkins, A. L., Larsen, R. A., and Williams, T. B. (2005). Characterization of amino acids using Raman spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 61, 1585–1594. doi: 10.1016/j.saa.2004.11.055
- Johnson, A. P., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Lazcano, A., and Bada, J. L. (2008). The Miller volcanic spark discharge experiment. *Science* 322:404. doi: 10.1126/science.1161527
- Jones, E., Oliphant, T., and Peterson, P. (2001). *SciPy: Open Source Scientific Tools for Python*. Available at: <http://www.scipy.org/> (accessed July 20, 2018).
- Klein, H. P., Lederberg, J., and Alexander, R. (1972). Biological experiments: the Viking Mars lander. *Icarus* 16, 139–146. doi: 10.1016/0019-1035(72)90141-8
- Kozłowski, L. P. (2017). Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.* 45, D1112–D1116. doi: 10.1093/nar/gkw978
- Kuhn, T. K., Krull, E. S., Bowater, A., Grice, K., and Gleixner, G. (2010). The occurrence of short chain n-alkanes with an even over odd predominance in higher plants and soils. *Organ. Geochem.* 41, 88–95. doi: 10.1016/j.orggeochem.2009.08.003
- Kumamoto, Y., Taguchi, A., Smith, N. I., and Kawata, S. (2011). Deep UV resonant Raman spectroscopy for photodamage characterization in cells. *Biomed. Opt. Express.* 2, 927–936. doi: 10.1364/BOE.2.000927
- Kumamoto, Y., Taguchi, A., Smith, N. I., and Kawata, S. (2012). Deep ultraviolet resonant Raman imaging of a cell. *J. Biomed. Opt.* 17, 0760011–0760014. doi: 10.1117/1.JBO.17.7.076001
- Levin, G. V., and Straat, P. A. (1977). Life on Mars? The Viking labeled release experiment. *Biosystems* 9, 165–174. doi: 10.1016/0303-2647(77)90026-0
- Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., and Milo, R. (2014). Visual account of protein investment in cellular functions. *PNAS* 111, 8488–8493. doi: 10.1073/pnas.1314810111
- Long, D. A. (1977). *Raman Spectroscopy*. London: McGraw Hill.
- Lovelock, J. E. (1965). A physical basis for life detection experiments. *Nature* 207, 568–570. doi: 10.1038/207568a0
- Martins, Z., Botta, O., Fogel, M. L., Sephton, M. A., Glavin, D. P., Watson, J. S., et al. (2008). Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett.* 270, 130–136. doi: 10.1016/j.epsl.2008.03.026
- McKay, C. P. (2004). What is life – and how do we search for it in other worlds? *PLoS Biol.* 2:e302. doi: 10.1371/journal.pbio.0020302
- McNeil, J. R., Reid, R. D., Gerstenberger, D. C., and Collins, G. J. (1978). “Ultraviolet ion lasers,” in *High-Power Lasers and Applications*, eds K. L. Kompa and H. Walther (Berlin: Springer), 89–95. doi: 10.1007/978-3-540-35942-5_14
- Meringer, M., Cleaves, H. J., and Freeland, S. J. (2013). Beyond terrestrial biology: charting the chemical universe of α -amino acid structures. *J. Chem. Inf. Model* 53, 2851–2862. doi: 10.1021/ci400209n
- Milani, A., Brambilla, L., Del Zoppo, M., and Zerbi, G. (2007). Raman dispersion and intermolecular interactions in unsubstituted thiophene oligomers. *J. Phys. Chem. B* 111, 1271–1276. doi: 10.1021/jp066933k
- Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science* 117, 528–529. doi: 10.1126/science.117.3046.528
- Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38, D750–D753. doi: 10.1093/nar/gkp889
- Muñoz Caro, G. M., Meierhenrich, U. J., Schutte, W. A., Barbier, B., Arcones Segovia, A., Rosenbauer, H., et al. (2002). Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* 416:403. doi: 10.1038/416403a
- Mustard, J. F., Adler, M., Allwood, A., Bass, D. S., Beaty, D. W., Bell, J. F., et al. (2013). *Report of the Mars 2020 Science Definition Team*. Available at http://mepag.jpl.nasa.gov/reports/MEP/Mars_2020_SDT_Report_Appendix.pdf (accessed July 26, 2018).
- Neidhardt, F. C. (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd Edn. Washington, DC: ASM Press.
