Article

Data-Driven UPLC-Orbitrap MS Analysis in Astrochemistry

Alexander Ruf [1,*] **, Pauline Poinot** [2] **, Claude Geffroy** [2] **, Louis Le Sergeant d'Hendecourt** [1] **and Gregoire Danger** [1,*]

[1] Laboratoire de Physique des Interactions Ioniques et Moléculaires (PIIM), Université Aix-Marseille, Saint Jérôme—AVE Escadrille Normandie Niemen, 13013 Marseille, France; alexander.ruf@univ-amu.fr (A.R.); ldh@ias.u-psud.fr (L.L.S.d.)

[2] Institut de Chimie des Milieux et Matériaux de Poitiers (IC2MP), Université de Poitiers, UMR CNRS 7285, 86073 Poitiers, France; pauline.poinot@univ-poitiers.fr (P.P.); claude.geffroy@univ-poitiers.fr (C.G.)

* Correspondence: alexander.ruf@univ-amu.fr (A.R.); gregoire.danger@univ-amu.fr (G.D.); Tel.: +33-491288285 (G.D.)

check for updates

Abstract: Meteorites have been found to be rich and highly diverse in organic compounds. Next to previous direct infusion high resolution mass spectrometry experiments (DI-HR-MS), we present here data-driven strategies to evaluate UPLC-Orbitrap MS analyses. This allows a comprehensive mining of structural isomers extending the level of information on the molecular diversity in astrochemical materials. As a proof-of-concept study, Murchison and Allende meteorites were analyzed. Both, global organic fingerprint and specific isomer analyses are discussed. Up to 31 different isomers per molecular composition are present in Murchison suggesting the presence of ≈440,000 different compounds detected therein. By means of this time-resolving high resolution mass spectrometric method, we go one step further toward the characterization of chemical structures within complex extraterrestrial mixtures, enabling a better understanding of organic chemical evolution, from interstellar ices toward small bodies in the Solar System.

Keywords: astrochemistry; meteorites; high-resolving analytical chemistry; data analysis; origin of life

1. Introduction

Organic chemistry is known to be rich and almost universal in astronomical objects, as observed from a huge molecular diversity in the interstellar medium (ISM) [1], in interstellar/precometary ices of laboratory analogs [2–4], in comets [5–7] or in meteorites [8–10]. When probing different steps in chemical evolution, from simple molecular ices toward complex organic materials, various prebiotically relevant compounds, such as amino acids, sugars or nucleobases are formed in precometay laboratory ices [11,12] or in meteorites [13–15]. In addition, organized molecular structures, such as peptides, have been proposed to be present in residues of interstellar ices [16–18]. These findings indicate that molecular assembly processes of astrochemical organic matter might have been present within early stages of the Solar System formation. Molecular organization is crucial in early prebiotic phases and might represent a marker to probe the transition between astro-/cosmochemistry and prebiotic chemistry [19,20].

The description of astrochemical organic matter as seeds for life and their interactions within various astrophysical environments may thus appear essential to further study problems regarding the emergence of life, in a given telluric planetary environment [19]. In-depth analysis is essential to understand its evolution, especially for testing the role of extraterrestrial organics in the fields of prebiotic chemistry and astrobiology (e.g., self-organization processes). Besides spectroscopic techniques [21–23],

mass spectrometry (MS) has been widely used in organic astro- and cosmochemistry over the last several years [19]. MS has also powerfully demonstrated to in situ study the chemical composition of extraterrestrial objects, e.g., for the comet 67P/Churyumov–Gerasimenko [24]. More recently, high resolution mass spectrometric (HR-MS) instruments have had a significant and an ongoing increasing impact in current astrochemical research [25]. For instance, Fourier transform ion cyclotron resonance mass spectrometric (FT-ICR-MS) analyses enabled to resolve mass signals which differ in mass differences by less than the mass of an electron, thanks to resolving power R > 10^6 and mass accuracy <200 ppb [9]. This type of sophisticated instrumentation allows one to study novel chemistry—as, for example, the one involving organomagnesium compounds in meteorites showing that a rich and complex chemistry has occurred in meteoritic parent bodies [9]. In context of space exploration, another type of high-resolving mass spectrometers, Orbitrap-type instruments, has been optimized in miniaturized form (called "Cosmorbitrap") [26,27]. Cosmorbitrap instruments could be directly sent on space missions for high resolution mass spectrometric in situ analysis to characterize the chemistry of planetary environments. In nowadays sample return missions (e.g., Hayabusa 2 [28] or OSIRIS-REx [29]), collected extraterrestrial samples are expected to be analyzed as comprehensive as possible. High sensitivity, high selectivity and high resolution should be achieved, but it is also intended to directly elucidate chemical structures. Indeed, most high-resolving mass spectrometric studies on astrochemical organic matter are operated in direct infusion electrospray ionization mass spectrometric mode (DI-HR-MS) [2–4,8,9]. By means of this technique, powerful insights on separate molecular formulas (isobars and isotopologues) are given, but no information on structural isomers can be obtained therewith [30,31].

By coupling chromatography with HR-MS instruments (e.g., liquid chromatography, LC-MS), the level of information on astrochemical complex mixtures would be increased [32]. Whereas mass spectrometry represents a physical analytical technique (determining the weight of a molecule), chromatography directly provides insights into chemical properties. The temporal dimension (retention time) gives information on the polarity of a molecule, by means of understanding the chemical interactions between analyte molecules and the chromatography's stationary phase. Thus, molecular isomers (compounds with the same exact molecular masses but different structures) can be distinguished [30,33]. The discrimination of isomers is not possible with DI-HR-MS only.

LC-HR-MS analyses generate information-rich and high-dimensional datasets that remain challenging to handle and to evaluate in a global, nontargeted manner [34]. This makes it difficult to describe a complex chemical fingerprint in a comprehensive manner. In complex astrochemical mixtures, each chromatographic data point represents a high resolution mass spectrum out of ten thousands of detected m/z signals. Thus, a proper data analytical methodology is critical but also crucial to generate meaningful results [19]. By the help of sophisticated data mining, these types of large, multivariate big data sets can be practically accessed. Furthermore, proper data visualization tools are required for interpretation and first-level comparison of different samples analyzed by high-resolution mass spectrometry [35,36].

Here, we will go one step further toward the analysis of chemical structures in complex astrochemical mixtures. Data-driven ultra high performance liquid chromatography (UPLC)-Orbitrap MS analysis of Murchison (CM2) and Allende (CV3) meteorite will be discussed. By that, we add a new analytical dimension (chromatography) to enable the mass spectrometric detection of isomers, next to the resolution of isobars by HR-MS. In this proof-of-concept study, we will (i) give the motivation of using LC-HR-MS in astro-/cosmochemistry, (ii) discuss different chemical families in Murchison and (iii) give a first-order comparison between Murchison and Allende soluble organic matter regarding their chemical isomers. Our aim is to develop a conceptual overview on the multi-level information provided by comprehensive UPLC-HR-MS analysis.

