Characterization of bio-organic and inorganic chemistries using Laser-based Mass Spectrometry

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Rustam A. Lukmanov

from Russia

Supervisor of the doctoral thesis:
Prof. Dr. Peter Wurz

Space Research & Planetary Sciences (WP)
Physics institute, University of Bern

Characterization of bio-organic and inorganic chemistries using Laser-based Mass Spectrometry

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Rustam A. Lukmanov

from Russia

Supervisor of the doctoral thesis:
Prof. Dr. Peter Wurz
University of Bern

Accepted by the Faculty of Science.

Bern, 26 of August 2021

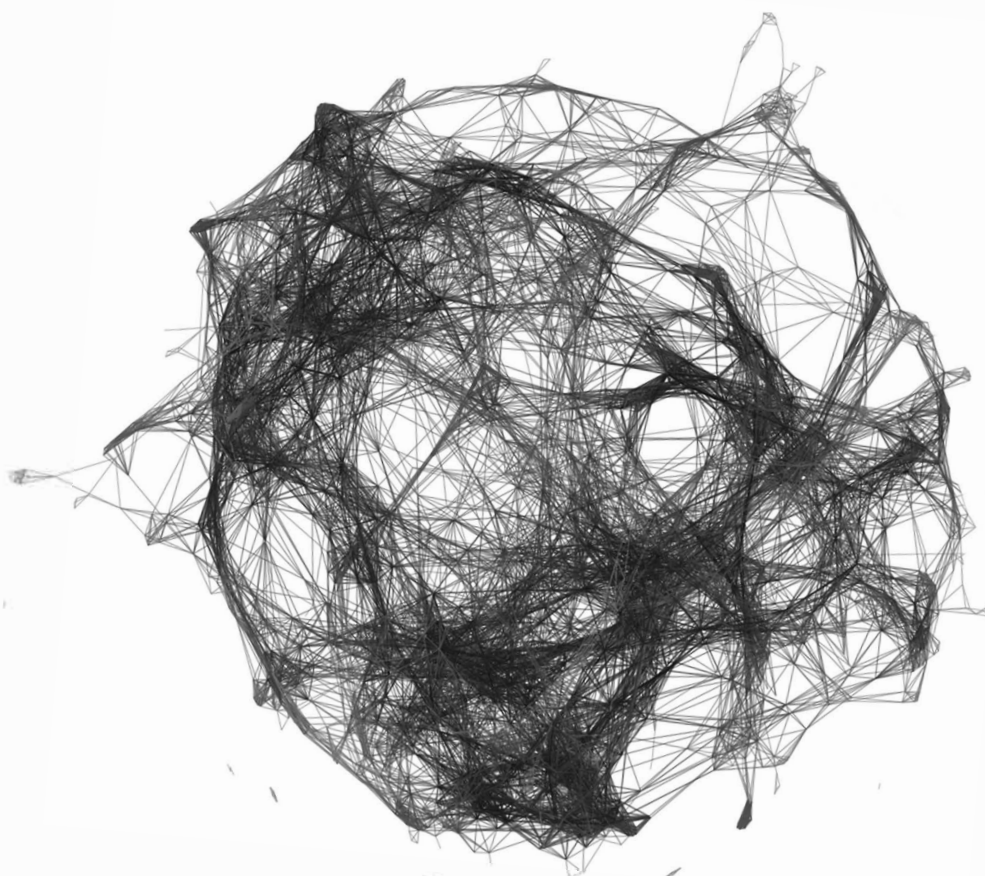The Dean
Prof. Dr. Zoltan Balogh

Acknowledgements

For my family

"The unlike is joined together, and from differences results the most beautiful harmony, and all things take place by strife"

Heraclitus, fragment 46

"During these vast, yet quite unknown, periods of time, the world swarmed with living creatures"

Charles Darwin, 1859

Abstract

This thesis contributes to the field of in-situ analytical chemistry by further expanding the measurement capabilities of a miniature laser ablation/ionization mass spectrometer (LIMS). The current version of LIMS used in this thesis was developed for in-situ space applications and represents the space-ready prototype of original size and figures of merit. The thesis expands on analytical problems related to the accurate measurement, classification, and identification of mass spectra registered from early (primitive) life and various minerals of inorganic origin. The investigated subjects cover specific aspects of ion generation in time-of-flight mass spectrometry, signal processing, and machine learning, applied to the large mass spectrometric data. The thesis outlines effective solutions for unsupervised characterization of compounds using graph theory and relational data analysis.

The current state of space exploration places the identification of signatures of life on Mars among the biggest challenges of our time and represents the frontier of science. The challenge of unambiguous and deterministic identification of biogenicity of the given sample (of unknown origin) is an outstanding problem that requires identification of the range of biological signatures (biosignatures) on the micrometer scale and demands the presence of unique patterns and characteristics that as a whole indicate biological processing.

This thesis explores the quality of chemical information that could be gathered from Precambrian microscopic fossils (microfossils) using three different wavelengths of the femtosecond laser radiation used as an ion source in the current LIMS system. The thesis discusses various aspects of mass spectrometric imaging and classification of spectra using the graph-theoretic approaches. Furthermore, the implementation of spectral similarity (proximity) networks will showcase the potential for the deterministic identification of bio-organic and inorganic chemistries. The current results provided in this thesis show high utility and perhaps, yet uncovered potential and importance of LIMS as an analytical method that can be used in future space exploration programs.

