Molecular screening with liquid chromatography coupled to ultra-high resolution mass spectrometry: Chromatographic methods and data treatment for application to complex organic matter in astrophysical materials

Cédric Wolters, Véronique Vuitton, Francois-Regis Orthous-Daunay, Laurène Flandinet, Chao He, Sarah E Moran, Sarah M. Hörst

Molecular screening with liquid chromatography coupled to ultra-high resolution mass spectrometry: Chromatographic methods and data treatment for application to complex organic matter in astrophysical materials

Cédric Wolters,[†] Véronique Vuitton*[,†], François-Régis Orthous-Daunay,[†] Laurène Flandinet,[†] Chao He,[‡] Sarah E. Moran[‡,§] and Sarah M. Hörst[‡]

[†]Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France

[‡]Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

[§]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA

KEYWORDS: Liquid Chromatography, Orbitrap, Exoplanet Analogues, Atmospheric Aerosols, Laboratory Simulation

ABSTRACT

Liquid and gas chromatography have been extensively applied to the analysis of extraterrestrial materials and their analogues. Methods are often highly specific and to obtain the largest possible description of a given sample, it must be analyzed over dozens of analytical methods and instruments, often requiring some heavy sample preparation.

However, because meteorites and analogue samples are usually available in relatively small quantities, their characterization requires an analytical method that can reveal their complex organic content, being their stoichiometric diversity or their isomeric variety, in one single run.

This article presents two liquid chromatography methods coupled to ultra-high resolution mass spectrometry (LC-HRMS) for the screening of organic matter in astrophysical samples. The liquid chromatography methods rely on orthogonal separations on a unique hydrophilic interaction liquid chromatography (HILIC) column using two different pH and both polarities. We developed our analytical methods using a laboratory-produced mixture of complex organics simulating the photochemical haze in a super-Earth atmosphere and processed the data with custom software.

The soluble fraction of only 0.2 mg of sample is used for analysis without further treatments or extraction steps. 1,160 peaks ($50 < m/z < 300$) have been efficiently separated, detected, and attributed to a CHNO molecular formula. Our results demonstrate the relevance of our methodology for the untargeted analysis of precious astrophysical samples returned by current and future space missions (Hayabusa2, OSIRIS-REx, MMX, MSR).

Introduction

Remote sensing is commonly used to detect molecules in extraterrestrial environments, from interstellar clouds to protoplanetary discs and planetary atmospheres[1–3]. Although these techniques only reveal the most abundant molecules, they still allow the detection of some biosignatures, as shown by the Galileo space-probe. While observing the Earth, the probe was able to quantify oxygen, methane, and a red-absorbing feature. Since all three could not be attributed to pure geological processes, together they indicated the possibility of life on the observed object[4]. While such observations of our own Earth prove that we can detect potential biosignatures on other planets, they are not sufficient to know if a planet is indeed hosting life. Bartlett and Wong[5] define the living state by four pillars: dissipation, autocatalysis, homeostasis, and learning. While the last pillar cannot be quantified by chemical analysis, the first three require a large diversity of chemical components to sustain life as we know it. Detecting and quantifying this chemical diversity is therefore one of the main goals of astrobiologically-focused remote sensing studies. However, given the limited information gained from such observations, the interpretation of potential biosignatures detected remotely can be highly ambiguous[6–8].

A more definitive method of characterizing such chemical diversity would be to sample an object and apply separative methods such as chromatography or mass spectrometry. In the Solar System, some in-situ analyses of small bodies have been performed in addition to remote sensing. For example, the COSIMA and ROSINA mass spectrometers on the Rosetta mission analyzed dust particles and volatile species from 67P/Churyumov-Gerasimenko, a Jupiter family comet[9,10]. However, due to power and volume constraints on a spacecraft, on-board performances are not comparable to what can be achieved on

Earth with state-of-the-art instruments and methods. Therefore, sample return missions have been developed for nearby objects such as Hayabusa2 and OSIRIS-REx[11,12]. Nevertheless, such samples returned to Earth by space probes are scarce and represent a very limited population of objects, while there are between 2900 and 7300 kg of meteoritic material falling to Earth each year[13]. Such a flux means that meteorites are available in reasonable quantities, enabling detailed laboratory analyses on a large variety of meteorites. Such detailed analysis allows comparison and classification to extract insights on solar system formation and evolution[14–17].

While in-situ molecular analysis provides one insight into the organic diversity of an object, such direct analyses are unfortunately not possible for outer Solar System objects and further. Our current technology already takes decades for a probe to reach our own outer Solar System, and reaching another planetary system is currently impossible. To alleviate such issues, Titan[18,19], Pluto[20,21], Triton[22], and exoplanet[23,24] atmospheric hazes, as well as the surfaces of icy bodies in the outer Solar System[25,26] have been simulated via laboratory experiments. The analogous materials generated under non-Earth-like conditions can then be characterized using many state-of-the-art analytical techniques to explore their structural diversity and to understand the evolution of organic matter on such objects.

There are multiple analytical techniques to characterize laboratory analogues, as well as meteorites and samples returned by space probes. These techniques include those also possible with remote sensing, such as spectrophotometry, which allows for direct comparison of laboratory results to observations[27,28]. More specific analytical techniques going deeper into the chemical characterization of the organic matter, beyond what is

possible with remote sensing, also exist. To provide the maximum information about a sample, the analytical chain must be strategic, with all non-destructive techniques coming prior to destructive preparations and analytical techniques. Before submitting a sample to very specific, i.e. targeted, analytical techniques, others must be applied for screening and initial classification of the diversity. For example, extracting the stoichiometric diversity of aerosol and ice analogs, and meteorites has been achieved with both Orbitrap mass spectrometers[29,30] and FTICR analyzers[16,18,31], allowing a detailed and sensitive description of the samples, with thousands of molecular formulas identified.

Nevertheless, although ultra-high resolution mass spectrometry can screen and classify molecular formulas, it cannot alone characterize and quantify molecular structures of interest. This is where combination of ultra-high resolution mass spectrometry with separation techniques such as chromatography is helpful, as the retention time of each detected component eluting from the column reveals information about the molecular structure, while the accurate mass measures determined by MS leads to molecular formulas hypotheses. Gas chromatography instruments can achieve a high chromatographic resolution due to their column length and geometry. This also allows multiple dimension gas chromatography (GCxGC) to achieve even better separations, as for the separation of amino acids enantiomeric species[32]. Such selectivity of gas chromatography is balanced by the need to vaporize the sample at relatively high temperature (<300°C), forbidding any temperature-sensitive or high boiling point molecular analysis. Some of this limitation can be mitigated by using derivatization methods, but these methods require heavy sample treatment and removal of partial to complete sample diversity. Thus, gas chromatography is perfect for quantification and

targeted structural identification of apolar and not temperature sensitive compounds. Liquid chromatography on the other hand is suited for molecular screening of almost all compounds, but lacks chromatographic resolution. This lack of resolution makes it less efficient for structural identification but nevertheless allows a comprehensive view of molecular diversity. Both techniques when coupled to mass spectrometry are complementary for structural identification and could be used for quantification. Thus, for a comprehensive view of molecular diversity, both methods should be used to cover the whole range of compounds from polar to apolar.

Liquid and gas chromatography have already been applied to the analysis of extraterrestrial materials and their analogues. Targeted analysis of analogous interstellar ice material has achieved the identification of amino acids by 2D-GC-MS[32,33] or specific HPLC-FD[33,34], of nucleobases by UPLC-SRM/MS[35], and sugars and their derivatives by GC-MS[36,37]. Similar techniques have been applied to meteorites for the detection of amino acids by HPLC-FD[15,38,39] or even nucleotides by HPLC-UV-MS[40]. However, all these techniques are highly specific and do not allow for the full understanding of a sample's wide molecular diversity simultaneously. To obtain the largest possible description of a given sample, it must be analyzed with dozens of analytical methods and instruments, with several of them requiring heavy sample preparation. However, because meteorites and analogue samples are usually available in relatively small quantities, an analytical method that can reveal their complex organic content, including both their stoichiometric diversity and their isomeric variety, in only a single run, is required. Such gas and liquid chromatography methods can be found in biology with omics-like methods. For example, metabolomics faces large variations in polarity for small molecules, and many are affected

by pH variations. Moreover, the stoichiometric composition of these compounds can be quite variable, with a large range of nitrogen and oxygen contents. Therefore, metabolomics methods use hydrophilic interaction liquid chromatography (HILIC) that often exhibits an intermediate retention between normal phase and reversed phase chromatography[41]. HILIC is characterized by several features that are advantageous over reversed phase chromatography for the analysis of polar soluble organic matter with electrospray-Orbitrap[42]:

- Good peak shape for basic solutes
- Possibility of direct injection of samples that are dissolved in a primarily organic solvent without further preparation
- Possibility of higher flow rates, because of the lower viscosity of the mobile phase and therefore enhanced mass-spectrometer sensitivity

This article presents two liquid chromatography methods coupled to high resolution mass spectrometry that aim to reveal the wider diversity possible of the soluble part of a sample by orthogonal separations using two different pHs on a unique HILIC column. Such an approach allows both molecular screening and the targeted analysis of molecules of interest for structural identification. Indeed, recent works have emphasized the necessity for orthogonal methods for the analysis of meteorites and analogous materials, either in-line[43] or with two different separations and comparisons during the data-processing step[44]. After a description of our methods and the data-processing, we present our results in two steps. First, we present a general data exploration, and second we present a classification of the assigned molecular formulas and annotation of specific

isomers. We then discuss our results and their implications for the molecular screening of complex organic matter in astrophysical environments.