2. Results

2.1. Complexity of Astrochemical Organic Matter—The Need for Additional Analytical Dimensions

Extraterrestrial organic matter is highly diverse, likely resulting in millions of molecules [8,19,37]. 14,000 various molecular formulas were described, including up to 150 different signals of molecular compositions per nominal mass [8]. Nevertheless, these type of DI-HR-MS experiments only determine a molecular composition but do not give any insights on the presence of isomers or chemical structures in general (without MS/MS). Complementary to DI-HR-MS analyses [8,9], we analyze here Murchison soluble organic matter by UPLC-Orbitrap MS. The total ion current (TIC) 1D chromatogram is rich in signals which reflects the overall molecular diversity present in Murchison (Figure 1A).
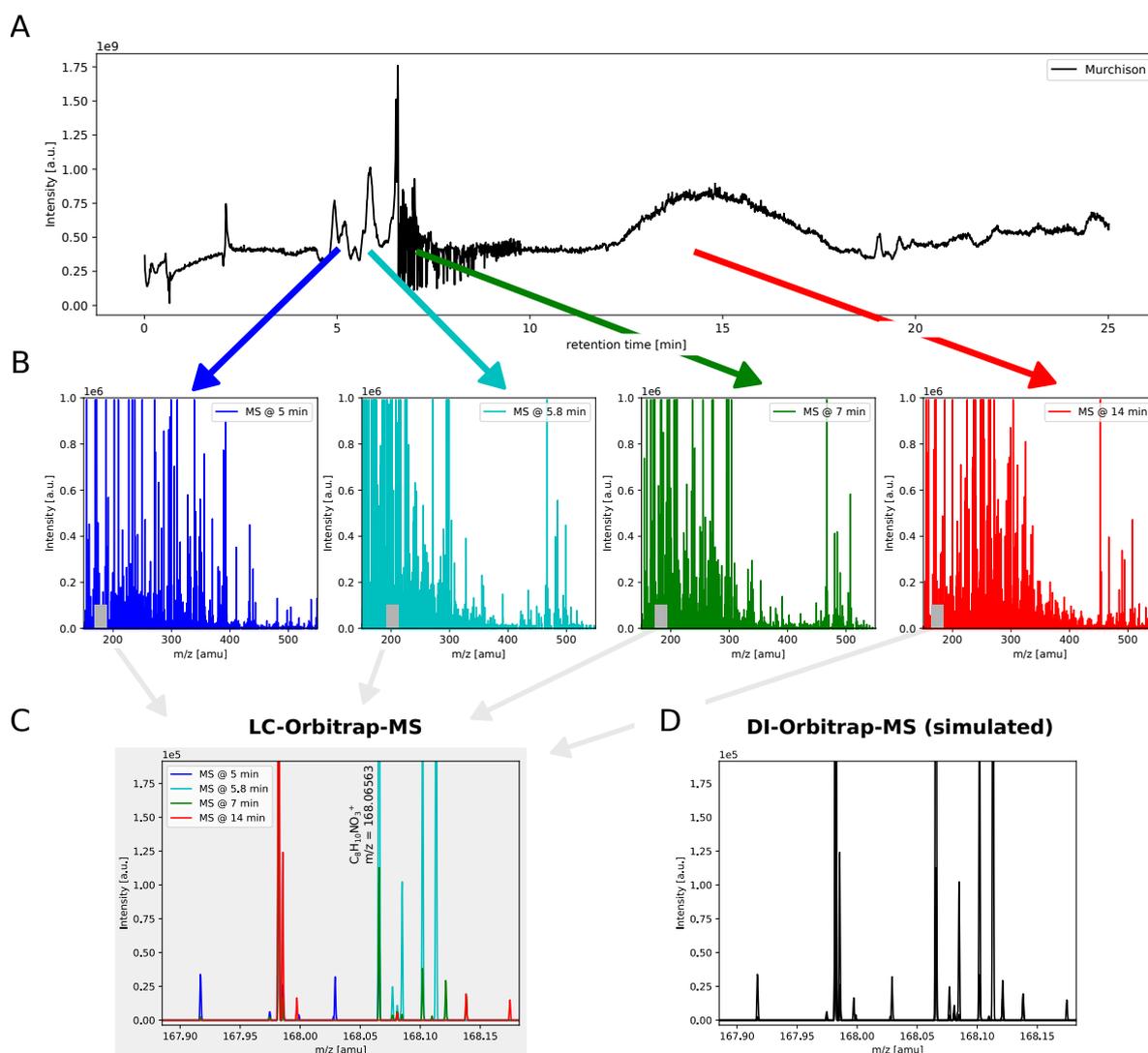
Figure 1. Complexity of astrochemical organic matter. (**A**) Positive ESI mode total ion current (TIC) 1D chromatogram of Murchison soluble organic matter; (**B**) different retention times result in different mass spectrometric fingerprints between 150 and 550 amu (atomic mass units); (**C**) the nominal mass m/z = 168 bears 17 different m/z signals, whose intensities vary for different retention times; (**D**) a simulated DI-Orbitrap mass spectrum of m/z = 168 amu doesn't give information on structural isomers. In the simulated DI mass spectrum, all mass spectra along the chromatographic run were averaged.

By adding the LC dimension to HR-MS, intensities of m/z signals vary for different retention times in a broad mass range between 150 and 550 amu (atomic mass units, Figure 1B). This gives

first insights into (i) the presence of isomers (intensity variation for the same exact mass in different time-resolved mass spectra, Figure 1C) and (ii) a general trend onto the relative chemical polarity scale of molecules (involving the identification of homologous series). Signals with small retention times bear higher polarity compared to the ones eluting at the end of a chromatographic run.

Thanks to chromatographic separation, the presence of isomers was elucidated in detail. As an example, for m/z = 168 amu, 17 different m/z signals could be detected (Figure 1C). The intensities of these mass peaks vary for different retention times. Thus, each m/z value corresponding to a molecular formula consists of various isomers. For example, m/z = 168.06549 corresponds to $C_8H_{10}NO_3^+$. Thirteen different isomers for this molecular formula have been detected in Murchison soluble organic matter. Because of their difference in chemical polarity, different isomers could be separated on a chromatographic column, while HR-MS only provides the resolution of isobars.

In comparison, direct infusion mass spectrometric (DI-HR-MS) experiments don't give any information on isomers (Figure 1D), as shown in the simulated direct infusion (DI) mass spectrum. The simulated DI mass spectrum was constructed by averaging all mass spectra along the chromatographic run. Thus, information on the retention time is vanished and only the differentiation of m/z signals (assignable to molecular compositions) remains. This directly demonstrates the importance of coupling LC to HR-MS.

In summary, DI-HR-MS gives the first level of information on the molecular diversity but only the coupling of MS with LC can describe isomers distinctly. This considerably increases our view onto the molecular diversity present in complex extraterrestrial samples.

2.2. Classifying Chemical Families—Bringing Order into the Chemical Diversity

The 2D chromatogram of Murchison soluble organic matter (m/z over retention time, Figure 2C), is rich in signals between m/z = 150–350 amu and from retention time = 5–20 min. Dealing with such a high degree of molecular diversity, it is important to systematically group chemical spaces. Here, chemical families were classified into CHO, CHNO, CHOS and CHNOS within LC-HR-MS data of Murchison. CHO, CHNO, CHOS and CHNOS compounds show different characteristics, both in m/z and in retention time (Figure 2C). This gives hints on chemical characteristics for each chemical class. In general, information on chemical polarity, structural isomers and homologous series can be obtained from this raw data representation. CHNO compounds represent the largest chemical subspace, distributed along the whole mass and retention time range which is in agreement with previous studies [3,8,10]. The significance of nitrogen chemistry in meteorites has been previously reported [38–45]. Otherwise, CHNOS compounds were found to group at fairly small retention times. This indicates a higher polarity of CHNOS species, compared to CHO, CHNO or CHOS classes. The contrary effect has been observed for CHO and CHOS compounds which elute toward the end of a chromatographic run suggesting more apolar properties.