Keywords

Mass spectrometry / Mars / Space Instrumentation / Laser Ionization

Table of Contents

Introduction

Precis

In this chapter, I will introduce a miniature Laser Ionization Mass Spectrometer, a state-of-the-art instrument for characterization of chemical compositions of solids, and discuss its importance in the context of space exploration. Further, I will present principles of operation and technical characteristics of the instrument. I will also introduce the Gunflint sample as a realistic testbed for space instrumentation. Furthermore, I will present the outline of the thesis with research goals and objectives.

1.1 Prospect of finding signatures of life in a Solar system with laser-based mass spectrometry

Identification of the extinct and/or extant forms of life on Mars is of the highest priority in current space research and remains the biggest challenge. Satellite imaging and in-situ research have shown that in the early days of Martian evolution, liquid water and clement conditions, suitable for microbial life, were present at the surface of the planet (Arvidson et al., 2014; Fassett & Head, 2011). Moreover, recent studies reveal radar evidence of modern subglacial liquid water on Mars (Orosei et al., 2018), providing a realistic habitat for microbes living in the subsurface (Onstott et al., 2019). All these observations form a strong rationale for a search of biosignatures on the Red Planet.

However, robotic exploration of Mars presents several technological challenges: the size, weight, and power consumption of a scientific payload have to be sufficiently small and at the same time capable of delivering highly accurate scientific information. Traditionally used in space research techniques such as thermal volatilization (TV) – gas chromatography (GC) – mass spectrometry (MS), capable of detecting organic compounds, might not be sufficiently sensitive to detect weak signals from the micrometer-sized objects of a potentially biological origin. Moreover, such methods provide bulk sample analysis and alter the chemistry of the sample, which can make an interpretation of retrieved data ambiguous (Navarro-González et al., 2006). Therefore, there is a growing consensus in the expert field that the design of future element and organic composition detection instruments for Mars and icy moons of Jupiter should include local and sensitive methods that are able to detect extinct and/or extant life (Arevalo Jr et al., 2020; Goesmann et al., 2017; Vago et al., 2017; Westall et al., 2015).

Laser-based mass spectrometry – is a technique that combines the capability to detect organic compounds (Laser Desorption Mass Spectrometry – LDMS) (Ligterink et al., 2020), provide element and isotope analysis (Laser Ablation/Ionization Mass spectrometry – LIMS)(Rohner et al., 2003; Tulej et al., 2020; Wurz et al., 2020), and a capability to operate on a microscopic level (diameter of the analytical spot is ~5-10 μm), without the necessity for any sample preparation (Tulej et al., 2014). Intense monochromatic light (laser) and matter interaction constitute the core of the LIMS/LDMS methods. A single laser pulse is focused and shot at the surface of the sample: if the power density of laser pulse is sufficient, the analyzed material gets ablated (excavated) and reaches very high temperatures (to the point when matter gets atomized and partially ionized). Formed ions from the sample further experience separation and detection with a mass analyzer. This is a basic principle of LIMS analysis, where power densities of a single laser pulse reach TW/cm^2 (Tera – 10^{12}) levels, and a single laser shot provides an analysis of solids on a femtogram scale (Femto – 10^{-15}). However, if significantly lower laser power densities are applied (MW/cm^2 - GW/cm^2) to the surface of the solid material, it is possible to desorb molecules and molecular fragments and register them with a mass analyzer. These processes describe in a nutshell core working principles of the LDMS method.

So, what does it take to identify signatures of life on the surface of Mars? To answer this question, we have to look into the terrestrial fossil record, using the assumption that decay and preservation of organic compounds follow the same principles on Mars. Community of Precambrian (older than 541 Ma) microfossils from the Gunflint

---

*This chapter is based on original contribution: https://boris.unibe.ch/152616/8/Vet_News_2_20_web_Lukmanov.pdf

formation (1.88 Ga), Ontario, Canada, are considered among the best examples of early life. The morphological and chemical quality of the microfossils preserved in the Quartz (SiO_2) matrix provides an unrivaled insight into the existence of microbial life almost 2 billion years ago. Gunflint Formation reveals a view into an abundant and widespread, already complex, and diversified community comprising 12 separate species (Barghoorn & Tyler, 1965; Cloud, 1965; Wacey et al., 2013). There is still an ongoing debate about the diversity and metabolism of some of the identified species. However, it is largely accepted that oxygenic photosynthetic bacteria (cyanobacteria) were a dominant part of the Gunflint microbiome (Barghoorn & Tyler, 1965; Lepot et al., 2017). By studying the fossilized microbial communities, like one from the Gunflint formation, we can infer an element, isotope, molecular and morphological characteristics of truly biogenic organic remains, which altogether can indicate biological origin. Moreover, using the Gunflint microfossils as a benchmark, we can understand how well our instruments are performing relative to each other and recognize what needs to be improved.

Therefore, if instrumentation developed for space exploration will be capable of measuring micro-sized inclusions of bio-organic remains and distinguish them from inorganic chemistries, we may come a little closer to the ultimate goal of finding signatures of life on the Red Planet. While this goal can be formulated in a very short manner, the complexity of the task should not be misinterpreted, as bio-organic chemistry is comprised of the same elements as inorganic materials.