Material and methods

Mass spectrometry and chromatography methods

We performed analyses with a hybrid ion trap-orbitrap mass spectrometer (LTQ Orbitrap XL™, ThermoScientific) coupled to a ultra-high performance liquid chromatography system (U3000 Liquid Chromatography, ThermoScientific) available at Institut de Planétologie et d'Astrophysique de Grenoble. The LTQ Orbitrap XL™ is a hybrid mass spectrometer that uses an ion trap -- the so-called Linear Trap Quadrupole (LTQ) -- connected to an Orbitrap cell. A C-trap is placed between the LTQ and the Orbitrap to store ions coming from the source and inject them in packets directly into the Orbitrap cell. Ion production is done using an IonMax™ electrospray (ESI) source at the interface between the chromatograph and the mass spectrometer. The liquid chromatography instrument consists of a binary high-pressure pump, an auto-sampler, and a column oven. The pump is equipped with a 350 µl mixer. The auto-sampler is equipped with a 50 µl injection loop, run on microliter pickup mode allowing for repeatable and reproducible 10 µl sample injections. The liquid chromatography column is a 150 mm x 2.1 mm SeQuant→ ZIC→ -pHILIC column (sulfobetaine functional group) equipped with the same type of guard column (Merck Millipore). Column selection is discussed in the chromatography development section. Mass spectrometry (MS) functions were controlled by LTQTune (ThermoScientific), while liquid chromatography functions were handled by Chromeleon Xpress (ThermoScientific). Acquisitions were achieved using XCalibur sequences and methods (ThermoScientific).

All analyses were performed at the maximum instrument resolution (100,000 at *m/z* 400 Da for a mass accuracy of ±2 ppm), using an Automatic Gain Control (AGC) setting of 500,000 ions storage. While coupled with liquid chromatography, the AGC accumulation time is limited to 10 ms. For accurate mass calibration, we use the commercial default LTQ-Orbitrap-XL instrument calibration mixtures: a mixture of Caffeine, Met-Arg-Phe-Ala peptide (MRFA), and Ultramark 1621 (a mixture of fluorinated phosphazine polymers) in positive mode, and a mixture of sodium dodecyl sulfate, sodium taurocholate and ultramark 1621 for negative mode. To obtain the best possible sensitivity, acquisitions are performed in SIM mode on 50 Da mass ranges, spanning from 50 to 300 Da with a 10 Da overlap between each mass range, requiring five independent analyses to acquire the entire mass range. The analysis is performed in positive and negative mode, with the following ESI source parameters, respectively: voltage of 3.5 and -3.0 kV, sheath gas at 50 and 60 (arb), aux gas at 15 and 30 (arb), capillary temperature at 300°C and voltage on the transfer tube at 35V.

Sample synthesis and preparation

We use the sample "400K, 1000x metallicity" described in section 3.2 of our previous study of several exoplanetary atmosphere analogues based on direct infusion Orbitrap mass spectrometry[45]. This sample presents a large diversity in nitrogen and oxygen compounds, with many potential amino acids, nucleotide bases, and other biologically related molecules.

The sample was synthesized at Johns Hopkins University (Maryland, USA) with the PHAZER apparatus[46]. A gaseous mixture, designed to be analogous to a 1000x solar metallicity super-Earth exoplanet atmosphere in thermodynamic equilibrium at 400 K, is comprised of 56% H_2O, 11% CH_4, 10% CO_2, 6.4% N_2, 1.9% H_2, and 14.7% He. This

mixture kept at 400 K is passed through an AC plasma, for 72 hours. Such conditions lead to the creation of particles analogous to atmospheric haze deposited on the wall and at the bottom of the reaction chamber. The particles are then collected in an oxygen-free, water-free glove box under N_2 atmosphere and can then be subjected to any analytical technique—more details about sample synthesis can be found in Hörst et al.[24].

The sample was prepared for chromatography and mass spectrometry analysis by adding approximately 1 mg of powder to 1 mL of pure acetonitrile (UPLC/MS grade, Carlo Erba) into a vial. The mixture is then shaken for at least 20 minutes and centrifuged at 10,000 rpm for 10 minutes to separate the remaining insoluble material from the liquid phase. The liquid phase is then recovered and used without further treatment for the analysis. This single preparation is enough for all the chromatographic analyses that are shown in this article.

Chromatography method developments

As we are interested in organic molecules, particularly in biology-related structures, we adapted methods developed by Creek et al.[47] that use the HILIC column with sulfobetaine functional groups to drive the separation of such molecules. Most biologically related molecules present acido-basic behavior; we therefore use one basic pH method and one acidic to benefit from this property. From Creek et al.'s work[47], such methods present a good orthogonal behavior, confirming the utility of both methods for our sample characterization. Method implementation and adaptation have been optimized by evaluating chromatographic resolution, column selectivity, and peak shapes for each method adjustment in order to select the best possible instrument settings. This is achieved by using a known mix of 7 amino acids and nucleobase compounds (Uracil,

$C_4H_4N_2O_2$ (Acros Organics, 99%), Adenine, $C_5H_5N_5$ (Sigma Aldrich, 99%), Glutamine, $C_5H_{10}N_2O_3$ (Sigma Aldrich, 99%), Asparagine, $C_4H_8N_2O_3$ (Sigma Aldrich, 99%), Histidine, $C_6H_9N_3O_2$ (Sigma Aldrich, 99%), Lysine, $C_6H_{14}N_2O_2$ (Sigma Aldrich, 99%), and Arginine, $C_6H_{14}N_4O_2$ (Sigma Aldrich, 99%)) chosen for their ESI response, as well as Toluene, C_7H_8 (Atrasol grade, Carlo Erba) as dead time marker. These 7 compounds were chosen because they contain a significant amount of oxygen, respond well in ESI, elute over the entire retention range, and are commercially available. Both methods have then been stress-tested on the complex sample described above. Final adjustments were made to prevent memory effects (i.e. “Carry-over”) by having a longer than needed column purge and a long enough equilibration time as HILIC columns are sensitive to poor equilibration conditions[48].

The basic method was run at pH=9.2 using as mobile phase A pure acetonitrile (UPLC/MS grade from Carlo Erba), and B pure water (from water purifier PureLab Flex (Veolia, France)) with 1.92 g/L Ammonium Carbonate. The pH is adjusted to 9.2 using a 32% Ammonia solution. The gradient starts with 80% of A and goes down to 60% of A from 0-11 min. 60% of A is kept up to 45 min. Then the column is washed at 35% of A for 5 min before being equilibrated again at 80% of A for 20 min. The total run time for the pH=9.2 method is 80 min. The acidic method was run at pH=3.2 using as mobile phase A acetonitrile (UPLC/MS grade from Carlo Erba) with 1 ml/L of formic acid (Carlo Erba) and B pure water (from water purifier PureLab Flex (Veolia, France)) with 1.72 g/L Ammonium Formate. The pH is adjusted to 3.2 using pure formic acid. The gradient starts with 85% of A and goes down to 35% for 40 min. 35% of A is then kept for 15 min for column washing, and the column is equilibrated for 20 min at 85% of A. The total run time

for the pH=3.2 method is 80 min. An “analytical sequence” is then composed of four “analytical runs” for pH=3.2 and pH=9.2 and positive and negative modes. Each block includes a column conditioning, a blank, a quality control injection, sample injections, and a quality control injection. The quality control sample is the development mixture (8 compounds) to evaluate a potential retention time drift. In the end, the analytical sequence requires a minimum of 200 µl of liquid sample, made out of 0.2 mg of solid sample, and approximately 3,000 minutes of instrument time, *i.e.,* a little more than two days of instrument time. No unexpected or unusually high safety hazards were encountered.

Data treatment

The data treatment was performed with a custom software package “Attributor” developed with IGOR Pro (WaveMetrics, Portland, OR, USA)[49]. For chromatography analysis, a new custom module for Attributor has specifically been designed and successfully tested and validated on this dataset as further described below. The processing consists of several consecutive steps: raw data loading and noise removal, mass alignment, mass and time 2D peak detection, and 1D fitting of time peaks.

Raw data from the analytical system is directly imported into the software. For a traditional 80-minute chromatographic run, the resulting field of mass/charge versus time is more than hundreds of millions of points, generating several gigabits of data to store, which is difficult for traditional desktop software to process in a reasonable time. One part of the problem is the noise that is generated and acquired by the instrument. A spectrum by spectrum noise level detection is made, and the average and standard deviation values are calculated. This automatic definition is supported by the gaussian-like distribution of noise levels, as presented in Figure 1 with the raw data in panel A, and the

sorted data by increasing noise level values in panel B. Experimentally, we observe that, at the exception of a few outliers, the noise level is similar for each spectrum inside a given analytical run. To ensure a consistent noise level across all samples of the same run, we apply the lowest determined noise level to all the other analyses.

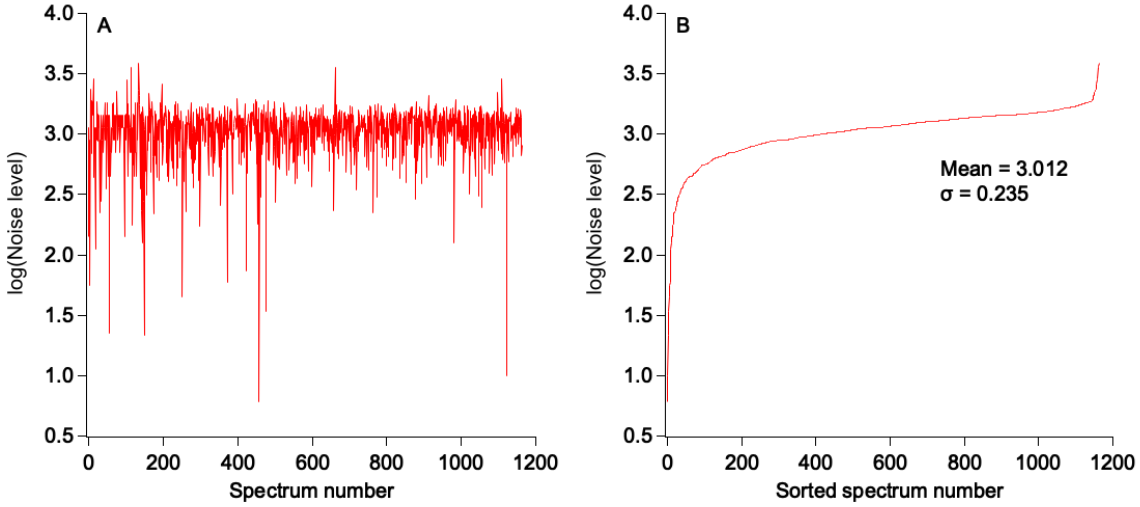


Figure 1 – Automatic noise level definition and distribution shape (A) before and (B) after sorting, showing a gaussian-like noise distribution.