Specific m/z vs. retention time windows were elucidated in more detail. Four spots were sampled and result in different van Krevelen diagrams (Figure 2A,B,E,F). Spot 1 represents molecules of highest polarity in Murchison (Figure 2A). They range from m/z 150–550 amu and mostly consist of CHNO and CHNOS species. Chemically, they show a high diversity in the degree of unsaturation (H/C ratio) indicating the presence of both aliphatic and unsaturated structures. In addition, the degree of oxygenation (O/C ratio) varies from 0.1 until 1.2. In comparison to spot 1, spot 2 represented more apolar compounds at high molecular masses (300–500 amu, Figure 2C). Chemical polarity can scale with the degree of oxygenation. This is reflected by a less widespread O/C ratio of spot 2 compounds when compared to those in spot 1 (O/C (average, spot 1) = 0.3, O/C (average, spot 2) = 0.2, Figure 2A,B). In addition, the H/C ratio is also narrower for higher apolar compounds (spot 2) in contrast to spot 1 compounds (H/C (average, spot 1) = 1.3, H/C (average, spot 2) = 1.5, Figure 2A,B). Spot 3 and spot 4 represent the same m/z range (200–300 amu) but differ in their retention time windows (Figure 2C). Spot 4 elutes later in the chromatographic run (higher retention time), thus indicating the presence of more apolar compounds. van Krevelen diagrams represent similar chemical

characteristics for both molecular subsets (Figure 2E,F). These two chemical subspaces consist of many structural isomers (same m/z value but different retention time) which cannot be discriminated in terms of van Krevelen-type representations (atomic ratio plots).
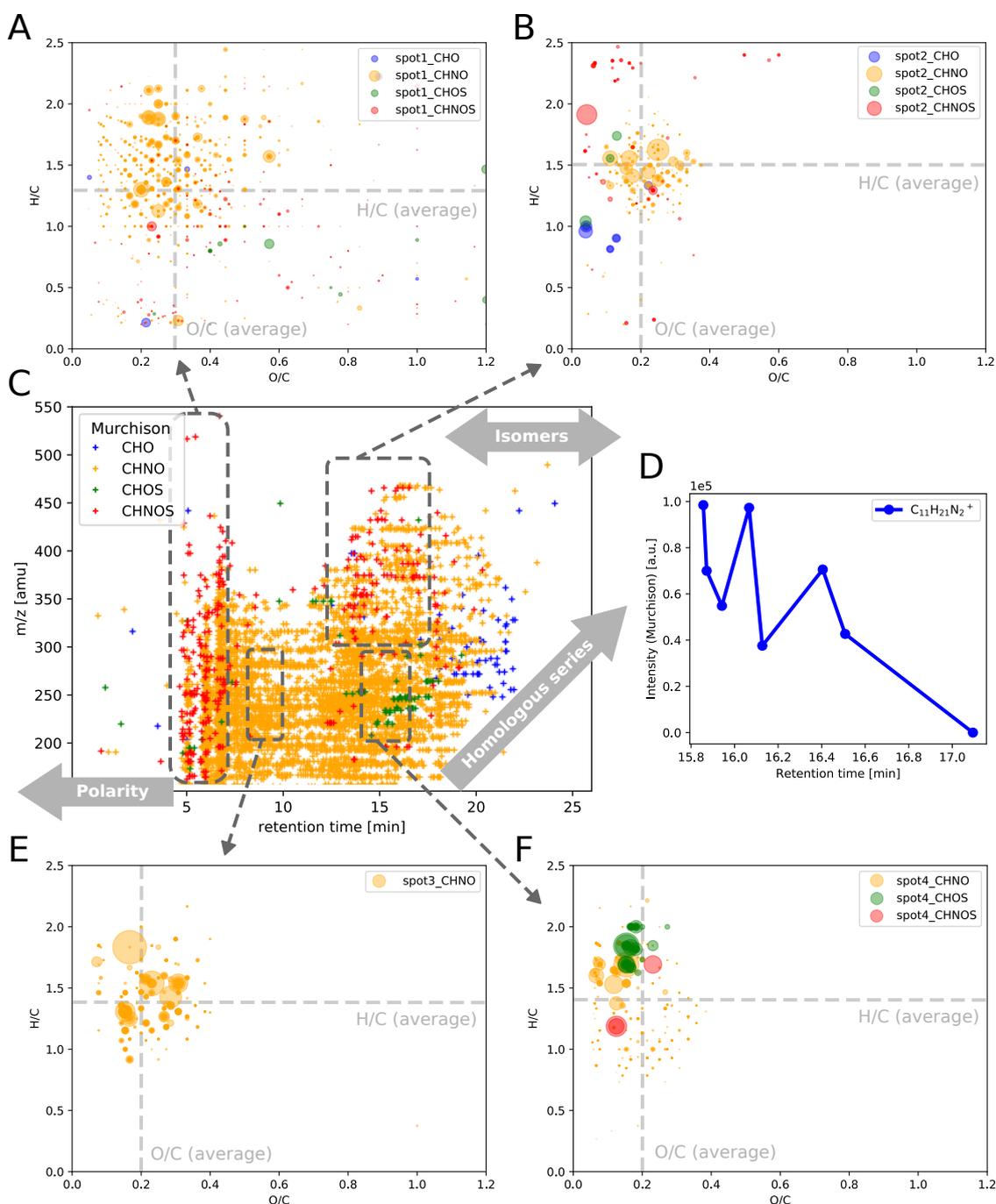
Figure 2. Classifying chemical families. (**A**) van Krevelen diagrams of spot 1 (RT = 4–7 min, m/z = 150–700 amu); (**B**) van Krevelen diagrams of spot 2 (RT = 12.5–17.5 min, m/z = 300–500 amu); (**C**) 2D chromatogram (m/z over retention time) of Murchison soluble organic matter depicts its chemical diversity in these two dimensions but also shows the grouping of distinct chemical subspaces CHO, CHNO, CHOS and CHNOS; (**D**) the resolved detection of eight $C_{11}H_{21}N_2^+$ isomers in Murchison gives first insights on the different chemical polarity of these compounds. (**E**) van Krevelen diagrams of spot 3 (RT = 8–10 min, m/z = 200–300 amu); (**F**) van Krevelen diagrams of spot 4 (RT = 14–17 min, m/z = 200–300 amu). Bubble sizes in van Krevelen diagrams are scaled with mass spectrometric intensites.

In-depth analyses on the number of isomers show that Murchison soluble organic matter bears a high number of isomers per molecular composition. As an example, eight different isomers of $C_{11}H_{21}N_2^+$ molecular compositions could be chromatographically resolved and separately detected. Different $C_{11}H_{21}N_2^+$ isomers showed variations in their detected mass spectrometric (+)-ESI intensities (Figure 2E). This intensity distribution might indicate differences in isomer stabilities (assuming similar ionization efficiency of the respective isomers).