Additionally, considering that only a small fraction of the organic material can be preserved due to the effects of heat, pressure, and time, all that remains is very fine inclusions that are often do not have any shapes due to morphological convergence. As a result of billions of years of sediment evolution, pristine chemistry can be replaced with secondary processes, and thus, we may have putative microstructures of undefined origin. Therefore, sensitive instruments are required to investigate the smallest pieces of evidence, which can directly or directly point towards the biogenicity of the given structures. This goal drives the instrument development teams around the world and pushes the envelope of progress further every year, increasing the sensitivity and spatial resolution of space-type instruments. If such faint signatures of life, like those from the Gunflint microfossils, can be routinely identifiable with space instrumentation, this will mean that we have the necessary tools for in-situ exploration of the Red Planet and beyond.

During the course of the thesis, I will demonstrate that LIMS as an analytical method is among few space-qualified methods that can successfully identify the Gunflint microfossils, distinguish them from bio-organic surface contamination and inorganic chemistry of the host minerals. Moreover, using LIMS, chemically distinct individual microfossils can be identified, together with ppm-level trace element compositions. As a robust and high-throughput method, LIMS allows a continuous record of large-scale mass spectrometric data, making mass spectrometric imaging (MSI) possible on the surfaces of planetary bodies. Although suitably flat samples are not yet possible to manufacture on the Martian surface, the capability for MSI is still valuable. In addition to the Martian exploration program, an integrated LIMS/LDMS approach will be demonstrated, which will showcase potential for the detection and structural characterization of bio-organic molecules such as amino acids. This is an important measurement capability that can be of high utility during in-situ investigations of icy moons of Jupiter and Saturn.

1.2 LIMS – principles of operation and technical characteristics.

The Laser ionization Time-of-Flight Mass Spectrometry (LI-TOF-MS or LIMS) is a direct solid sample characterization technique that provides a highly localized chemical composition analysis with high sensitivity. Historically, the first experiments on laser mass spectrometry were published by (Honig & Woolston, 1963), where effective mass analysis was limited due to considerable space charge broadening and hence limited mass resolution. The early instruments also suffered from long pulse durations and power fluctuations; henceforth, the reproducibility of measurements was rather poor. The first ns laser-based microprobe was reported by (Fenner & Daly, 1966), where they utilized a nanosecond laser focused to ~20 μm. Further advances in laser-based mass spectrometry came from (Hillenkamp et al., 1975), where authors reported focused beam diameter of 0.5 μm (on the border of the diffraction limit), in-line microscopic imaging, and detection limits as low as 10^{-19} g (~0.2 ppm), measured from the thin sample of epoxy resin. At the same publication, the author mentioned that laser-based mass spectrometer (LAMMA-500) has an impressive range of measurement capabilities: "In principle, it should enable not only high sensitivity but also isotope analysis and the 'fingerprinting' of organic molecules, provided they do not decompose into fragments too small for recognition of the parent molecules". Not much later, mass spectrometric analysis of organic samples was reported by (Nitsche et al., 1978), where they showed measurements of the organic standards and heart muscle cells with detection limits corresponding to 0.1-10 ppm. Characteristic mass spectra of different organic substances obtained with the laser microprobe have illustrated the utility of applying laser-based mass spectrometry in the characterization of solid samples.

Although early instruments have shown impressive analytical capabilities, they were large laboratory-based instruments with limitations such as lack of computing power for quantification purposes, vacuum arrangement issues, isobaric interferences due to small resolving powers, long laser pulse durations, and lack of control over laser powers leading to large fluctuation in the ion yields. These limitations,
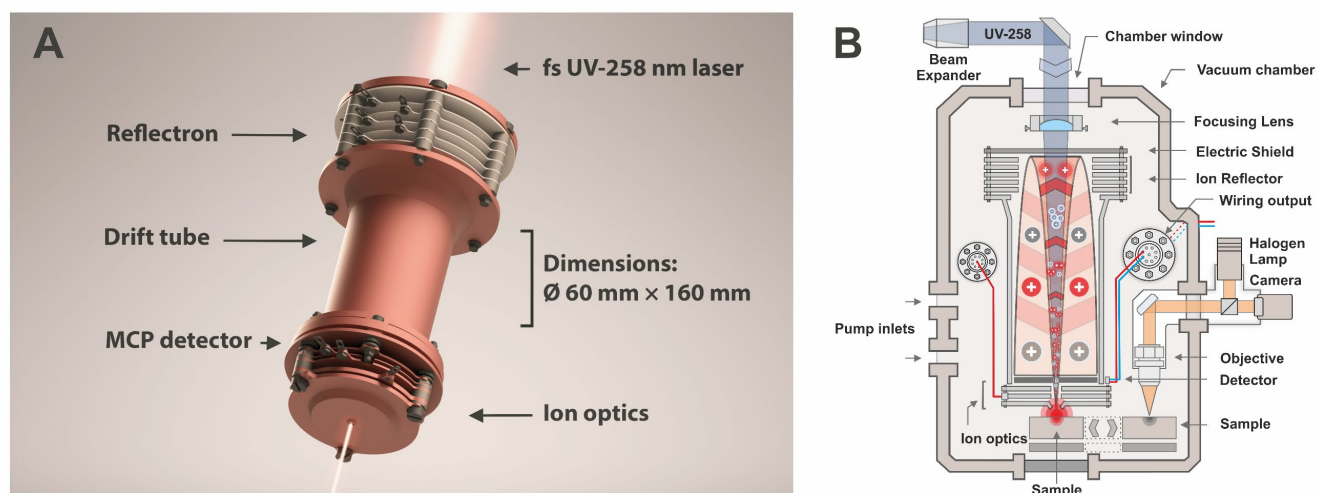
Figure 1A – 3D render of the LIMS instrument. The main parts of the instrument are denoted on the render. Figure 1B – Schematic illustration of the main parts of the integrated LIMS system.