Another reason for such a large data volume is that bin sizes sampling the mass/charge dimension are not equivalent between each spectrum, inducing an artificially high number of very close but different data points. This is presented in Figure 2A, where there are more than 40 different data points to sample the signal for the given peak at *m/z* 141.1385. All these different and very close points in mass/charge can be represented as shown in Figure 3, where we have calculated all differences between one data point and its immediate higher point in mass/charge. Three behaviors are identified: a discrete distribution, an increase of the average mass/charge difference with increasing mass/charge, and random values. The discrete distribution indicates that the bin width is defined over restricted mass/charge ranges, consistent with the overall instrument precision. The increase in mass/charge differences means that the bin size is larger with

mass/charge increasing, consistent with resolution loss with increasing mass/charge. Then, random differences indicate a mass/charge cluster switch, consistent with the sample not populating every possible mass/charge but only those related to specific CHNO combinations. The discrete distribution is fitted with a pure degree three polynomial function. When we invert the different functions from the fit back to exact points in mass/charge, each exact experimental point in mass/charge can be attributed to a unique bin size and mass/charge. This treatment aligns every mass spectrum over the entire time range with a unique bin size, reducing by more than 40 times the number of data points on the mass/charge dimension. This allows the generation of a several million point field (instead of hundreds of millions), allowing desktop computers to process the data (in a couple of minutes). As shown in Figure 2B, some artifacts are present where the intensity seems to be distorted, which is due to oversampling the mass/charge dimension. Such an issue does not create critical problems for other treatments applied later on, as signals are continuous (i.e. not going back to zero in the middle of a peak) and aligned (i.e. the mass/charge of a given peak is constant over time).

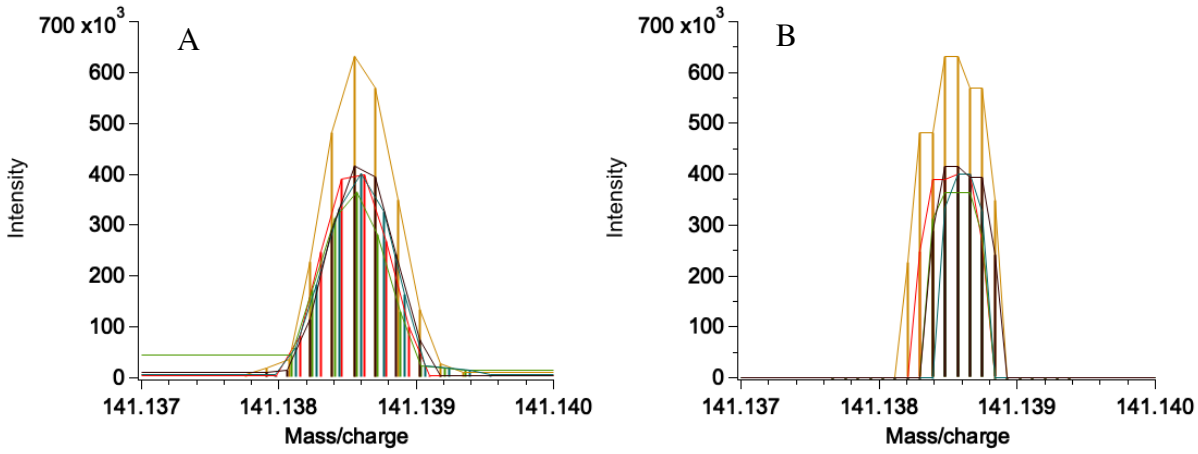


Figure 2 – Zoom on a given peak, showing five consecutive spectra. (A) Before data treatment, points in mass/charge are not aligned and represent 40+ different data points. (B) After data treatment, points in mass/charge are aligned and present ten unique data points.

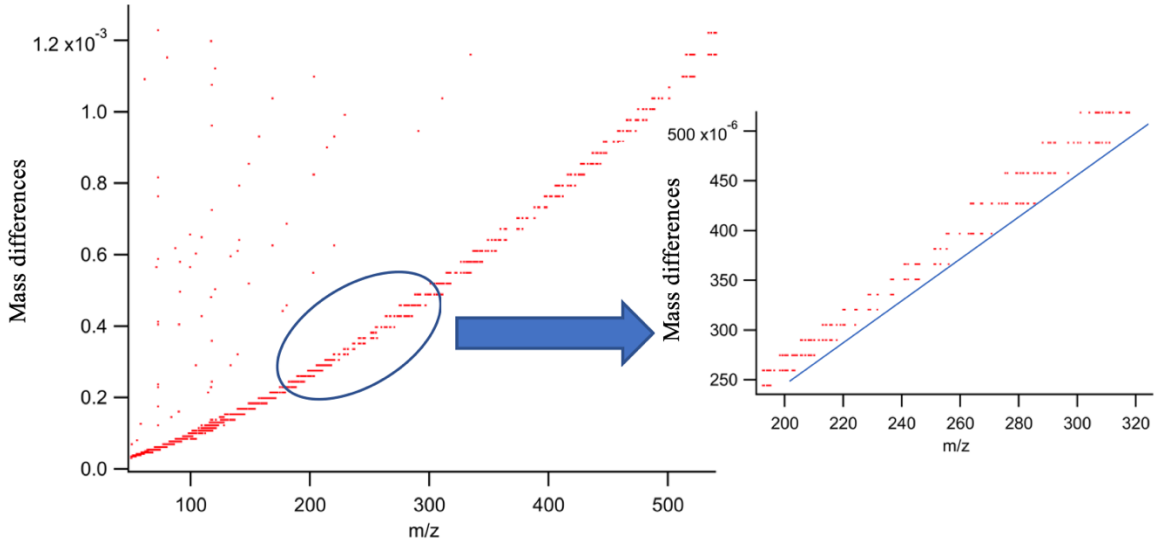


Figure 3 – Representation of mass/charge differences evaluated for each consecutive mass/charge doublet. The blue line is a proxy for the new bin width for the corresponding exact mass/charge.

The field of mass/charge versus time can be directly generated after both previous steps. We present in Figure 4 a comparison before and after the treatment of mass/charge versus time field. The right field is called an “Ion Map”. The mass index is a non-linear dimension, while the time index is directly converted into time values, also called “Retention time”, through a simple scaling. If we zoom inside the map, as shown in Figure 4, we can see what we call “Islands”, a mass/charge versus time subfield of non-zero intensity values circled with zero intensity values. Such islands are detected and listed using an algorithm adapted from the Hoshen-Kopelman algorithm[50] that allows the exhaustive detection and listing of all mass/charge versus time signals in the two-dimensional field. This so-called “island” data set will be used for other treatments and representations in this article.

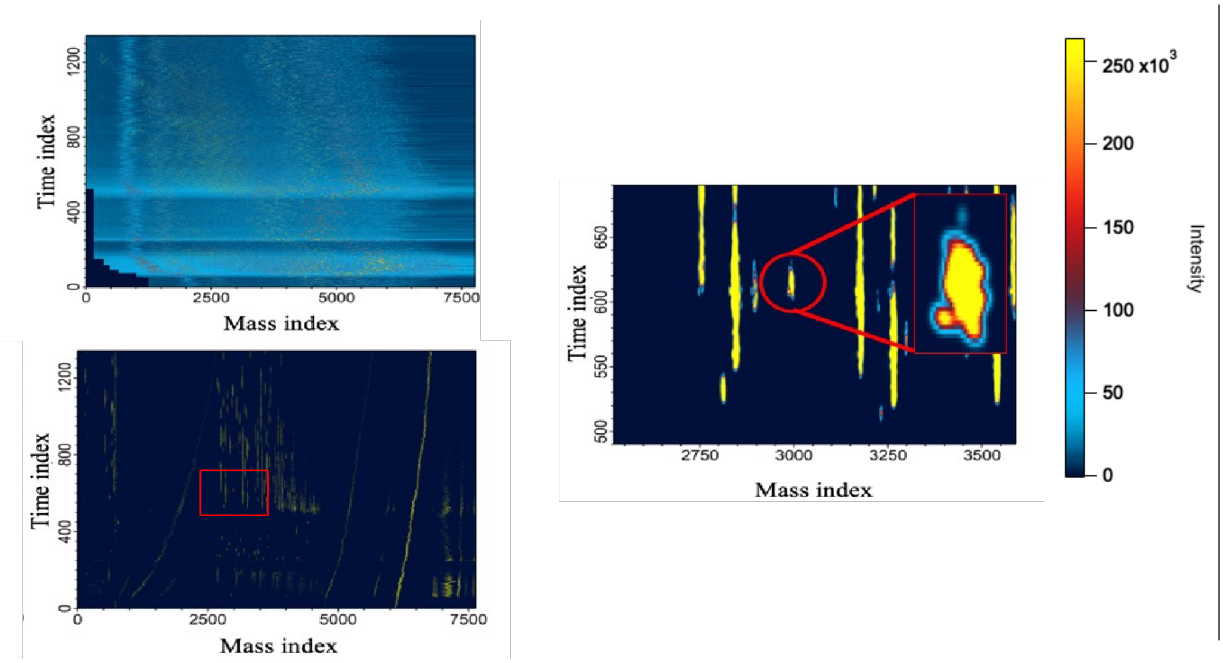


Figure 4 – Comparison of a sample run before and after the data treatment. The same intensity color scale is used.

A problem with this algorithm is that islands can be composed of multiple peaks, such as non-resolved peaks. To make a list of single chromatographic peaks, another treatment is therefore required. Each island is projected on the time dimension on the island list, and a one-dimension peak detection algorithm adapted from Persistent-Homology algorithm[51] is run to determine the number of peaks per island. Then, we fit the peaks with the Exponentially Modified Gaussian model (EMG), a mathematical model that has been used for chromatographic peak fitting for decades[52,53]. We present an example of an island and its fitted model in Figure 5. Each fitted peak is then listed, and its retention time, area, height, and asymmetry estimation are calculated. We then obtain a so-called "fitted" data set that will be used in further data treatment and representation.