These results underline the importance of studying isomers to get further insights into chemical structures in complex astrochemical mixtures. In the next section, we further study the isomeric diversity of Murchison and Allende in a more comprehensive manner.

2.3. Comparison of Murchison with Allende—A Difference in Isomeric Diversity

To illustrate the power of comprehensive UPLC-Orbitrap MS analyses in complex extraterrestrial mixtures, we compared Murchison with Allende meteorite in a coarse, first-order manner. Murchison represents a primitive CM2 carbon-rich carbonaceous chondrite [46]. It has become a reference material for extraterrestrial organic matter [8,10,13,47,48]. This CM2-type meteorite has been reported to have experienced weak secondary, aqueous alteration processes [49,50]. Allende (CV3 chondrite) is assigned as a petrologic type 3 meteorite and has undergone thermal metamorphism, indicating a modification/alteration of its organic matter [51]. Compared to Murchison, Allende has lower organic contents [52]. Anyhow, a rigorous direct comparison between Murchison and Allende in terms of chemical evolution is difficult. Both objects are supposed to have different parent bodies and might be different in their primary organics.

The molecular diversity of Murchison soluble organic matter rules the overall chemical space of the two samples (blue crosses, Figure 3A). Allende soluble organic matter seems to be more diverse in m/z range (red crosses, Figure 3A). Whereas Murchison organic matter shows a continuum in chemical polarity, Allende is characterized by more distinct properties. Allende bears either fairly polar molecules (retention time 5–7 min) or apolar compounds, widespread over a large mass range (retention time 14–24 min, m/z 150–600 amu).

The unique power of UPLC-Orbitrap MS analyses is the combination of accurate mass measurement with resolving structural isomers. By monitoring isomer counts for both Murchison and Allende meteoritic soluble organic matter, a significant difference in isomeric diversity has been revealed (Figure 3B). Unique features in Murchison are present with a maximum of 31 different isomers. The number of isomer counts decreases exponentially which might indicate a statistical process in the diversification of molecular structures (Supplementary Materials, Figure S3). Allende is much less diverse in isomers than Murchison. Only up to three different isomers were found herein. Both isomeric distributions differ in their mass ranges. Most Murchison isomer candidates were found at small m/z values (200–300 amu). In contrast, Allende isomers were present over a larger mass scale (350–650 amu). Based on this finding, ≈440,000 different organic compounds could be present in Murchison. The estimated number of 440,000 different compounds is based on the following extrapolation: 14,197 molecular compositions [8] × 31 different isomers.

The top 15 molecular compositions with the highest diversity in isomers have both large and small differences in respective retention times (Figure 3C,D). Herein, the top 15 molecular compositions sorted by the number of isomers are depicted. Both isomeric sets of Murchison and Allende reflect different chemical characteristics (Figure 3E,F). Murchison isomers show more similar properties than those of Allende. In Murchison, the extracted top 15 compositions, representing the most diverse isomers, are exclusively CHNO compounds in a fairly narrow H/C and m/z range (H/C = 1–1.7, m/z = 180–300 amu). On the contrary, Allende isomers are widespread from H/C = 0.5–2.5 and m/z = 200–700 amu. In addition, Allende isomers consist of CHO, CHNO and CHNOS molecules.

Murchison and Allende are significantly distinct in their sets of chemical isomers. Murchison bears a significant higher number of isomers, while their chemical characteristics seem to be less diverse than those of Allende. This finding might give hints on the different secondary processing

(aqeuous alteration/thermal metamorphism) in Murchison and Allende and/or in their parent bodies. These findings are consistent with previous studies, indicating an inverse relation between molecular diversity and degree of meteoritic alteration [8,9,44].
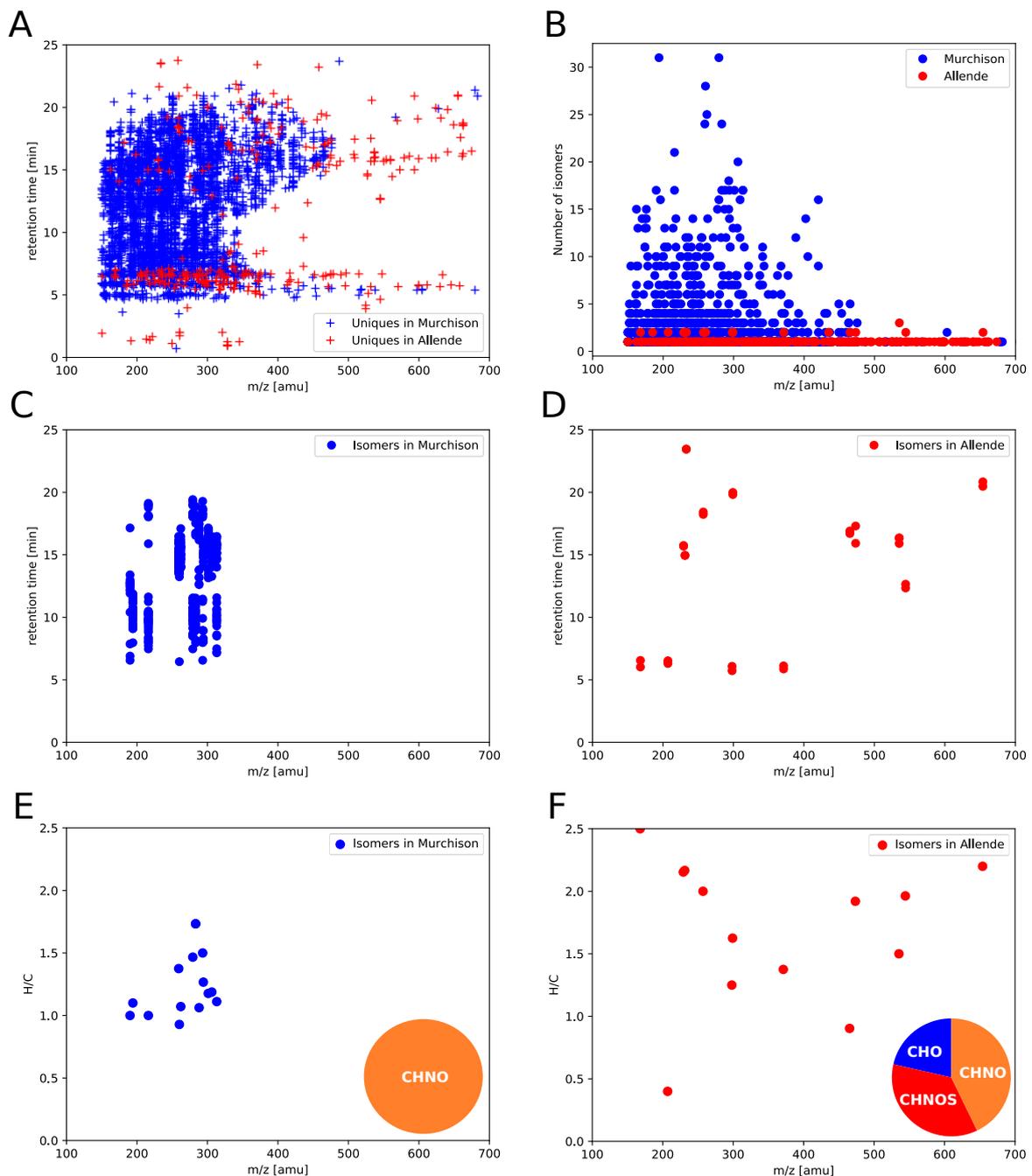
Figure 3. Comparison of Murchison with Allende. (**A**) 2D chromatogram highlighting the compound distributions of Murchison and Allende; (**B**) number of isomers as a function of m/z; (**C,D**) 2D chromatogram of the top 15 molecular compositions sorted by the number of isomers in Murchison (**C**) and Allende (**D**); (**E,F**) mass-edited van Krevelen diagrams of the top 15 molecular compositions sorted by the number of isomers in Murchison (**E**) and Allende (**F**). In all plots, only unique features present in Murchison or Allende, respectively.