therefore, made laser-based mass spectrometry a less popular method for characterization of solid samples, making the inductively coupled plasma mass spectrometry (ICP-MS) an analytical technique of choice of many laboratories around the world for a number of years. However, advances in laser technologies (Akhmanov et al., 1988), electronics (Schaller, 1997), and the development of ion simulation software (Dahl, 2000) made the LIMS technique reemerge in different scientific fields (Cui et al., 2012; Liang et al., 2017; Managadze et al., 2010; Wurz et al., 2012; Zavalin et al., 2015), facilitating the development of new laboratory-based and miniaturized instruments (Tulej et al., 2014; Wiesendanger et al., 2019).

A number of authors proposed new versions of LIMS microprobes as a viable solution for the needs of space exploration programs, utilizing various instrument designs (Arevalo Jr et al., 2018; Brinckerhoff et al., 2000; Goesmann et al., 2017; Managadze et al., 2010; Rohner et al., 2003; Wurz et al., 2012). The compactness of developed instruments allowed them to be deployable on planetary landers or rovers (Arevalo Jr et al., 2018; Li et al., 2017; Wurz et al., 2020; Wurz et al., 2009).

The Mass Spectrometry group from the University of Bern developed the LIMS instrument called LMS (Laser Mass Spectrometer), which was used in this thesis. Initially, the LIMS microprobe was developed as part of the Mercurian exploration program (Rohner et al., 2003). However, due to the changes in the mission design, the lander part of the mission was excluded from the final arrangement. Since 2003, the instrument has experienced continuous development and improvement, adding on measurement capabilities. In the current state, the LIMS instrument utilizes a UV-258 nm femtosecond laser ion source that can be exchanged to IR-775, or UV 387 nm wavelength with a possibility to use a picosecond delayed post ionization double-pulse scheme. It also encompasses an integrated microscopy system, sample introduction, and 3D translation system. The state-of-the-art vacuum system supports the measurements conducted with LIMS. The laboratory setting provides computer-controlled temperature and humidity, allowing for stable conditions in the laboratory. The high vacuum that reaches 10^{-8} mBar, is achieved by using the joint molecular turbo-pump and ion-getter pump. High vacuum conditions allow for direct transmission of all produced ions into the instrument.

The LIMS system illustrated in Fig. 1 utilizes the time-of-flight principle first introduced by (Wolff & Stephens, 1953). The pulsed laser (see the blue line in fig. 1B) passes through the instrument and ionizes material on the surface of the sample. The plume of ions radially expands and reaches the point where the extraction field of the instrument attracts only positively charged ions. The acceleration field gives an equal push to all ions, directing them into the field-free area (drift tube) where ions experience mass/charge separation. Further, ions enter a retarding field (ion reflectron), where they are reflected in the opposite direction and finally fall on the surface of the microchannel plate detector (MCP). The incoming time-separated flux of ions (see the parabolic ion paths in Fig.1b) launches an electron avalanche, which causes a measurable current on the output anodes. The current is further digitized using the analog-to-digital converter. Thus, the element and isotope composition of the sample will be directly proportional to the amount of current registered on the output anodes. The recorded time-of-flight spectrum will indicate the presence of elements with specific masses, allowing to decipher the chemical composition of the investigated materials. Although the diameter of the analytical spot is wavelength dependent, using the current UV-258 nm fs laser, it is possible to characterize matter with spatial resolution around 5 μm.

1.2.1 Fundamentals of Time-of-Flight Mass Spectrometry

Suppose that singly charged ions are subjected to the pulsed extraction field U, which accelerates and transports ions to the entrance of the field-free region, which is called a "drift tube". During the ion extraction stage, all ions receive an equal amount of acceleration:

$$U_{acc} = qU, \qquad (1)$$

Where q denotes the charge of the ion, and U acceleration voltage. The acceleration energy of each ion is converted into kinetic energy:

$$E_{kin} = \frac{m_i v_i^2}{2}, \qquad (2)$$

Where m_i and v_i Corresponds to the mass and velocity of ion i. From the equivalence of equations 1 and 2, it is possible to derive velocity of the ion i:

$$v_i = \sqrt{\frac{2qU}{m_i}}, \qquad (3)$$

Therefore, considering the dependence of velocity on the mass of the ion, we can observe that lighter ions reach the detector in a shorter time compared to heavier ions. The time that ions need to pass the field-free area of the mass analyzer and reach the detector can be calculated as:

$$t_i = \frac{d}{v_i} = d * \sqrt{\frac{m_i}{2qU}} \Rightarrow t_i \propto \sqrt{\frac{m_i}{q}}, \qquad (4)$$

The total time that ion needs to reach the detector, including the time t_0, can be given as:

$$t_{tot} = t_0 + t_i, \qquad (5)$$

Therefore, from equation 4, it is clear that ion arrival time is proportional to the square root of its mass over charge ratio. However, it is assumed that ions have similar initial energy distributions, initial speeds, and similar directions. In reality, these parameters can significantly vary, limiting the mass resolution of the instrument.