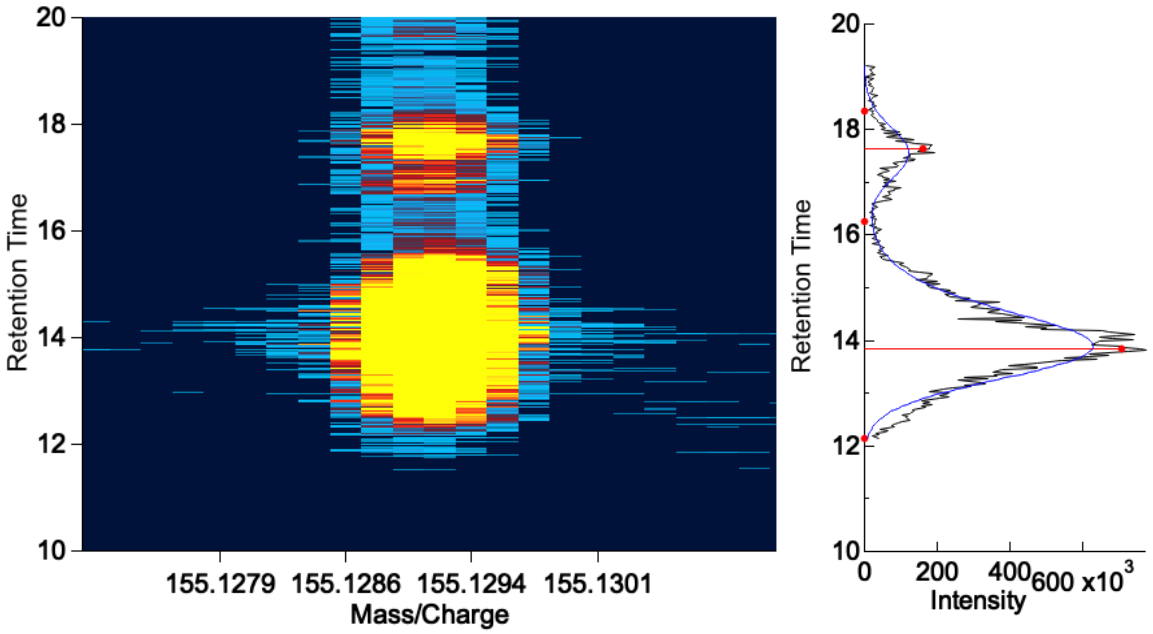


Figure 5 – An island and its fitted model. The red ticks are the persistent-homology results, allowing initial fit guesses. The blue curve on the right panel is the fit result. The fitted model provides all parameters of interest for chromatography use: retention time, peak area, peak height, peak asymmetry estimation, and peak exact mass.

In the end, two different data sets can be used: the “island” and the “fitted” data sets. Special care has to be taken with the island data set as each island can be composed of multiple peaks, implying that it cannot be used for isomer counting as conclusions would be erroneous. The fitted data set has been filtered to ensure that the fitting can be adequately made, excluding weak signals having less than 10 points in time or less than 30 pixels in the 2D mass/charge and time field.

Chromatography and mass spectra molecular assignment

After a complete treatment, a chromatographic peak is defined by its retention time and mass/charge. A mass spectrum can then be reconstructed from each mass/charge value, without the time dimension, to attribute to each mass/charge a unique molecular formula in CHNO. This treatment results in four sets (basic/positive, basic/negative, acidic/positive, acidic/negative) of attributed data that can then be linked again to their respective retention time. With experience on several chromatographic treatments, we

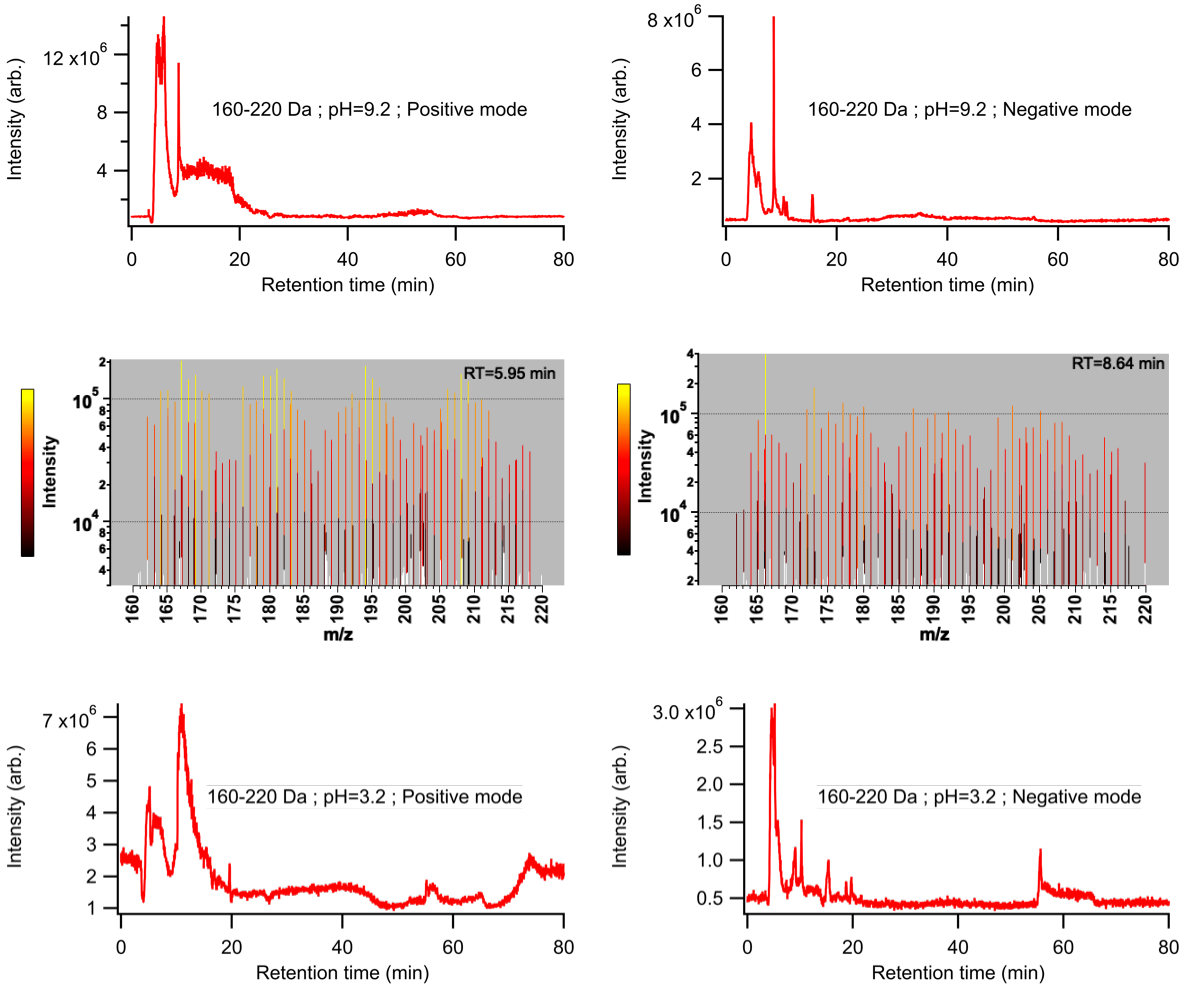
observe that the chromatographic process eliminates noise and artifacts that we usually must deal with in direct infusion mass spectrometry. We observe no discrepancy in the error function generated from the Δ*ppm* of all molecular assignments, with the error function being continuous and at a ±5 *ppm* spread for this dataset. A way to remove incorrect molecular assignments is to use the polymeric behavior of the sample by checking that N and O distributions are dense and continuous over the entire mass range. For acidic and basic methods in positive analysis, molecular assignments that show more than six and eight nitrogen atoms, respectively, and more than three oxygen atoms are removed. For acidic and basic methods in negative analysis, molecular assignments that show more than five and seven nitrogen atoms, respectively, and more than six and four oxygen atoms, respectively, are removed. The final molecular assignments are provided in Figure S1. We note that the acidic runs in negative mode do not allow any molecular assignments for the lower mass range (i.e., < 120 Da) and almost no molecular assignments for the higher mass range (i.e., > 250 Da). Such lack of data is attributed to the low sensitivity of the acidic method in negative mode.

Results

Data exploration

Before any particular treatments, datasets are investigated to check if they present an organic matter diversity. We illustrate in Figure 6 the Total Ion Counts (TIC) and the mass spectrum at the retention time of highest intensity for all four different runs and the mass range 160-220 Da. We can observe that all four runs present information at least in the time range of 5 to 25 min. Highly structured mass spectra are observed in both positive and negative polarity, which is comparable to what we saw in direct infusion in our previous study of the same sample[45]. Such an intensity pattern in the mass spectrum is

characteristic of analogue samples and can be linked to a polymeric-like structure[49,54], confirming the presence of complex molecules. TIC representations show different time features across all analyses. This indicates that the methods successfully separate molecules over a wide range of time and mass/charge, as they were designed to achieve. The extracted mass spectra at multiple retention times also suggest the presence of multiple homologous series, as other works using similar conditions have revealed[55].



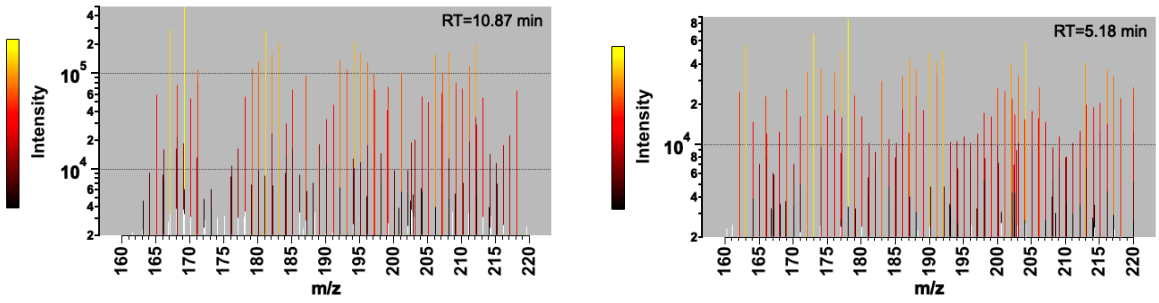
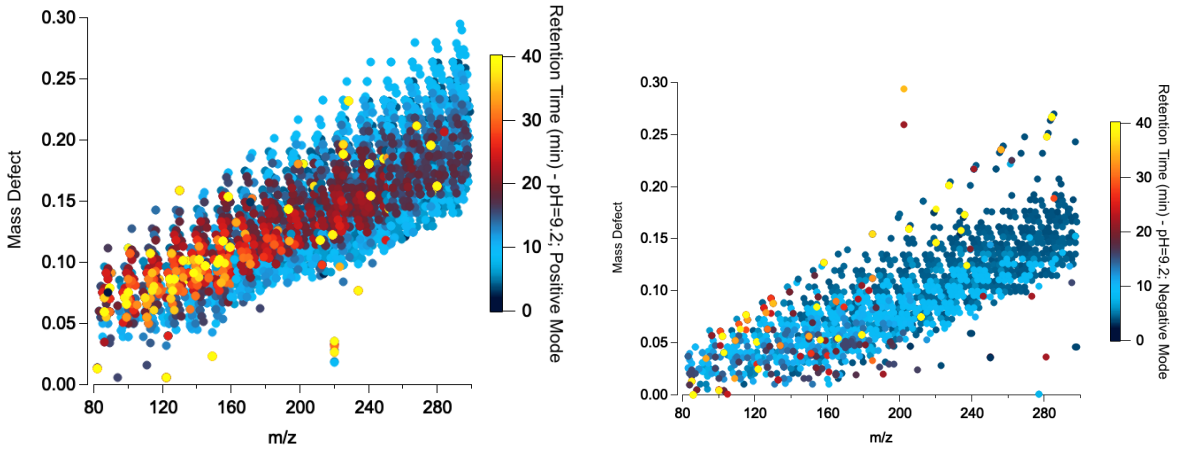


Figure 6 – Total ion counts for the four 160-200 Da chromatographic runs and their associated mass spectra at the retention time with maximum intensity.