3. Materials and Methods

3.1. Extraction of Meteorites

All chemicals were purchased from Sigma-Aldrich, Steinheim, Germany. Purified water was generated by a Purelab UHQ II water purification System (ELGA Lab Water, Veolia, Paris, France). Both meteorites were provided by Musée Cantonal de Géologie (Lausanne, Switzerland). Glassware used in sample processing were pyrolyzed at 500 °C in air overnight prior to use. For both Murchison and Allende, a sample (1.5 g) was sonicated with water (2.5 mL) at 50 °C for 20 h. The resulting solution was recovered by centrifugation (4500 rotations per minute (rpm), 10 min) and discarded. The sample (approximately 500 mg) was then powdered in an agate mortar and extracted with water (2 mL) at 100 °C for 20 h in an evacuated and N_2 purged vial. After centrifugation, the supernatant was removed, dried down with a speedvac and resuspended in water (2 mL). One millilitre aliquot of the resuspended sample was desalted through a cation-exchange resin (Dowex 50W-X8 resin, 50–100 mesh, hydrogen form) and the eluates of ammonium hydroxide solution were collected and dried under vacuum. It was stored at −20 °C until use. It was dissolved in 100 µL of purified water and diluted in methanol (1/3) for LC-MS analysis.

3.2. UPLC-Orbitrap MS Analysis

Analyses were performed with two quaternary Accela LC pumps (pump 1: Accela 600 pump; pump 2: Accela 1250 pump) working together and interfaced with a Q-Exactive (Hybrid quadrupole-OrbitrapTM) mass spectrometer equipped of an HESI source (Thermo Fisher Scientific, Waltham, MA, USA). Before injection, samples were stored at 4 °C using a Stack cooler CW (CTC Analytics AG, Zwingen, Switzerland). MS functions and LC solvent gradients were controlled by the Xcalibur data system (Thermo Fisher Scientific). Acetonitrile with 0.1% formic acid was used as buffer A and water with 0.1% formic acid as buffer B. Instrument calibration in positive mode was done every day via the direct infusion of positive ions calibration solutions. Compounds elution was performed via a Dual-LC configuration composed of two distinct units that are each composed of two stationary phases coupled in series. Unit 1 corresponds to the dual-trap unit. It combines a semi-polar Hypersil Gold aQ (20 mm × 2.1 mm, 12 µm, 175 Å; Thermo Fisher Scientific) and a porous graphitic carbon Hypercarb (20 mm × 2.1 mm, 7 µm, 250 Å; Thermo Fisher Scientific). They are connected to a second analytical unit which combines a Pentafluorophenyl reverse-phase Kinetex F5 (100 mm × 2.1 mm, 1.7 µm, 100 Å, Phenomenex) and a C18 reverse-phase Hypersil Gold Q (50 mm × 1 mm, 1.9 µm, 175 Å, Thermo Fisher Scientific)). One thousand microliters of the sample are first transferred to the dual trap unit with water—0.1% formic acid for 2 min at 100 µL·min^{-1}. Once the loading completed, the second elution phase (5% buffer A and 95% buffer B) passes through the whole system for 4 min at 100 µL·min^{-1}. Trapped analytes are then refocused and eluted into the serially coupled columns Kinetex F5—Gold Q in a backflush mode to narrow the peaks. Molecules are separated in the series of analytical columns through two successive gradient steps; the first one reaching 30% buffer A and 70% buffer B in 4 min and the second one reaching 100% buffer A in 12 min. The columns are then rinsed with 100% buffer A during 3 min and reconditioned for 4 min with 5% buffer A and 95% buffer B.

Mass detection was performed in positive ion mode with a full scan mode with a resolution of 140,000. Electrospray voltage was set at 4.0 kV. Capillary and heater temperatures were 275 °C and 300 °C. The relative flow rate of the sweep gas (nitrogen) which aids in solvent declustering and adduct reduction was set to 10 (flow in arbitrary units as defined by Thermo Fisher Scientific). To help nebulise sample solution into a fine mist in the ESI nozzle, the relative flow rate of the sheath gas (nitrogen) was set to 35 (flow in arbitrary units as defined by Thermo Fisher Scientific). Finally, the relative flow rate of the auxiliary gas (nitrogen) which assists the sheath gas in dispersing and/or evaporating sample solution was set at 20 (flow in arbitrary units as defined by Thermo Fisher Scientific). Instrument calibration in positive and negative mode was done every 3 days via the direct infusion of positive and negative ion calibration solutions. Peaks integration and MS spectra acquisition were performed with Thermo Xcalibur™ Qualitative Browser (Thermo Xcalibur 2.2 SP1.48, Thermo Fisher Scientific Inc.).

We have used solvent blanks which signatures were subtracted from endogenous analyte signals from each sample (Murchison and Allende). The method used here was optimized to study the molecular diversity of astrochemical samples [53].

3.3. Data Analysis

The whole LC-HR-MS data analysis workflow (Figure 4) was performed using MZmine 2.37 [54]. All computations were done on a stand-alone computer. Further processing (calculations and filtering) were done in Python and in LibreOffice Calc. The major part of this work is expressed by handling the complexity of LC-HR-MS big data sets. It is a challenge to mine these highdimensional data resulting from $\approx$40,000 m/z signals per mass spectrum, multiplied by $\approx$2600 chromatographic data points, multiplied by the number of samples (here 3, Blank, Murchison and Allende). Thus, $\approx 3 \times 10^8$ m/z signals need be handled. The treatment is problematic, both in computational costs and in terms of error propagation. Thus, a strict validation procedure to test for errors along the whole data analytical workflow is absolutely necessary. The preprocessing procedure (raw file toward unfiltered matrix) was tested to three types of errors; (i) errors in m/z during matrix generation, (ii) errors in retention time (RT) during matrix generation and (iii) error in m/z during molecular formula assignment. Errors in m/z (no i and iii) are mainly within $\pm$ 2.5 ppm (Supporting Materials, Figures S1 and S2). Thus, a m/z tolerance of 5 ppm within the whole workflow is justified. Within formula assignment, 54.4% of all formulas have an error of $\pm$0.2 ppm, 90.3% of formula assignments within $\pm$0.5 ppm and 99.6% of all formulas within an error window $\pm$1 ppm.