The resolving power of the time-of-flight instrument is directly proportional to the length of the ion flight and depends on the initial ion parameters. Therefore, implementation of the retarding field, i.e., reflectron, can significantly improve the resolving power of the given TOF mass analyzer. The ions with a significant spread of the initial kinetic energies will penetrate the retarding field with varying depths until the retarding field will compensate their kinetic energies and reflect them back. Consequently, the introduction of an ion mirror improves the mass resolving power by a factor of 2 without extending the length of the instrument. Also, the retarding field can compensate for the spread of initial kinetic energies (Boesl, 2017).

1.2.2 Technical Characteristics

Although the LIMS system was developed in 2003 (Rohner et al., 2003), various instrument parts were significantly modified, improving the measurement capabilities. Here I will outline some of the significant changes that instrument setup experienced. For example, implementation of the high-speed detector system with a high dynamic range (Riedo et al., 2017) allowed improving the composite signal's dynamic range to the level of 8 orders of magnitude. The application of a high-voltage pulser system improved the detection sensitivity of the instrument (Wiesendanger et al., 2017), allowing to identify ppm-level chemistries from lunar meteorite sample. Improvement of the ion source was reported by (Riedo *et al.*, 2013b) from ns to a state-of-the-art femtosecond laser system with ‰ level of energy stability. Implementation of the double-pulse ionization scheme (Riedo et al., 2021; Tulej et al., 2018) revealed that ionization characteristics of plasmas could be significantly improved. The introduction of a fully automated data analysis platform was reported by (Meyer et al., 2017), improving the speed of data analysis. A contribution by (Wiesendanger et al., 2018) reported the implementation of a microscopy system that allowed precise targeting of micrometer-sized features. Additionally, the knowledge and expertise collected during years of work on small LIMS prototype lead to the construction of the new LIMS instrument with high mass resolution (Wiesendanger et al., 2019). Summarizing all these developments, table 1 outlines the current LIMS properties and figures of merit[1].

LIMS properties	
Dimensions of mass analyzer	160mm * ø60 mm
Laser wavelenght λ	775, 387, 258 nm
Laser pulse length τ	~190 fs
Laser pulse repetition rate	1 kHz, controllable
Time of flight	20 µs, controllable
Mass range	1-1000, controllable
Voltage on detector system	~2000 V
Voltage on ion optics	<2000 V
Active channels for data acquisition	4, controllable
ADC sampling rate	3.2 Gs/s
ADC vertical resolution	12 bit
ADC storage speed	Up to 1000 spectra/s
Figures of merit of LIMS	
Lateral resolution	5-15 µm
Vertical resolution	nm-range
Dynamic range	4 orders of magnitude
Mass resolution	~800 m/Δm
Detection sensitivity	~ppm range
Isotope ratio accuracies	‰ to % range

Table 1. LIMS properties and main figures of merit.

---

[1] The dynamic range can be imporved to the range of 10^8 by recording the composite spectrum from all 4 channels of the detector (see Riedo *et.al.,* 2017)

1.3 Thesis overview

The central goal of the thesis is to further advance measurement capabilities of the LIMS space-prototype in characterization of bio-organic and inorganic chemistries. This goal is addressed by a combination of hardware modifications and the application of computational approaches. The hardware modifications include improvement of the ion source - implementation of the far UV fs laser that allowed to increase ionization efficiencies of investigated materials. Implementing a double-pulse ionization setup allowed us to gain additional improvements on the stoichiometry of investigated mass spectrometric signals. And lastly, further work on the high voltage pulser regime allowed to increase the instrument's sensitivity, which resulted in a larger range of identified species. Once the hardware-based signal quality improvement was achieved, I focused on computational methods that can further increase the signal quality and allow to infer structural inhomogeneity present in the data, in an unsupervised fashion. Such structural diversity identified from large LIMS observations was used as a proxy to understand the mineralogical and chemical compositions of investigated samples.

The thesis is organized in chapters: Introduction, Methods, Results, and Conclusions. Additional information for specific sections from scientific results is provided in the appendix. The methods section presents an overview of computational methods used in the thesis, describes main equations and provides a short description of algorithmic procedures.

The results section of the thesis presents an extensive study of Precambrian Gunflint microfossils using the miniature laser ionization mass spectrometer (LIMS) developed for space applications and high-resolution laser ionization mass spectrometer developed for laboratory use. The Gunflint microfossils were considered as Martian analog material. During the course of the thesis, I will demonstrate that by consequently improving the mass spectrometric signal quality and recording large spectrometric data, it is possible to identify individual mass spectra that are indicative of bio-organic inclusions (microfossils). Using the spectral data reduction techniques, I will demonstrate structural characteristics and data patterns that are measured from individual microfossils. Further, I will classify an extensive collection of bio-organic spectra from inorganic host material in a probabilistic manner, using various binary machine learning models. Furthermore, using the graph-theoretic approaches, I will demonstrate that it is possible to classify bio-organic and inorganic chemistry from the Gunflint chert in a deterministic way.

Similarly, I'll present the results of the mineralogical investigations of microscopic inclusions from the amygdaloidal pillow basalt, where complex mineralization patterns were deciphered using the LIMS instrument.