Nevertheless, a mass spectrum alone is not a good metric to discuss all the molecular diversity within a sample. For this purpose, we use a mass defect versus mass diagram. This is a representation of the difference between the mass/charge of each data point and the closest integer versus its exact mass/charge. An island dataset is used for the mass defect representation as it is a non-filtered set, and we are interested in seeing the widest possible organic diversity. In Figure 7, we present the four datasets that each include the five different mass ranges. All four representations show a dense organic trend and a variety of retention times depending on the mass/charge and the mass defect value. Negative mode graphs show a less diverse retention time behavior that can be explained by a lower sensitivity of the mass spectrometer in this polarity.



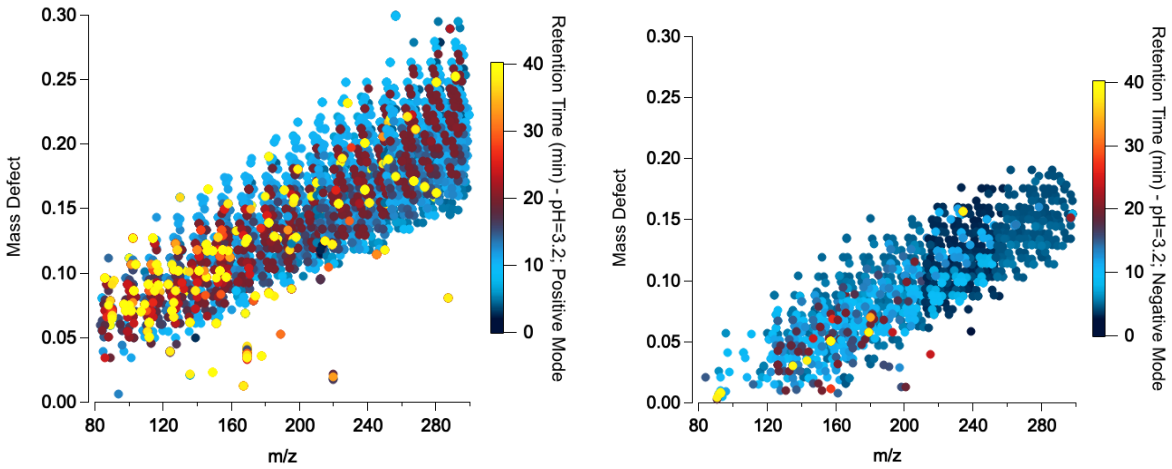


Figure 7 – Mass defect diagrams in the positive and negative mode of the four types of run, color-coded with the associated retention times values. Data are from island detections.

We can observe higher retention times for lower mass/charge, while the higher mass/charge range shows a lower retention time on average. This behavior can be explained by the column being sensitive to polar compounds. With the increasing size of molecules due to their carbon chain elongation, the apparent polarity of the molecules decreases with mass overall, explaining the lowering of retention times with mass increase.

Molecular assignments and chromatography

Before discussing molecular structures, we compare in Figure 8 the molecular assignments from chromatography and from direct infusion mass spectrometry already published in Moran et al.[45] to see if chromatography buffered solutions impact observed molecular formulas. From 55% (in basic mode) to 67% (in acidic mode) of assigned molecular formulas are shared between the direct infusion molecular assignments and the chromatography molecular assignments. Molecular formulas unique to chromatography are suspected to be ion-molecule clusters formed with mobile phase solvents and additives. Such clusters need a specific treatment to dissociate the ion from

the cluster signal, yet to be implemented. Some molecular assignments (not shown here) are also unique to direct infusion because of its higher sensitivity. Indeed, the signal is accumulated during 30 min in direct infusion, corresponding to 250 µl of solution analyzed while in chromatography only 10 µl are injected.

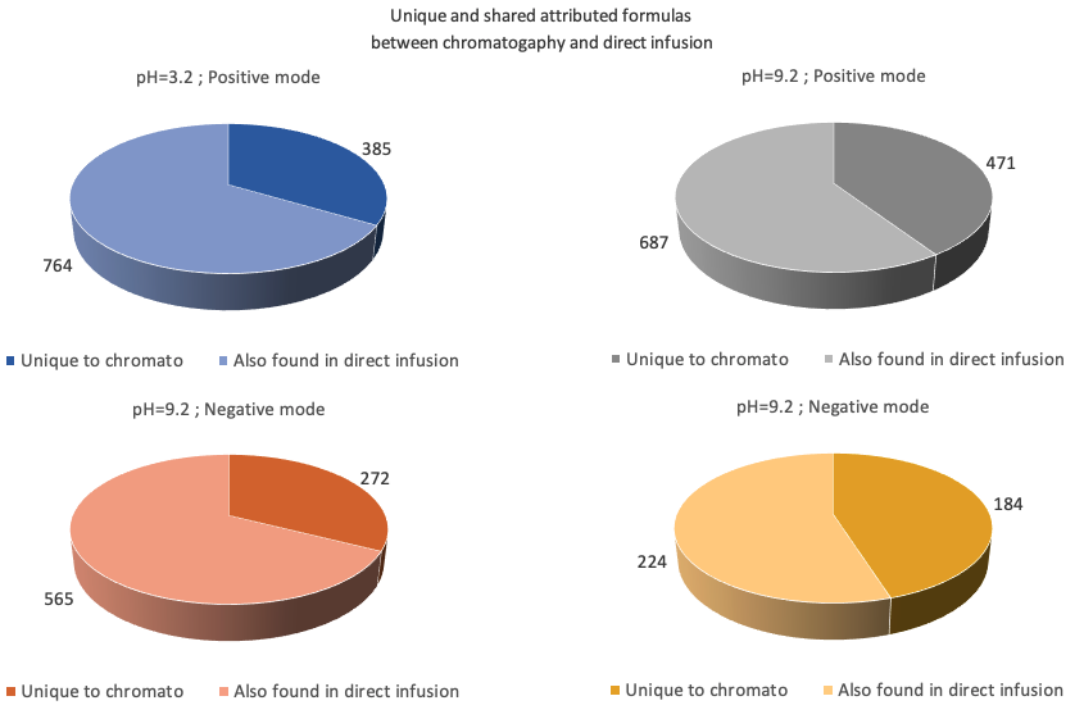


Figure 8 – Comparison of unique molecular formulas from direct infusion and from chromatography, with molecular formulas present in both datasets and molecular formulas unique to chromatography. Molecular formulas that are unique to chromatography have the darkest color. Molecular formulas unique to chromatography are suspected to be ion-molecule clusters formed with mobile phase solvents and additives.

The chromatographic treatment was developed on the hypothesis of orthogonal methods, with the objective to detect the same molecular formulas with the different analytical runs and exploit the constraints provided by each of them. In Figure 9, we present three different pie charts from the enumeration of unique molecular formulas. We can see that 60% of the molecular formulas are seen in at least two analytical runs. To go further, we can see that almost 60% of the unique molecular formulas are seen in acidic and basic pH methods, confirming the interest in running different pH methods to

analyze a complex sample. The use of negative and positive polarities also seems to be an essential point, as about half of the molecules are seen in both negative and positive mode, as was already known from direct injection MS. Such crossed or single-mode observations add more constraints on the possible structures as not all functions can be seen in both polarities.

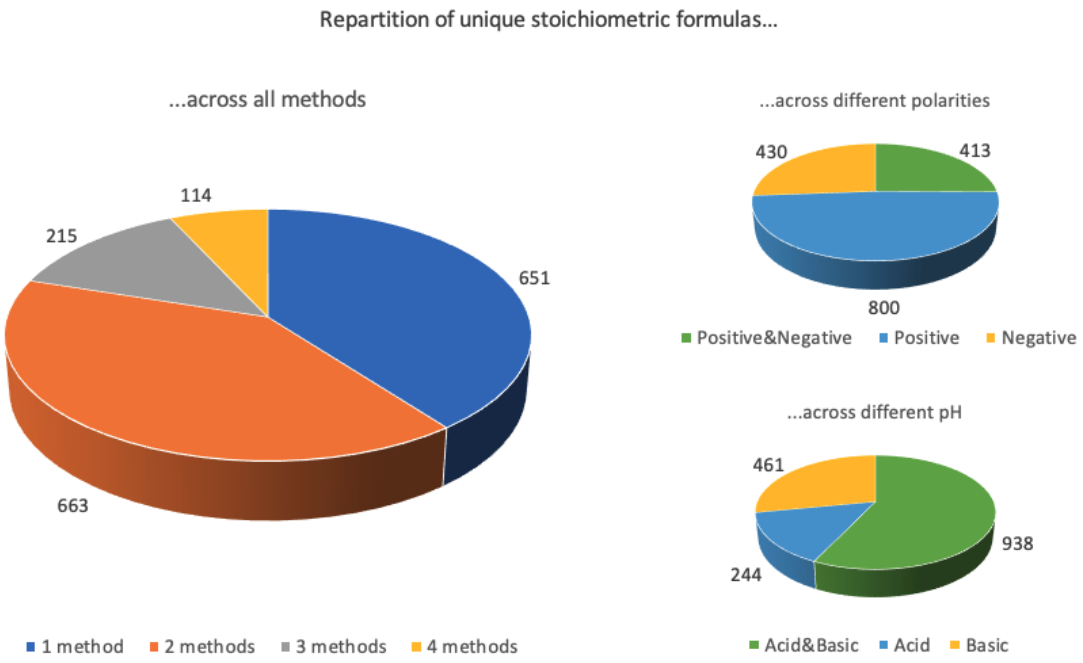


Figure 9 – Enumerated unique molecular formulas for their appearance on 1 to 4 different analytical runs, on positive, negative, or both polarities, and on acid, basic, or both pH methods.