The data analytical workflow consists of two main parts, (i) a preprocessing step and (ii) a filtering step (Figure 4). First, Thermo RAW files were converted via ProteoWizard msconvert into a mzML format [55,56]. Then, mzML data were processed in MZmine 2.37 [54]. Within MZmine, the workflow can be mainly divided into three steps, (i) feature detection, (ii) sample comparison and (iii) identification of m/z signals. The finalized preprocessing procedure ended in a matrix including 429,040 features (Figure 4B). This first part of analysis was tested to three types of errors; (i) errors in m/z during matrix generation, (ii) errors in retention time (RT) during matrix generation and (iii) error in m/z during molecular formula assignment. Details are given in the Supplementary Materials section. The second main part of the data processing workflow, filtering invalid features, consisted of four steps: (i) not assigned m/z features (e.g., noise peaks); (ii) testing for the validation of a molecular formula according Senior's rules [57] and with respect to the Seven Golden Rules [58]; (iii) remaining features were removed which were in common with solvent blanks. Finally, 6760 valid features containing endogenous signals from Murchison and Allende remained for further analysis (Figure 4C).
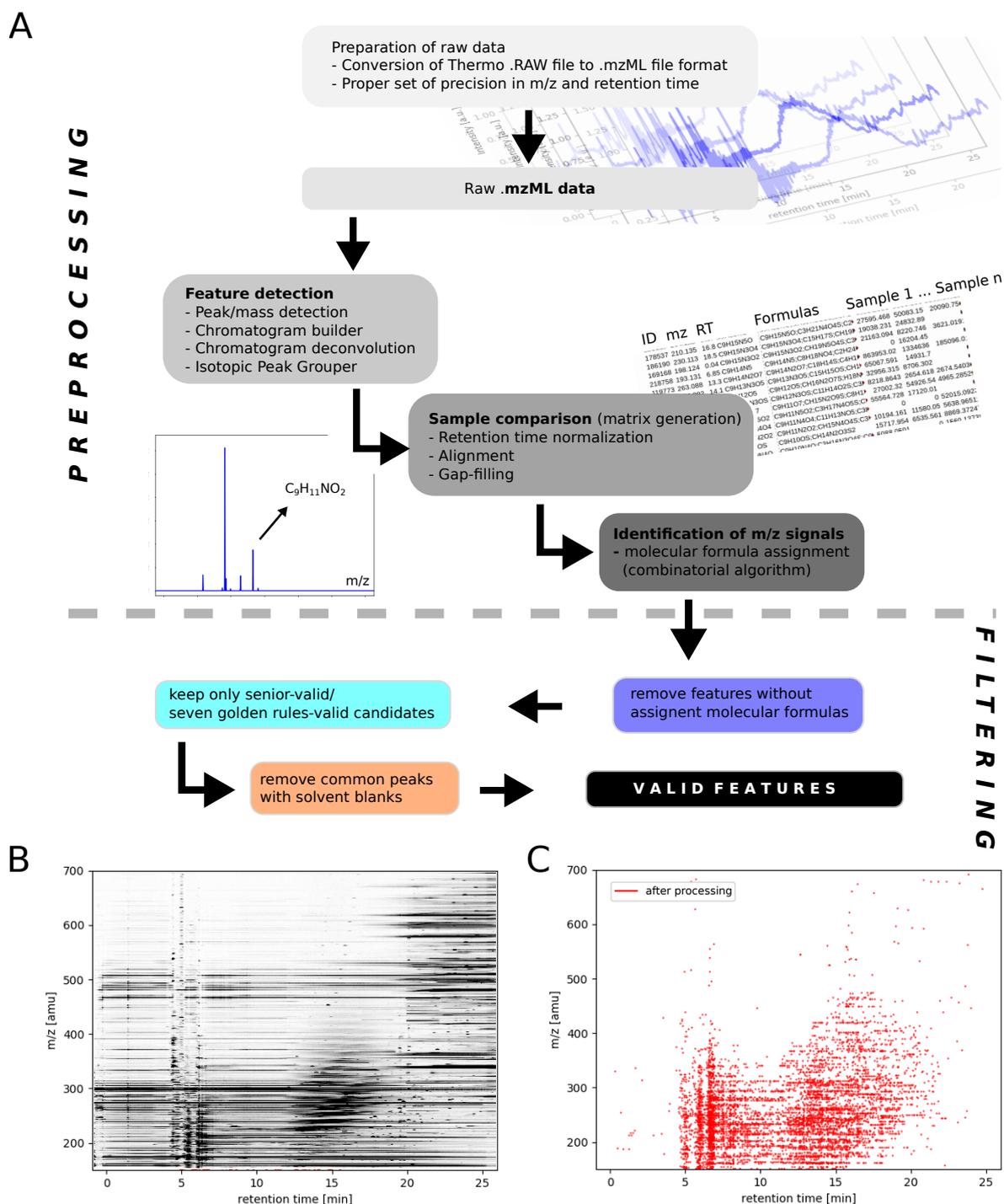
Figure 4. Data processing workflow—a protocol to merge highdimensional LC-HR-MS data. (**A**) the filtering and optimized process of this data set reduced the data set from ≈400,000 features to ≈7000 features; (**B**,**C**) *m/z* over retention time plot shows the difference in *m/z* signals before and after processing.

4. Conclusions

We present in-depth data-analytical strategies to comprehensively evaluate UPLC-Orbitrap MS analysis. As a proof-of-concept study, we studied Murchison and Allende meteorites. These findings indicate the power of this type of analysis but also state the challenge in terms of big data processing.

In summary, (i) LC-HR-MS extends the level of information of astrochemical molecular diversity, (ii) a relative polarity scale of compounds out of complex mixtures can be determined, (iii) structural

isomers can be resolved by chromatography and separately detected; (iv) and the combination of UPLC and HR-MS merges both accurate isomer and isobar detection.

By means of this time-resolving high resolution mass spectrometric method, we go one step further toward the characterization of chemical structures within complex extraterrestrial mixtures, enabling a better understanding of organic chemical evolution.

References

1. Öberg, K.I. Photochemistry and astrochemistry: Photochemical pathways to interstellar complex organic molecules. *Chem. Rev.* **2016**, *116*, 9631–9663. [CrossRef]

2. Danger, G.; Orthous-Daunay, F.R.; de Marcellus, P.; Modica, P.; Vuitton, V.; Duvernay, F.; Flandinet, L.; d'Hendecourt, L.L.S.; Thissen, R.; Chiavassa, T. Characterization of laboratory analogs of interstellar/cometary organic residues using very high resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2013**, *118*, 184–201.