In addition to the Gunflint studies, I will present the results of molecular desorption experiments conducted on a set of amino acids that will showcase the versatility and high value of integrated LIMS/LDMS technology in the structural characterization of complex bio-organic molecules. Also, I'll address specific aspects of ion generation in time-of-flight mass spectrometry, where a double-pulse ionization scheme will showcase that plasma stoichiometry can be significantly improved, allowing for more stoichiometric signals to be recorded.

Computational methods

Precis

Here I will provide technical notes on computational methods used in the results section of the thesis. First, I'll introduce the spectral preprocessing routines to emphasize the importance of signal quality in characterization of microscopic inclusions. Further, I'll present matrix factorization methods as a foundation of multivariate analysis. In the third section, I present graph-based data processing methods that will demonstrate that it is possible to extract mineralogical composition from LIMS observations using an unsupervised data-driven approach. Lastly, I present the machine learning frameworks that can be used to classify mass spectra from previously unseen samples. Although the essential information is provided, some algorithmic details are omitted from the thesis. Thus, for the full description of algorithms, I refer to original publications.

2.1 Spectral preprocessing routines and decomposition methods

Recovering the intensities of the given mass lines from the raw time-of-flight mass spectra represents the first step of uncovering the chemical dependencies present in the acquired dataset. The mass spectra should be accurately preprocessed, removing the noise and artefactual signal change. However, this goal might seem simple; in practice, it requires a substantial effort in signal modeling and fine-tuning of preprocessing parameters to retrieve high-quality signals from extensive mass spectrometric data. Moreover, mass spectra from different campaigns and with different histogramming numbers can require separate tunning procedures. In this chapter, I provide an overview of spectral preprocessing techniques utilized in this thesis. I also outline the importance of accurate signal retrieval as a foundation for multivariate analysis and pattern recognition. This chapter is based on functions and methods proposed in (Gil & Marco, 2007) and (Meyer *et al.*, 2017) and describes a complete set of preprocessing routines required to accurately analyze mass spectrometric data.

2.1.1 Denoising

The experimental Mass Spectrometry begins with data acquisition, which can often be noisy and pertain to some degree artifacts that are not directly related to the sample's chemistry. Spectral preprocessing aims to achieve a high signal-to-noise ratio (SNR), an important parameter measured from the mass spectra and identified for any given mass peak. Higher the SNR, better the sensitivity, and richer the information that can be retrieved from the given measurement. Therefore, the first subset of preprocessing procedures is aimed to increase this parameter. However, the problems in the analytical signal can be found in the experiment design and sample preprocessing. Therefore, the acquisition of high-quality mass spectra should be considered as a complex problem, addressing the ionization, sample preprocessing, and experiment design in the first place, and computational methods improving the signal quality on the second. If experiment settings were set to the level that produced the optimal signal quality, i.e., the instrument provides the required signal without excessive noise and distorted peak shapes, one could further continue with spectral preprocessing techniques. The follow-up step aims to remove the remaining artefactual signal change and minimize the effects distorting the analytical signal.

The mass spectra usually contain some level of noise intermixed with a signal of varying intensity. Therefore, an accurate assessment of the noise bandwidth should be done before the noise filtering. There are many ways to filter out the noise, and signal processing, in general, represents the separate research field itself. Here I will present the analytical pipeline that provides the conservative noise removal and preserves most of the analytical signal. Need to mention that interventions such as noise removal always mean some loss of signal.

The first step of the noise removal starts with subtracting the empty waveform from the mass spectra. This step removes the low-frequency noise present in the acquisition line (from the detector all the way to the data acquisition), which in any laboratory setting represents an antenna that captures the artefactual signals emitted from other electronic devices (electromagnetic interference, i.e., EMI). The empty waveform should be recorded with the same settings and with the same histogramming number (single-shot averages), capturing only the artefactual noise present on the acquisition line. Further, the noise subtracted mass spectra are preprocessed

using the low-pass filter (LPF). The signal's frequency content reaches 200 MHz, but sampling is performed at 3.2 Gs/s, with an analog bandwidth of ~1.5 GHz, which allows for the removal of high-frequency noise. Therefore, The LPF applied to the mass spectra removes the high-frequency component from the observation.
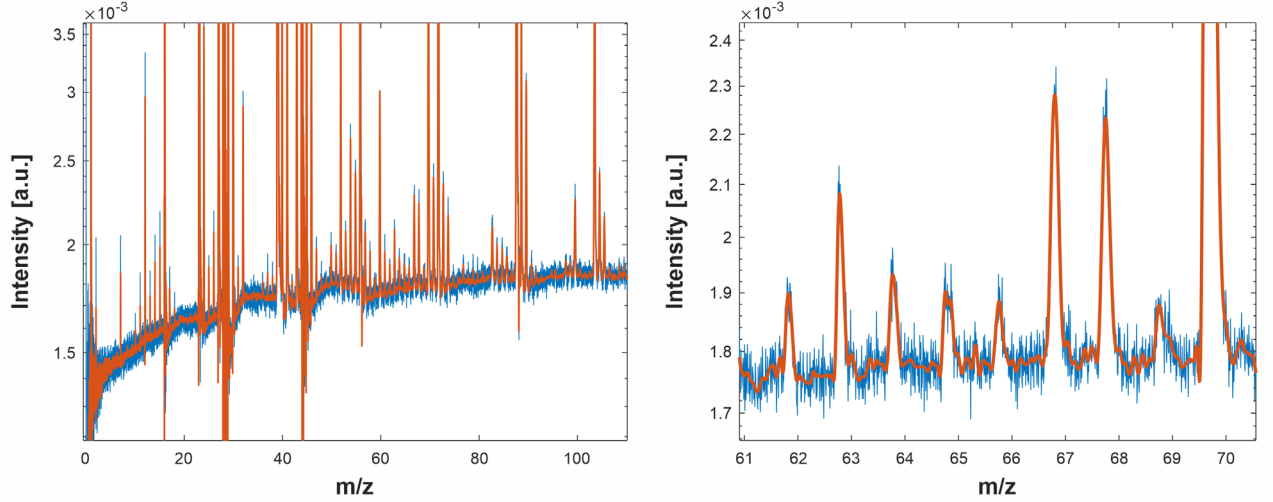
Fig. 2 The exemplary mass spectrum preprocessed with LPF. The noise subtracted mass spectrum is shown with the blue line. The LPF preprocessed mass spectrum is shown with an orange line. The zoomed-in version of the spectra is shown in the right panel. Note that the intensity scale is logarithmic.