Chromatography is also useful to separate isomers, as we expect different retention times for isomers with the same molecular formula. In Figure 10, two charts compile the average number of isomers and the fraction of attributed molecules with at least two isomers, per mass range and analytical run. These results highlight that on top of the important diversity in molecular formula in this analogue sample, there is also some diversity in molecule structure. A clear difference is observed between positive and negative mode, with less detected isomers in the negative mode than the positive mode. This can be

explained by the instrument being less sensitive in the negative mode than in the positive mode, as already seen in direct infusion[45]. We also observe a higher number of isomers at lower molecular masses compared to higher molecular masses. This can be explained by the general intensity distribution of the sample in direct infusion, which shows lower intensities at higher mass ranges compared to lower mass ranges. Although molecules with higher masses are expected to have as many isomers or more than molecules with lower masses, their concentration are likely below the instrument's detection limit, explaining the counter-intuitive apparent low number of isomers for higher mass ranges. Such observations also confirm the importance of running the analysis across different chromatographic methods and polarities, in line with previous comments.

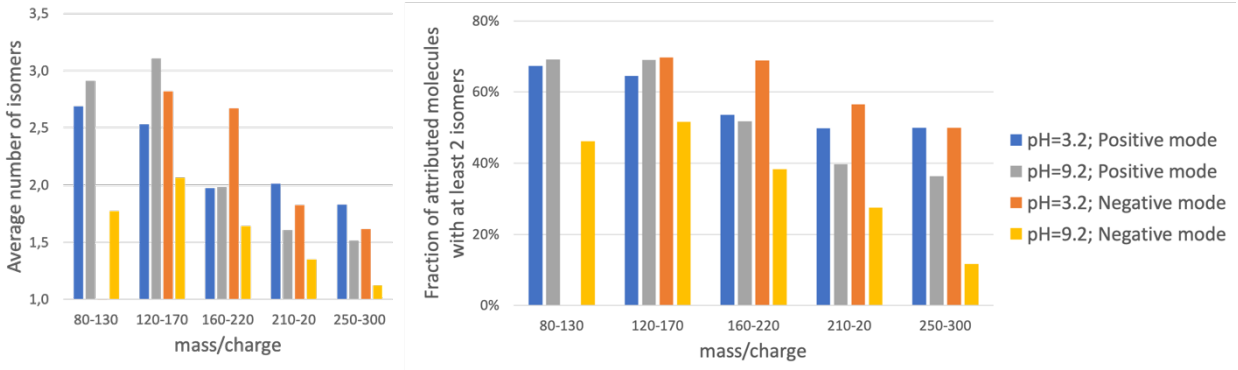


Figure 10 – Statistics on isomers from chromatography data, with the average number of isomers per mass range and the fraction of attributed formulas with at least two isomers for the four analytical runs

From now on, to avoid any discrepancies, we only consider molecular formulas that are seen in both chromatography and direct infusion. Assigned molecules are then compared against the HMDB database of known molecular formulas for small molecule metabolites found in the human body[56]. By comparing our 1,160 molecular assignments to the database, we match 62 molecular formulas present in the database, including 27 amino acids, 15 di-peptides, and 4 nucleobases. These molecular formulas, along with the mass measures, mass precisions and retention times obtained for each analytical run, are

provided in Table S1. We also compared our molecular assignments to those presented in Table 7 in Moran et al.[45] that do include a wider range of molecules of possible interest for exobiology. We match 19 molecular formulas present in this Table, including 9 non-proteinogenic amino acids and one amino acid used in some methanogenic archaea and bacteria. They are available in Table 1, along with the mass measures, mass precisions and retention times obtained for each analytical run. We stress out that although the assigned molecular formulas match that of biologic molecules, we did not confirm molecular structures with MS/MS or coelution as further discussed below. Indeed, the level of annotation obtained correspond to a level G[57]: confirmation of molecular formulae. For any confident annotation against biological molecules, further experiments are needed.

Formula	Exact mass	pH=3,2 - Positive			pH=3,2 - Negative			pH=9,2 - Positive			pH=9,2 - Negative			Potential molecule	Relevance
		Mass	Error	RT	Mass	Error	RT	Mass	Error	RT	Mass	Error	RT		
$C_5H_5N_5O$	151.0494	-	-	-	150.0421	-0.61	10.3	-	-	-	150.0421	-0.46	9.22	Guanine	Nucleotide base
$C_8H_9NO_2$	151.0633	152.0706	-0.31	10.15	-	-	-	152.0706	0.39	9.1	150.0560	-0.66	8.52	2-phenylglycine	Non-proteinogenic amino acid
$C_8H_{11}NO_2$	153.0789	154.0862	0.49	5.05 5.94	-	-	-	154.0862	1.18	8.9 17.5	-	-	-	Dopamine	Non-proteinogenic amino acid
$C_6H_9N_3O_2$	**155.0694**	**156.0767**	**0.18**	**19.54**	**154.0622**	**0.17**	**10.2**	**156.0767**	**1.16**	**5.05 8.79 15.8**	**154.0622**	**-0.46**	**5.46 9.72 11.2**	**Histidine**	**Biological amino acid**
$C_5H_8N_4O_2$	156.0647	-	-	-	-	-	-	157.0722	1.83	10.4 11.6 16.3	155.0573	-0.53	11.1 12.5	1,2,4-tyriazole-3-alanine	Non-proteinogenic amino acid
$C_7H_{11}NO_3$	157.0738	-	-	-	156.0665	-0.63	9.02	-	-	-	-	-	-	Furanomycin	Non-proteinogenic amino acid
$C_9H_{11}NO_2$	165.0789	166.0862	-0.19	10.2	164.0717	0.73	4.94 5.00	166.0862	0.36	5.34 5.67 8.60	-	-	-	phenylalanine	Non-proteinogenic amino acid
$C_7H_9N_3O_2$	167.0694	-	-	-	166.0622	0.16	5.07 9.89 11.2	168.0767	-0.29	9.81	166.0619	-1.38	5.9	β-pyrazinyl-L-alanine	Non-proteinogenic amino acid
$C_7H_{11}N_3O_2$	169.0851	-	-	-	168.0778	0.70	9.86	170.0924	-0.20	5.69 17.3	168.0778	2.48	10.8	3-methylhistidine	Non-proteinogenic amino acid
$C_6H_9N_3O_3$	171.0644	-	-	-	170.0568	-1.75	9.31	-	-	-	-	-	-	β-hydroxyhistidine	Non-proteinogenic amino acid
$C_{10}H_{13}NO_2$	179.0946	180.1016	-1.53	4.78	178.0871	-1.13	4.73	180.1017	-0.60	4.4 5.9	-	-	-	Homophenylalanine	Non-proteinogenic amino acid
$C_7H_{10}N_4O_2$	182.0803	-	-	-	-	-	-	183.0874	-1.10	8.79	-	-	-	Lathyrine	Non-proteinogenic amino acid
$C_9H_{12}N_2O_3$	196.0847	-	-	-	195.0773	-0.66	4.44 9.01 5.21	197.0920	-0.11	5.48	195.0775	0.25	5.80	Pyridinylmethylserine	Non-proteinogenic amino acid
$_{11}H_{12}N_2O_2$	**204.0898**	**-**	**-**	**-**	**203.0826**	**0.36**	**4.40 5.00**	**-**	**-**	**-**	**-**	**-**	**-**	**Tryptophan**	**Biological amino acid**
$_{10}H_{11}N_3O_2$	205.0851	-	-	-	204.0778	0.20	4.25 8.79	-	-	-	204.0777	-0.31	4.40	Tryptazan	Non-proteinogenic amino acid
$_9H_{10}N_4O_2$	206.0803	-	-	-	205.0730	-0.25	5.04 10.0	-	-	-	205.0732	0.65	8.56	Benzotriazolylalanine	Non-proteinogenic amino acid
$_9H_{14}N_4O_3$	226.1065	-	-	-	-	-	-	227.1138	0.71	8.47 9.81	-	-	-	Alanylhistidine	Biological amino acid metabolit
$_{10}H_{18}N_2O_5$	246.1215	-	-	-	245.1142	-1.12	5.09	-	-	-	-	-	-	Boc-L-glutamine	Non-proteinogenic amino acid
$_{12}H_{21}N_3O_3$	255.1582	256.1658	1.40	5.01 10.4 13.0	-	-	-	256.1659	1.52	4.23 8.57	-	-	-	Pyrrolysine	Biological amino acid

Table 1 – Molecular formulas, mass measures, mass precisions and retention times obtained for each analytical run for the subset of molecules presented in Table 7 of Moran et al.[45]*. We advise that we only assigned the molecular formulas presented in this Table, we did not confirm their molecular structure.*

Annotating all 19 molecular formulas listed in Table 1 is beyond the scope of this paper but, as an example, we compared the retention times obtained for $C_{11}H_{12}N_2O_2$ and $C_6H_9N_3O_2$, shown in Table 2, to the experimental retention times of Tryptophan and Histidine standards, respectively. Tryptophan standard times are very different from times we observe in the sample, indicating that Tryptophan is probably not present, or if so, at very low concentrations that cannot be reached with this analytical system (see discussion about detection limits). Histidine standard times are only close to one retention time of $C_6H_9N_3O_2$, at pH=9.2 in positive mode (measured retention times of 15.83 and 16.34 min, and standard retention time of 16.2 min). In Figure 11 we show an Ion Map and a chromatogram with its related EMG fit for the tentatively annotated Histidine peak. We observe a weak non-resolved peak, indicating that Histidine might be present in the sample but selective techniques such as GC-MS would have to be run to confirm its presence or not.