3. Danger, G.; Fresneau, A.; Mrad, N.A.; De Marcellus, P.; Orthous-Daunay, F.R.; Duvernay, F.; Vuitton, V.; d'Hendecourt, L.L.S.; Thissen, R.; Chiavassa, T. Insight into the molecular composition of laboratory organic residues produced from interstellar/pre-cometary ice analogues using very high resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2016**, *189*, 184–196. [CrossRef]

4. Fresneau, A.; Mrad, N.A.; d'Hendecourt, L.L.; Duvernay, F.; Flandinet, L.; Orthous-Daunay, F.R.; Vuitton, V.; Thissen, R.; Chiavassa, T.; Danger, G. Cometary Materials Originating from Interstellar Ices: Clues from Laboratory Experiments. *Astrophys. J.* **2017**, *837*, 168. [CrossRef]

5. Altwegg, K.; Balsiger, H.; Bar-Nun, A.; Berthelier, J.J.; Bieler, A.; Bochsler, P.; Briois, C.; Calmonte, U.; Combi, M.R.; Cottin, H.; et al. Prebiotic chemicals—Amino acid and phosphorus—In the coma of comet 67P/Churyumov–Gerasimenko. *Sci. Adv.* **2016**, *2*, e1600285. [CrossRef]

6. Fray, N.; Bardyn, A.; Cottin, H.; Altwegg, K.; Baklouti, D.; Briois, C.; Colangeli, L.; Engrand, C.; Fischer, H.; Glasmachers, A.; et al. High-molecular-weight organic matter in the particles of comet 67P/Churyumov–Gerasimenko. *Nature* **2016**, *538*, 72–74. [CrossRef] [PubMed]

7. Sandford, S.A.; Aléon, J.; Alexander, C.M.; Araki, T.; Bajt, S.; Baratta, G.A.; Borg, J.; Bradley, J.P.; Brownlee, D.E.; Brucato, J.R.; et al. Organics captured from comet 81P/Wild 2 by the Stardust spacecraft. *Science* **2006**, *314*, 1720–1724. [CrossRef]

8. Schmitt-Kopplin, P.; Gabelica, Z.; Gougeon, R.D.; Fekete, A.; Kanawati, B.; Harir, M.; Gebefuegi, I.; Eckel, G.; Hertkorn, N. High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2763–2768. [CrossRef] [PubMed]

9. Ruf, A.; Kanawati, B.; Hertkorn, N.; Yin, Q.Z.; Moritz, F.; Harir, M.; Lucio, M.; Michalke, B.; Wimpenny, J.; Shilobreeva, S.; et al. Previously unknown class of metalorganic compounds revealed in meteorites. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2819–2824. [CrossRef]

10. Hertkorn, N.; Harir, M.; Schmitt-Kopplin, P. Nontarget analysis of Murchison soluble organic matter by high-field NMR spectroscopy and FTICR mass spectrometry. *Magn. Reson. Chem.* **2015**, *53*, 754–768. [CrossRef]

11. Caro, G.M.; Meierhenrich, U.; Schutte, W.; Barbier, B.; Segovia, A.A.; Rosenbauer, H.; Thiemann, W.P.; Brack, A.; Greenberg, J. Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **2002**, *416*, 403–406. [CrossRef]

12. Meinert, C.; Myrgorodska, I.; De Marcellus, P.; Buhse, T.; Nahon, L.; Hoffmann, S.V.; d'Hendecourt, L.L.S.; Meierhenrich, U.J. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science* **2016**, *352*, 208–212. [CrossRef]

13. Elsila, J.E.; Aponte, J.C.; Blackmond, D.G.; Burton, A.S.; Dworkin, J.P.; Glavin, D.P. Meteoritic amino acids: Diversity in compositions reflects parent body histories. *ACS Cent. Sci.* **2016**, *2*, 370–379. [CrossRef]

14. Cooper, G.; Kimmich, N.; Belisle, W.; Sarinana, J.; Brabham, K.; Garrel, L. Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature* **2001**, *414*, 879–883. [CrossRef] [PubMed]

15. Callahan, M.P.; Smith, K.E.; Cleaves, H.J.; Ruzicka, J.; Stern, J.C.; Glavin, D.P.; House, C.H.; Dworkin, J.P. Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13995–13998. [CrossRef]

16. 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* **2012**, *77*, 186–191. [CrossRef]

17. Kaiser, R.; Stockton, A.; Kim, Y.; Jensen, E.; Mathies, R. On the formation of dipeptides in interstellar model ices. *Astrophys. J.* **2013**, *765*, 111. [CrossRef]

18. Ligterink, N.; Terwisscha van Scheltinga, J.; Taquet, V.; Jørgensen, J.; Cazaux, S.; van Dishoeck, E.; Linnartz, H. The formation of peptide-like molecules on interstellar dust grains. *Mon. Not. R. Astron. Soc.* **2018**, *480*, 3628–3643. [CrossRef]

19. Ruf, A.; d'Hendecourt, L.; Schmitt-Kopplin, P. Data-Driven Astrochemistry: One Step Further within the Origin of Life Puzzle. *Life* **2018**, *8*, 18. [CrossRef] [PubMed]

20. Hordijk, W.; Steel, M.; Kauffman, S.A. Molecular Diversity Required for the Formation of Autocatalytic Sets. *Life* **2019**, *9*, 23. [CrossRef]

21. Gibb, E.; Whittet, D.; Boogert, A.; Tielens, A. Interstellar ice: The infrared space observatory legacy. *Astrophys. J. Suppl. Ser.* **2004**, *151*, 35. [CrossRef]

22. McCarthy, M.; Chen, W.; Travers, M.; Thaddeus, P. Microwave spectra of 11 polyyne carbon chains. *Astrophys. J. Suppl. Ser.* **2000**, *129*, 611. [CrossRef]

23. Hewins, R.H.; Bourot-Denise, M.; Zanda, B.; Leroux, H.; Barrat, J.A.; Humayun, M.; Göpel, C.; Greenwood, R.C.; Franchi, I.A.; Pont, S.; et al. The Paris meteorite, the least altered CM chondrite so far. *Geochim. Cosmochim. Acta* **2014**, *124*, 190–222. [CrossRef]

24. Goesmann, F.; Rosenbauer, H.; Bredehöft, J.H.; Cabane, M.; Ehrenfreund, P.; Gautier, T.; Giri, C.; Krüger, H.; Le Roy, L.; MacDermott, A.J.; et al. Organic compounds on comet 67P/Churyumov–Gerasimenko revealed by COSAC mass spectrometry. *Science* **2015**, *349*, aab0689. [CrossRef]

25. Somogyi, Á.; Thissen, R.; Orthous-Daunay, F.R.; Vuitton, V. The role of ultrahigh resolution fourier transform mass spectrometry (FT-MS) in astrobiology-related research: Analysis of meteorites and tholins. *Int. J. Mol. Sci.* **2016**, *17*, 439. [CrossRef]

26. Briois, C.; Thissen, R.; Thirkell, L.; Aradj, K.; Bouabdellah, A.; Boukrara, A.; Carrasco, N.; Chalumeau, G.; Chapelon, O.; Colin, F.; et al. Orbitrap mass analyser for in situ characterisation of planetary environments: Performance evaluation of a laboratory prototype. *Planet. Space Sci.* **2016**, *131*, 33–45. [CrossRef]

27. Selliez, L.; Briois, C.; Carrasco, N.; Thirkell, L.; Thissen, R.; Ito, M.; Orthous-Daunay, F.R.; Chalumeau, G.; Colin, F.; Cottin, H.; et al. Identification of organic molecules with a laboratory prototype based on the Laser Ablation-CosmOrbitrap. *Planet. Space Sci.* **2019**, *170*, 42–51. [CrossRef]

28. Tsuda, Y.; Yoshikawa, M.; Abe, M.; Minamino, H.; Nakazawa, S. System design of the Hayabusa 2—Asteroid sample return mission to 1999 JU3. *Acta Astronaut.* **2013**, *91*, 356–362. [CrossRef]

29. Lauretta, D.; Balram-Knutson, S.; Beshore, E.; Boynton, W.V.; d'Aubigny, C.D.; DellaGiustina, D.; Enos, H.; Golish, D.; Hergenrother, C.; Howell, E.; et al. OSIRIS-REx: Sample return from asteroid (101955) Bennu. *Space Sci. Rev.* **2017**, *212*, 925–984. [CrossRef]

30. Forcisi, S.; Moritz, F.; Lucio, M.; Lehmann, R.; Stefan, N.; Schmitt-Kopplin, P. Solutions for low and high accuracy mass spectrometric data matching: A data-driven annotation strategy in nontargeted metabolomics. *Anal. Chem.* **2015**, *87*, 8917–8924. [CrossRef]

31. Moritz, F. Deep Metabotyping of Exhaled Breath Condensate (EBC)–Characterization of Surrogate Markers for Systemic Metabolism and Non-Invasive Diagnostics in Diabetes. Ph.D. Thesis, Technische Universität München, München, Germany, 2014.