The mass spectrum shown in Fig. 2 represents the effect of applying the LPF to the observations recorded with the LIMS instrument. As it could be seen from the figure, the LPF does not change too much the peak shapes and improves the SNR by a factor of 2-5. The overall improvement of the SNR depends on the particular character of the high-frequency component of the observation. The frequency cutoff value also can be used to attenuate the output result of the low-pass filtering. Typically the normalized passband frequencies at 0.1 – 0.15 $[\frac{\pi rad}{data\ point}]$ are enough to remove the high-frequencies from the signal. The stopband attenuation can be chosen around 60 dB, and the steepness of LPF can be chosen at around 0.85 dB/oct, providing a relatively smooth cutoff of high frequencies. Thus, the signals buried in the high-frequency noise can be retrieved, and the overall quality of mass spectra can be improved using the LPF. However, the remaining noise components can still be significant, i.e., sharp spikes in the baseline (see Fig. 2, right panel), which are parts of the noise bandwidth that survives low-pass filtering. Although such signals are typically very small, they can introduce an additional uncertainty while working with trim signals.

Overall, filters such as LPF are generally very helpful and better in retrieving more resolved data in comparison to conventional smoothing approaches (such as moving average). While within the single mass spectra, an improvement of the SNR can be not that significant, the more extensive data preprocessed with LPF can reveal a substantial improvement in the determination of the noise floor. Moreover, a notch filters (cutoff of specific frequencies) can be of high value as well. Therefore, an accurate assessment of the frequency range is required to understand which part of the frequency spectrum can be preserved and which part represents the noise bandwidth. Similar to the LPF the steepness of the filter can play a big role in the overall signal preservation. Therefore, a signal modeling with synthetic spectra (and a real noise component) can be of high utility in the analysis of mass spectra.

2.1.2 Parametric polynomial smoothing

The preprocessed with noise subtraction and LPF spectra can still contain low-frequency levels of noise that can further be flattened using parametric peak preserving smoothing. Polynomial filtering (smoothing) is an alternative to conventional noise reduction techniques and uses subsamples of data and tries to fit them into a predefined model. Thus, the sharp high-frequency peaks from mass spectra can be preserved, whereas the small noise contributions can be removed from the output spectrum. The Lowess filtering smooths the spectrum using the locally weighted linear regression in a given filter window. The method uses the user-defined evenly spaced windows of the mass spectra and fits the linear (Lowess) or quadratic function (Loess). Thus, by collecting the statistics on neighboring windows, it is possible to identify the data points that are outliers and data points that lie in the range of the actual signal.

The generalization of the Lowess filtering represents the Svazitsky Golay filtering (SGF) (Press & Teukolsky, 1990). The SG filter smoothes the data using the least-squares polynomial filter. The method derives the filter coefficients by performing the linear least-squares fit using a polynomial function of a given degree. The fit is iteratively repeated in small filter windows; thus, the size of the filter function provides the critical constraint to the overall smoothing result. For typical LIMS observations with 64'000 digitized data points per spectrum, a filter function with 50 data points provides a relatively good fit to the signal. However, with an increased number of data points in the spectrum (i.e., 320'000 data points per spectrum in HR-LIMS-GT measurements), an appropriate adjustment should be made. Another parameter that affects the smoothing result is the degree of a polynomial, where higher degrees can adapt to more complex shapes, and zero degree is an equivalent of the weighted mean estimator on a given window.

Figure 3 shows the previously denoised mass spectrum (see Fig. 2) further smoothed with parametric SGF. The mass spectrum was regressed using the 50 samples wide sliding window and degree 4 of the polynomial. The output result is a smoothed mass spectrum that does not contain the jittery signal between masses. However, the output SNR of peaks is slightly reduced. Therefore, the SGF smoothing operator requires balancing between preserving shapes of the peaks and the smoothness of the baseline. Therefore, the degree of smoothing is dependent on the scientific task and can be adjusted accordingly.

Another drawback of the SGF is that polynomial fit forms artefactual negative spikes in the vicinity of intense peaks. Such spikes are generally not harmless in further downstream peak data analysis; however, it is important to keep in mind that the negative spikes can be significant for some peaks. It is also important to note that the baseline regression on smoothed spectra can be complicated with artefactual signals like SGF-induced negative spikes; therefore, it is recommended to perform the baseline extraction prior to the smoothing.