Formula	Name	pH=3,2 - Positive	pH=3,2 - Negative	pH=9,2 - Positive	pH=9,2 - Negative
$C_{11}H_{12}N_2O_2$	Tryptophan	16.3	16.3	12.9	12.9
$C_6H_9N_3O_2$	Histidine	29.7	29.7	16.2	16.2

Table 2 – Standard retention times for Histidine and Tryptophan.

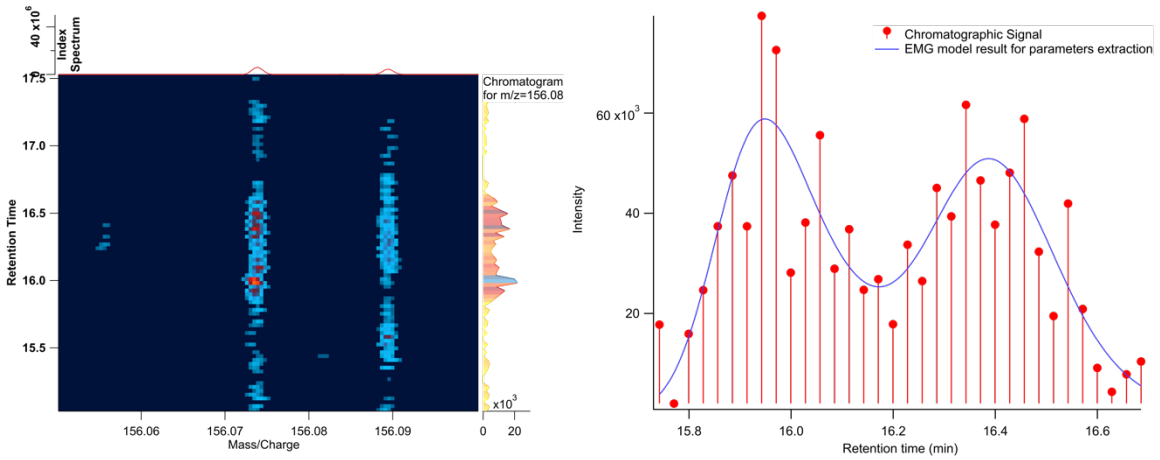


Figure 11 – Ion Map, chromatogram, and EMG automatic fitting for m/z 156.077 ($C_6H_{10}N_3O_2^+$) between 15 min and 17.5 min and for the pH=9.2 in positive mode.

Discussion

The classical "Full-Loop" injection mode traditionally used with chromatography uses up to 5 times the real injected sample volume. Such consumed volumes would not be sustainable with the sample volume we effectively have for our use in the case of extraterrestrial materials and analogues. Instead, we used the "Microliter Pickup" mode from Thermo Scientific automatic sampler instrument where 10 µl of the solution is effectively used when injecting 10 µl in the system. Such a low sample consumption allows precious extraterrestrial materials available in small quantities to be analyzed with several analytical runs and over multiple mass ranges.

When there is no analyte out of the chromatography instrument, the AGC can store ions up to 500 ms, while when there are compounds, the AGC usually reaches the required ion count in less than 10 ms. We set the AGC accumulation time to less than 10 ms to ensure the time stability of the chromatography. Lowering the maximum AGC accumulation time also avoids a low-level background to be artificially increased by waiting longer to stockpile ions. This makes data treatment faster by reducing data size and having an almost constant spectrum-to-spectrum time step, allowing a proper scaling of the retention time axis for all representations without distortions.

We have acquired data for higher mass ranges, i.e., for mass ranges higher than *m/z* 300, but such results cannot be exploited as the instrument precision and resolution do not allow for reliable molecular assignment due to the multiplication of ions with very close *m/z* in this mass range. For direct infusion molecular assignments, we usually acquire a single mass range from 150-800 *m/z*, and we use mathematical ways to propagate the molecular assignments from low *m/z* up to high *m/z* with high confidence[23,58] . In the case

of chromatography runs, the mass range is sliced into 50 Da blocks, and with the way the instrument works, calibration shifts slightly, making it impossible to use a continuous polynomial function to recalibrate all analyses. New data treatment and mathematical tools are then needed to extrapolate to higher mass ranges, hence they are not presented and discussed in this work.

In order to unveil the isomeric ambiguity behind a given molecular formula, it is necessary to rely on chemical standards and to compare their retention time and/or fragmentation pattern (tandem mass spectrometry) to that measured for the species in the complex sample[35,59]. Since the number of isomers may be prohibitive, the identification may be supported by comparison with standard tandem mass spectral databases, network-based approaches[60]. The prediction of retention times could also be performed to narrow down the possibilities and select only relevant molecules for standard comparison[47].

The analysis of the MS/HRMS fragmentation patterns after direct infusion of aerosol analogues has proven to be difficult[61]. First, for instrumental reasons, the selection window for the parent ion(s) cannot be narrower than 1 Da and, considering the complexity of the samples, a mixture of ions with different molecular formulas is actually selected. This implies that most of the fragment ions can originate from more than one parent ion, limiting accurate interpretation. Next, as discussed above, it is likely that, even at a given same exact mass, isomeric molecules are present and since each exhibit a specific fragmentation signature in the MS/HRMS, the interpretation is again difficult. The obvious advantage of HPLC is that by spreading the molecular complexity over time, it

should become easier to select pure or quasi-pure ions, rendering the identification of the isomers easier.

Some non-published experimental data that have been obtained during the analytical development of this work indicate that detection limits in the positive mode are around the ppm level for biological amino acids. A better quantification would be challenging as a calibration curve made with pure standard compounds will not be equivalent to what will occur with a complex mixture made of hundreds to thousands of compounds because of ion suppression in the ESI source. Another possibility to obtain a better estimate of amino acid (or any other compound) concentrations would be to use the standard addition method, where the chromatographic peak area of a given compound in the mixture is measured before and after adding a known amount of the corresponding standard[44]. However, since haze analogue samples are not 100% soluble[45,62] and the solubility of one compound might differ from one another, it would be impossible to interpret these values as actual concentrations within the bulk sample.

Conclusion

We have developed two liquid chromatography analytical methods coupled to ultra-high resolution mass spectrometry that use orthogonal conditions to analyze complex organic samples with nitrogen and oxygen content. We have shown that by combining basic and acidic methods and ionization with both polarities, i.e. four analytical runs, we can establish a list of over a thousand molecular formulas. While most of them are detected on multiple different runs, others can only be seen on one of them, justifying the complementary nature of all four runs to extract as much information as possible with one single LC-HRMS setup. Such work then allows the creation of a list of molecular formulas

to be annotated, either on the same system with pure standards or by more selective techniques such as GC-MS, for final identifications.

ASSOCIATED CONTENT

Figure S1 (PDF): Assigned molecular formulas after filtering for positive and negative runs with acidic and basic chromatography methods. The numbers of N and O are color-coded.

Table S1 (PDF): Potential matches between molecular formulas assigned in our sample and molecular formulas of known biologic CHNO compounds (https://hmdb.ca/). Corresponding retention times are provided for positive and negative runs with acidic and basic chromatography methods. We advise that we did not confirm molecular structure and assign only the formulas for all of the molecules in this Table.

AUTHOR INFORMATION

Corresponding Author

* veronique.vuitton@univ-grenoble-alpes.fr

Present Addresses

Author Contributions

CW and VV designed this study and wrote the paper. CW and LF performed LC/HRMS experiments. CW and FROD performed LC/HRMS data treatment. CH and SEM performed the synthesis of the laboratory analogues under the supervision of SH. All authors have approved the final version of the manuscript.

Funding Sources

Agence Nationale de la Recherche, Grant/ Award Numbers: ANR-15-IDEX-02, ANR-16-CE29-0015 2016-2021

ACKNOWLEDGMENT

This work is supported by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX- 02), through the funding of the “Origin of Life” project of the Université Grenoble Alpes and the French Space Agency (CNES) under its Exobiology, Exoplanets and Planetary Protection program (VV). CW acknowledges a PhD fellowship from CNES/ANR (ANR-16-CE29-0015 2016-2021). The samples were synthesized with support from NASA Earth and Space Science Fellowship grant 80NSSC18K1109 (SEM) and NASA Exoplanets Research Program grant NNX16AB45G (CH, SMH). A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the ANR grants’ open access conditions.

ABBREVIATIONS

REFERENCES

1. Mifsud, D. V. *et al.* The Role of Terahertz and Far-IR Spectroscopy in Understanding the Formation and Evolution of Interstellar Prebiotic Molecules. *Front. Astron. Space Sci.* **8**, 757619 (2021).

2. Ercolano, B. *et al.* Observations of PAHs in the atmospheres of discs and exoplanets. *Mon. Not. R. Astron. Soc.* **512**, 430–438 (2022).

3. Cordiner, M. A. *et al.* ALMA Spectral Imaging of Titan Contemporaneous with Cassini’s

Grand Finale. *Astron J* **158**, 76 (14pp) (2019).

4. Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D. & Hord, C. A search for life on Earth from the Galileo spacecraft. *Nature* **365**, 715–721 (1993).

5. Bartlett, S. & Wong, M. L. Defining Lyfe in the Universe: From Three Privileged Functions to Four Pillars. *Life* **10**, 42 (2020).

6. Meadows, V. S. *et al.* Exoplanet Biosignatures: Understanding Oxygen as a Biosignature in the Context of Its Environment. *Astrobiology* **18**, 630–662 (2018).

7. He, C. *et al.* Gas Phase Chemistry of Cool Exoplanet Atmospheres: Insight from Laboratory Simulations. *ACS Earth Space Chem.* **3**, 39–50 (2019).

8. Krissansen-Totton, J., Fortney, J. J., Nimmo, F. & Wogan, N. Oxygen False Positives on Habitable Zone Planets Around Sun-Like Stars. *AGU Adv.* **2**, (2021).

9. Bardyn, A. *et al.* Carbon-rich dust in comet 67P/Churyumov-Gerasimenko measured by COSIMA/Rosetta. *Mon. Not. R. Astron. Soc.* **469**, S712–S722 (2017).