- Neidhardt, F. C., Ingraham, J. L., and Schaecter, M. (1990). Physiology of the bacterial cell - A molecular approach. *Trends Genet.* 7:341. doi: 10.1016/0168-9525(91)90427-R
- Nelson, W. H., Manoharan, R., and Sperry, J. F. (1992). UV resonance raman studies of bacteria. *Appl. Spectrosc. Rev.* 27, 67–124. doi: 10.1080/05704929208018270
- Newville, M., Stensitzki, T., Allen, D. B., and Ingargiola, A. (2014). *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. Available at: <https://lmlfit.github.io/lmlfit-py/index.html> (accessed March 10, 2018).
- Nierlich, D. P. (1972). Regulation of ribonucleic acid synthesis in growing bacterial cells. II. Control over the composition of the newly made RNA. *J. Mol. Biol.* 72, 765–777. doi: 10.1016/0022-2836(72)90190-8
- Otto, A., Bernhardt, J., Meyer, H., Schaffer, M., Herbst, F.-A., Siebourg, J., et al. (2010). Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis*. *Nat. Commun.* 1:137. doi: 10.1038/ncomms1137
- Pahlow, S., Meisel, S., Cialla-May, D., Weber, K., Rösch, P., and Popp, J. (2015). Isolation and identification of bacteria by means of Raman spectroscopy. *Adv. Drug Deliv. Rev.* 89, 105–120. doi: 10.1016/j.addr.2015.04.006
- Parker, E. T., Cleaves, H. J., Callahan, M. P., Dworkin, J. P., Glavin, D. P., Lazcano, A., et al. (2011). Prebiotic synthesis of methionine and other sulfur-containing organic compounds on the primitive earth: a contemporary reassessment based on an Unpublished 1958 Stanley Miller Experiment. *Orig. Life Evol. Biosph.* 41, 201–212. doi: 10.1007/s11084-010-9228-8
- Phillips, C. B., and Pappalardo, R. T. (2014). Europa clipper mission concept: exploring jupiter’s ocean moon. *EOS Trans. Am. Geophys. Union* 95, 165–167. doi: 10.1002/2014EO200002
- Pizzarello, S., Cooper, G. F., and Flynn, G. (2006). The nature and distribution of the organic material in carbonaceous chondrites and interplanetary dust particles. *Meteorites Early Sol. Syst. II* 1, 625–651.
- Razzell-Hollis, J., Wade, J., Tsoi, W. C., Soon, Y., Durrant, J., and Kim, J.-S. (2014). Photochemical stability of high efficiency PTB7:PC70BM solar cell blends. *J. Mater. Chem. A* 2, 20189–20195. doi: 10.1039/C4TA05641H
- Scalan, E. S., and Smith, J. E. (1970). An improved measure of the odd-even predominance in the normal alkanes of sediment extracts and petroleum. *Geochim. Cosmochim. Acta* 34, 611–620. doi: 10.1016/0016-7037(70)90019-0
- Schmitt-Kopplin, P., Gabelica, Z., Gougeon, R. D., Fekete, A., Kanawati, B., Harir, M., et al. (2010). High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proc. Natl. Acad. Sci.* 107, 2763–2768. doi: 10.1073/pnas.0912157107
- Sephton, M. A. (2002). Organic compounds in carbonaceous meteorites. *Nat. Prod. Rep.* 19, 292–311. doi: 10.1039/b103775g

- Sephton, M. A., and Botta, O. (2005). Recognizing life in the solar system: guidance from meteoritic organic matter. *Int. J. Astrobiol.* 4, 269–276. doi: 10.1017/S1473550405002806
- Soffen, G. A., and Snyder, C. W. (1976). The first Viking mission to Mars. *Science* 193, 759–766. doi: 10.1126/science.193.4255.759
- Steele, A., McCubbin, F. M., and Fries, M. D. (2016). The provenance, formation, and implications of reduced carbon phases in Martian meteorites. *Meteorit. Planet. Sci.* 51, 2203–2225. doi: 10.1111/maps.12670
- Stoks, P. G., and Schwartz, A. W. (1979). Uracil in carbonaceous meteorites. *Nature* 282:709. doi: 10.1038/282709a0
- Summons, R. E., Albrecht, P., McDonald, G., and Moldowan, J. M. (2008). “Molecular biosignatures,” in *Strategies of Life Detection Space Sciences Series of ISSI*, eds O. Botta, J. Bada, J. G. Elvira, E. Javaux, F. Selsis, and R. Summons (Boston, MA: Springer), 133–159.