32. Gautier, T.; Schmitz-Afonso, I.; Touboul, D.; Szopa, C.; Buch, A.; Carrasco, N. Development of HPLC-Orbitrap method for identification of N-bearing molecules in complex organic material relevant to planetary environments. *Icarus* **2016**, *275*, 259–266. [CrossRef]

33. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt-Kopplin, P. Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *J. Chromatogr. A* **2013**, *1292*, 51–65. [CrossRef]

34. Tugizimana, F.; Steenkamp, P.; Piater, L.; Dubery, I. A conversation on data mining strategies in LC-MS untargeted metabolomics: Pre-processing and pre-treatment Steps. *Metabolites* **2016**, *6*, 40. [CrossRef]

35. Van Krevelen, D. Graphical-statistical method for the study of structure and reaction processes of coal. *Fuel* **1950**, *29*, 269–284.

36. Tziotis, D.; Hertkorn, N.; Schmitt-Kopplin, P. Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: Improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur. J. Mass Spectrom.* **2011**, *17*, 415–421. [CrossRef]

37. Kwok, S. *Organic Matter in the Universe*; John Wiley & Sons: Hoboken, NJ, USA, 2011.

38. Popova, O.P.; Jenniskens, P.; Emel'yanenko, V.; Kartashova, A.; Biryukov, E.; Khaibrakhmanov, S.; Shuvalov, V.; Rybnov, Y.; Dudorov, A.; Grokhovsky, V.I.; et al. Chelyabinsk airburst, damage assessment, meteorite recovery, and characterization. *Science* **2013**, *342*, 1069–1073. [CrossRef]

39. Schmitt-Kopplin, P.; Harir, M.; Kanawati, B.; Tziozis, D.; Hertkorn, N.; Gabelica, Z. Chemical footprint of the solvent soluble extraterrestrial organic matter occluded in Sołtmany ordinary chondrite. *Meteorites* **2012**, *2*, 79–92.

40. Bartoschewitz, R.; Appel, P.; Barrat, J.A.; Bischoff, A.; Caffee, M.W.; Franchi, I.A.; Gabelica, Z.; Greenwood, R.C.; Harir, M.; Harries, D.; et al. The Braunschweig meteorite- a recent L6 chondrite fall in Germany. *Chemie der Erde* **2017**, *77*, 207–224. [CrossRef]

41. Jenniskens, P.; Rubin, A.E.; Yin, Q.Z.; Sears, D.W.; Sandford, S.A.; Zolensky, M.E.; Krot, A.N.; Blair, L.; Kane, D.; Utas, J.; et al. Fall, recovery, and characterization of the Novato L6 chondrite breccia. *Meteorit. Planet. Sci.* **2014**, *49*, 1388–1425. [CrossRef]

42. Yamashita, Y.; Naraoka, H. Two homologous series of alkylpyridines in the Murchison meteorite. *Geochem. J.* **2014**, *48*, 519–525. [CrossRef]

43. 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.* **2017**, *1*, 540–550. [CrossRef]

44. Hashiguchi, M.; Naraoka, H. High-mass resolution molecular imaging of organic compounds on the surface of Murchison meteorite. *Meteorit. Planet. Sci.* **2019**, *54*, 452–468. [CrossRef]

45. Naraoka, H.; Hashiguchi, M. Distinct distribution of soluble N-heterocyclic compounds between CM and CR chondrites. *Geochem. J.* **2019**, *53*, 33–40. [CrossRef]

46. Sephton, M.A. Organic compounds in carbonaceous meteorites. *Nat. Prod. Rep.* **2002**, *19*, 292–311. [CrossRef] [PubMed]

47. Remusat, L.; Derenne, S.; Robert, F.; Knicker, H. New pyrolytic and spectroscopic data on Orgueil and Murchison insoluble organic matter: A different origin than soluble? *Geochim. Cosmochim. Acta* **2005**, *69*, 3919–3932. [CrossRef]

48. Cronin, J.; Pizzarello, S.; Epstein, S.; Krishnamurthy, R. Molecular and isotopic analyses of the hydroxy acids, dicarboxylic acids, and hydroxydicarboxylic acids of the Murchison meteorite. *Geochim. Cosmochim. Acta* **1993**, *57*, 4745–4752. [CrossRef]

49. Rubin, A.E.; Trigo-Rodríguez, J.M.; Huber, H.; Wasson, J.T. Progressive aqueous alteration of CM carbonaceous chondrites. *Geochim. Cosmochim. Acta* **2007**, *71*, 2361–2382. [CrossRef]

50. Vinogradoff, V.; Le Guillou, C.; Bernard, S.; Binet, L.; Cartigny, P.; Brearley, A.; Remusat, L. Paris vs. Murchison: Impact of hydrothermal alteration on organic matter in CM chondrites. *Geochim. Cosmochim. Acta* **2017**, *212*, 234–252. [CrossRef]

51. Kojima, T.; Tomeoka, K. Indicators of aqueous alteration and thermal metamorphism on the CV parent body: Microtextures of a dark inclusion from Allende. *Geochim. Cosmochim. Acta* **1996**, *60*, 2651–2666. [CrossRef]

52. Pearson, V.; Sephton, M.; Franchi, I.; Gibson, J.; Gilmour, I. Carbon and nitrogen in carbonaceous chondrites: Elemental abundances and stable isotopic compositions. *Meteorit. Planet. Sci.* **2006**, *41*, 1899–1918. [CrossRef]

53. Eddhif, B.; Allavena, A.; Liu, S.; Ribette, T.; Mrad, N.A.; Chiavassa, T.; d'Hendecourt, L.L.S.; Sternberg, R.; Danger, G.; Geffroy-Rodier, C.; et al. Development of liquid chromatography high resolution mass spectrometry strategies for the screening of complex organic matter: Application to astrophysical simulated materials. *Talanta* **2018**, *179*, 238–245. [CrossRef] [PubMed]

54. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]

55. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534–2536. [CrossRef] [PubMed]

56. Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W.H.; Römpp, A.; Neumann, S.; Pizarro, A.D.; et al. mzML—A community standard for mass spectrometry data. *Mol. Cell. Proteom.* **2011**, *10*, R110.000133. [CrossRef]

57. Senior, J.K. Partitions and their representative graphs. *Am. J. Math.* **1951**, *73*, 663–689. [CrossRef]

58. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform.* **2007**, *8*, 105. [CrossRef]