10. Le Roy, L. *et al.* Inventory of the volatiles on comet 67P/Churyumov-Gerasimenko from Rosetta/ROSINA. *Astron. Astrophys.* **583**, A1 (2015).

11. March, R. E. OSIRIS-REx: a NASA Asteroid Space Mission. *Int. J. Mass Spectrom.* **469**, 116677 (2021).

12. Watanabe, S. *et al.* Hayabusa2 Mission Overview. *Space Sci. Rev.* **208**, 3–16 (2017).

13. Bland, P. A. *et al.* The flux of meteorites to the Earth over the last 50 000 years. *Mon. Not. R. Astron. Soc.* **283**, 551–565 (1996).

14. Cooper, J. Energetic Ion and Electron Irradiation of the Icy Galilean Satellites. *Icarus* **149**, 133–159 (2001).

15. Botta, O., Martins, Z. & Ehrenfreund, P. Amino acids in Antarctic CM1 meteorites and their relationship to other carbonaceous chondrites. *Meteorit. Planet. Sci.* **42**, 81–92 (2007).

16. Schmitt-Kopplin, P. *et al.* High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *P Natl Acad Sci USA* **107**, 2763–2768 (2010).

17. Naraoka, H., Yamashita, Y., Yamaguchi, M. & Orthous-Daunay, F.-R. Molecular Evolution of N-Containing Cyclic Compounds in the Parent Body of the Murchison Meteorite. *ACS Earth Space Chem.* **1**, 540–550 (2017).

18. Maillard, J., Carrasco, N., Schmitz-Afonso, I., Gautier, T. & Afonso, C. Comparison of soluble and insoluble organic matter in analogues of Titan's aerosols. *Earth Planet. Sci. Lett.* **495**, 185–191 (2018).

19. Sebree, J. A., C. Roach, M., Shipley, E. R., He, C. & Hörst, S. M. Detection of Prebiotic Molecules in Plasma and Photochemical Aerosol Analogs Using GC/MS/MS Techniques. *Astrophys. J.* **865**, 133 (2018).

20. Fayolle, M. *et al.* Testing tholins as analogues of the dark reddish material covering Pluto's Cthulhu region. *Icarus* **367**, 114574 (2021).

21. Jovanović, L. *et al.* Chemical composition of Pluto aerosol analogues. *Icarus* **346**, 113774 (2020).

22. Moran, S. E. *et al.* Triton Haze Analogs: The Role of Carbon Monoxide in Haze Formation. *J. Geophys. Res. Planets* **127**, e2021JE006984 (2022).

23. Vuitton, V. *et al.* H_2SO_4 and Organosulfur Compounds in Laboratory Analogue Aerosols of Warm High-metallicity Exoplanet Atmospheres. *Planet. Sci. J.* **2**, 2 (2021).

24. Hörst, S. M. *et al.* Haze production rates in super-Earth and mini-Neptune atmosphere experiments. *Nat. Astron.* **2**, 303–306 (2018).

25. Urso, R. G., Baklouti, D., Djouadi, Z., Pinilla-Alonso, N. & Brunetto, R. Near-infrared Methanol Bands Probe Energetic Processing of Icy Outer Solar System Objects. *Astrophys. J.* **894**, L3 (2020).

26. Ruf, A. *et al.* Organosulfur Compounds Formed by Sulfur Ion Bombardment of Astrophysical Ice Analogs: Implications for Moons, Comets, and Kuiper Belt Objects. *Astrophys. J.* **885**, L40 (2019).

27. He, C., Hörst, S. M., Radke, M. & Yant, M. Optical Constants of a Titan Haze Analog from 0.4 to 3.5 μm Determined Using Vacuum Spectroscopy. *Planet. Sci. J.* **3**, 25 (2022).

28. Gavilan, L., Carrasco, N., Hoffmann, S. V., Jones, N. C. & Mason, N. J. Organic Aerosols in Anoxic and Oxic Atmospheres of Earth-like Exoplanets: VUV-MIR Spectroscopy of CHON Tholins. *Astrophys J* **861**, #110 (2018).

29. Danger, G. *et al.* Characterization of laboratory analogs of interstellar/cometary organic residues using very high resolution mass spectrometry. *Geochim Cosmochim Acta* **118**, 184–201 (2013).

30. Hörst, S. M. *et al.* Formation of amino acids and nucleotide bases in a Titan atmosphere simulation experiment. *Astrobiology* **12**, 809–817 (2012).

31. Danger, G. *et al.* The transition from soluble to insoluble organic matter in interstellar ice analogs and meteorites. *Astron. Astrophys.* **667**, A120 (2022).

32. Meinert, C., Filippi, J.-J., de Marcellus, P., Le Sergeant d'Hendecourt, L. & Meierhenrich, U. J. N-(2-Aminoethyl)glycine and Amino Acids from Interstellar Ice Analogues. *ChemPlusChem* **77**, 186–191 (2012).

33. Elsila, J. E., Dworkin, J. P., Bernstein, M. P., Martin, M. P. & Sandford, S. A. Mechanisms of Amino Acid Formation in Interstellar Ice Analogs. *Astrophys. J.* **660**, 911–918 (2007).

34. Dworkin, J. P., Deamer, D. W., Sandford, S. A. & Allamandola, L. J. Self-assembling amphiphilic molecules: Synthesis in simulated interstellar/precometary ices. *Proc. Natl. Acad. Sci.* **98**, 815–819 (2001).

35. Ruf, A. *et al.* The Challenging Detection of Nucleobases from Pre-accretional Astrophysical Ice Analogs. *Astrophys. J.* **887**, L31 (2019).

36. Nuevo, M., Cooper, G. & Sandford, S. A. Deoxyribose and deoxysugar derivatives from photoprocessed astrophysical ice analogues and comparison to meteorites. *Nat. Commun.* **9**, 5276 (2018).

37. Meinert, C. *et al.* Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **352**, 208–212 (2016).

38. Bada, J. L., Glavin, D. P., McDonald, G. D. & Becker, L. A Search for Endogenous Amino Acids in Martian Meteorite ALH84001. *Science* **279**, 362–365 (1998).

39. Kojo, S. S-Isovaline Contained in Meteorites, Induces Enantiomeric Excess in D,L-glutamic Acid During Recrystallization. *Orig. Life Evol. Biospheres* **45**, 85–91 (2015).

40. Callahan, M. P. *et al.* Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *PNAS* **108**, 13995–13998 (2011).

41. Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **499**, 177–196 (1990).

42. McCalley, D. V. Is hydrophilic interaction chromatography with silica columns a viable alternative to reversed-phase liquid chromatography for the analysis of ionisable compounds? *J. Chromatogr. A* **1171**, 46–55 (2007).

43. Eddhif, B. *et al.* Development of liquid chromatography high resolution mass spectrometry strategies for the screening of complex organic matter: Application to astrophysical simulated materials. *Talanta* **179**, 238–245 (2018).

44. Gautier, T. *et al.* Development of HPLC-Orbitrap method for identification of N-bearing molecules in complex organic material relevant to planetary environments. *Icarus* **275**, 259–266 (2016).

45. Moran, S. E. *et al.* Chemistry of Temperate Super-Earth and Mini-Neptune Atmospheric Hazes from Laboratory Experiments. *Planet. Sci. J.* **1**, 17 (2020).

46. He, C. *et al.* Carbon Monoxide Affecting Planetary Atmospheric Chemistry. *Astrophys. J.* **841**, L31 (2017).

47. Creek, D. J. *et al.* Toward Global Metabolomics Analysis with Hydrophilic Interaction Liquid Chromatography–Mass Spectrometry: Improved Metabolite Identification by Retention Time Prediction. *Anal. Chem.* **83**, 8703–8710 (2011).

48. Snyder, L. R., Kirkland, J. J. & Dolan, J. W. *Introduction to modern liquid chromatography*. (Wiley, 2010).

49. Orthous-Daunay, F.-R. *et al.* Ultraviolet-photon fingerprints on chondritic large organic molecules. *Geochem. J.* **53**, 21–32 (2019).

50. Hoshen, J. & Kopelman, R. Percolation and cluster distribution. I. Cluster multiple labeling

technique and critical concentration algorithm. *Phys. Rev. B* **14**, 3438–3445 (1976).

51. Verri, A., Uras, C., Frosini, P. & Ferri, M. On the use of size functions for shape analysis. *Biol. Cybern.* **70**, 99–107 (1993).

52. Grushka, Eli. Characterization of exponentially modified Gaussian peaks in chromatography. *Anal. Chem.* **44**, 1733–1738 (1972).

53. Kalambet, Y., Kozmin, Y., Mikhailova, K., Nagaev, I. & Tikhonov, P. Reconstruction of chromatographic peaks using the exponentially modified Gaussian function. *J. Chemom.* **25**, 352–356 (2011).

54. Pernot, P., Carrasco, N., Thissen, R. & Schmitz-Afonso, I. Tholinomics: Chemical analysis of nitrogen-rich polymers. *Anal Chem* **82**, 1371–1380 (2010).

55. Yamashita, Y. & Naraoka, H. Two homologous series of alkylpyridines in the Murchison meteorite. *Geochem. J.* **48**, 519–525 (2014).

56. Wishart, D. S. *et al.* HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.* **50**, D622–D631 (2022).

57. Rampler, E. *et al.* Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics—Standardization, Coverage, and Throughput. *Anal. Chem.* **93**, 519–545 (2021).

58. Hörst, S. M. Post-Cassini investigations of Titan atmospheric chemistry. (University of Arizona, 2011).

59. Naraoka, Hashiguchi, Sato, & Hamase. New Applications of High-Resolution Analytical Methods to Study Trace Organic Compounds in Extraterrestrial Materials. *Life* **9**, 62 (2019).

60. Cai, Y., Zhou, Z. & Zhu, Z.-J. Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics. *TrAC Trends Anal. Chem.* **158**, 116903 (2023).

61. Vuitton, V. *et al.* Very high resolution mass spectrometry of HCN polymers and tholins. *Faraday Discuss* **147**, 495–508 (2010).

62. Carrasco, N. *et al.* Chemical characterization of Titan's tholins: Solubility, morphology and molecular structure revisited. *J Phys Chem A* **113**, 11195–11203 (2009).

FOR TABLE OF CONTENTS ONLY