- Summons, R. E., Amend, J. P., Bish, D., Buick, R., Cody, G. D., Des Marais, D. J., et al. (2011). Preservation of martian organic and environmental records: final report of the mars biosignature working group. *Astrobiology* 11, 157–181. doi: 10.1089/ast.2010.0506
- Taguchi, A., Smith, N. I., Kawata, S., and Kumamoto, Y. (2011). Deep UV resonant Raman spectroscopy for photodamage characterization in cells. *Biomed. Opt. Express* 2, 927–936. doi: 10.1364/BOE.2.000927
- Tarcea, N., Harz, M., Roesch, P., Frosch, T., Schmitt, M., Thiele, H., et al. (2007). UV Raman spectroscopy - A technique for biological and mineralogical in situ planetary studies. *Spectrochim. Acta Part A* 68, 1029–1035. doi: 10.1016/j.saa.2007.06.051
- Toyama, A., Hanada, N., Ono, J., Yoshimitsu, E., and Takeuchi, H. (1999). Assignments of guanosine UV resonance Raman bands on the basis of 13C, 15N and 18O substitution effects. *J. Raman Spectrosc.* 30, 623–630. doi: 10.1002/(SICI)1097-4555(199908)30:8<623::AID-JRS407>3.0.CO;2-9
- Toyama, A., Takino, Y., Takeuchi, H., and Harada, I. (1993). Ultraviolet resonance Raman spectra of ribosyl C(1')-deuterated purine nucleosides: evidence of vibrational coupling between purine and ribose rings. *J. Am. Chem. Soc.* 115, 11092–11098. doi: 10.1021/ja00077a005
- Uckert, K., and Bhartia, B. (2019). A semi-autonomous method to detect cosmic rays in raman hyperspectral datasets. *Appl. Spectrosc.* (in press).
- Vago, J. L., Westall, F., Coates, A. J., Jaumann, R., Korabely, O., Ciarletti, V., et al. (2017). Habitability on early mars and the search for biosignatures with the ExoMars rover. *Astrobiology* 17, 471–510. doi: 10.1089/ast.2016.1533
- Wade, J., Wood, S., Collado-Fregoso, E., Heeney, M., Durrant, J., and Kim, J.-S. (2017). Impact of fullerene intercalation on structural and thermal properties of organic photovoltaic blends. *J. Phys. Chem. C* 121, 20976–20985. doi: 10.1021/acs.jpcc.7b05893
- Wen, Z. Q., Overman, S. A., and Thomas, G. J. (1997). Structure and interactions of the single-stranded DNA genome of filamentous virus fd: investigation by ultraviolet resonance Raman spectroscopy. *Biochemistry* 36, 7810–7820. doi: 10.1021/bi970342q
- Wen, Z. Q., and Thomas, G. J. (1998). UV resonance Raman spectroscopy of DNA and protein constituents of viruses: assignments and cross sections for excitations at 257, 244, 238, and 229 nm. *Biopolymers* 45, 247–256. doi: 10.1002/(SICI)1097-0282(199803)45:3<247::AID-BIP7>3.0.CO;2-R
- Williford, K. H., Farley, K. A., Stack, K. M., Allwood, A. C., Beaty, D. W., Beegle, L. W., et al. (2018). “The NASA Mars 2020 rover mission and the search for extraterrestrial life,” in *From Habitability to Life on Mars*, eds N. A. Cabrol and E. A. Grin (Amsterdam: Elsevier).
- Wong, J. T. (1975). A coevolution theory of the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* 72, 1909–1912. doi: 10.1073/pnas.72.5.1909
- Wong, J. T. (2005). Coevolution theory of the genetic code at age thirty. *BioEssays* 27, 416–425. doi: 10.1002/bies.20208
- Wood, S., Hollis, J. R., and Kim, J.-S. (2017). Raman spectroscopy as an advanced structural nanoprobe for conjugated molecular semiconductors. *J. Phys. D Appl. Phys.* 50:73001. doi: 10.1088/1361-6463/50/7/073001
- Wu, Q., Hamilton, T., Nelson, W. H., Elliott, S., Sperry, J. F., and Wu, M. (2001). UV Raman spectral intensities of *E. Coli* and other bacteria excited at 228.9, 244.0, and 248.2 nm. *Anal. Chem.* 73, 3432–3440. doi: 10.1021/ac001268b
- Zhu, G., Zhu, X., Fan, Q., and Wan, X. (2011). Raman spectra of amino acids and their aqueous solutions. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 78, 1187–1195. doi: 10.1016/j.saa.2010.12.079

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sapers, Razzell Hollis, Bhartia, Beegle, Orphan and Amend. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms. The U.S. Government retains a nonexclusive license to this work for non-commercial purposes.