The overall strategy for smoothing and noise reduction is dependent on the scientific tasks. For example, the calculation of isotope fractionation factors between given isotope pairs can be affected by noise contribution or overly smoothed data. For that reason, actual calculations on given isotope intensities are usually calculated on unprocessed raw mass spectra. In contrast to the isotopic studies, the LIMS spatial imaging of various elements is usually done on preprocessed spectra. Baseline corrected and cleaned from noise mass spectra better reflect the actual structure of elemental

intensities and provide overall improved mass spectrometric images. Concluding this section, it is important to mention that all spectral preprocessing procedures should be applied with care because some of the important features of mass spectra can be oversmoothed, and some new artifactual features introduced.
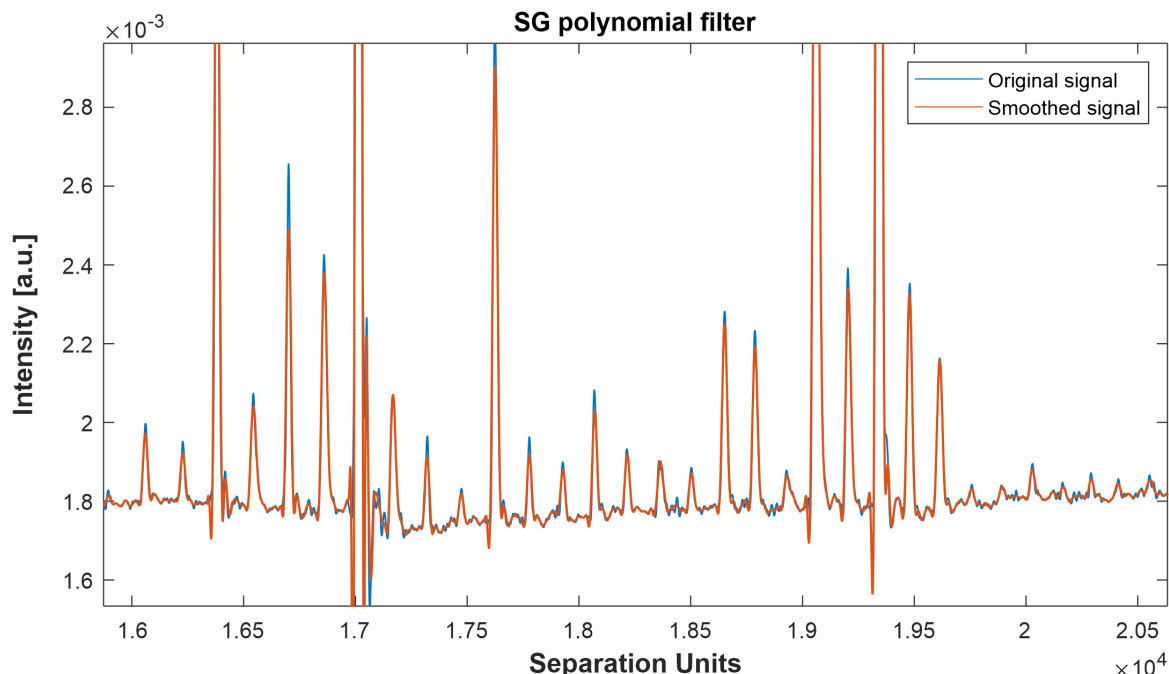
Figure 3. The exemplary mass spectrum smoothed using the Svazitsky-Golay polynomial filter. Note the jittery pulses between masses that are smoothed using the SGF.

The change of the signal shape after application of any smoothing operator is inevitable. Therefore, it is up to the user to define to which degree the change is acceptable. The extent of the distortion can be controlled by checking the SNR of the output signal. Since mass spectrometry deals with pulsed signal shapes, the moving average is a bad choice for smoothing due to the overly aggressive smoothing of peaks. Typically, combination of the SGF and LPF can retrieve relatively small signals and improve the output SNR up to two orders of magnitude, if applied properly. Also need to mention that the SGF can be applied to the data only with evenly spaced intervals, which is not the case with mass scale calibrated spectra. Therefore, the SGF filter needs to be applied in the preprocessing stage, while the mass scale assignment is not yet performed.

Other smoothing approaches exist i.e., the Gaussian filter which uses the Gaussian kernel to smooth the data (widely used in the image processing). In contrast to the SGF, the Gaussian kernel does not preserve the widths of peaks, thus, lowering the output mass resolution. A relatively new advancement in the signal processing is application of wavelets and neural networks in noise suppression. The deep neural networks can be used to learn the noise statistics from data, therefore allowing to tailor models to the lab-specific noise contributions. Such approaches can significantly improve the SNR with minimal changes in the shapes of the output waveforms. Although if data specific biases are introduced it might be more challenging to remove those from the analyzed spectra.

2.1.3 Baseline correction

In general, the varying baseline of any given mass spectrum is caused by the drift of DC voltage on the anodes and can be detected on every observation with slightly differing profiles. It can also reflect the ion overload or chemical noise in the matrix. These effects lead to the positioning of mass peaks in ascending or descending regions of the baseline, affecting the investigated peak areas. An example of a varying baseline registered from the composite mass spectrum (200 single laser shot spectra) is illustrated in Fig. 4, where the blue line indicates the original mass spectrum and the red line denotes the regressed baseline. The baseline subtracted mass spectra bring all baselines to the same level, thus, making them comparable.
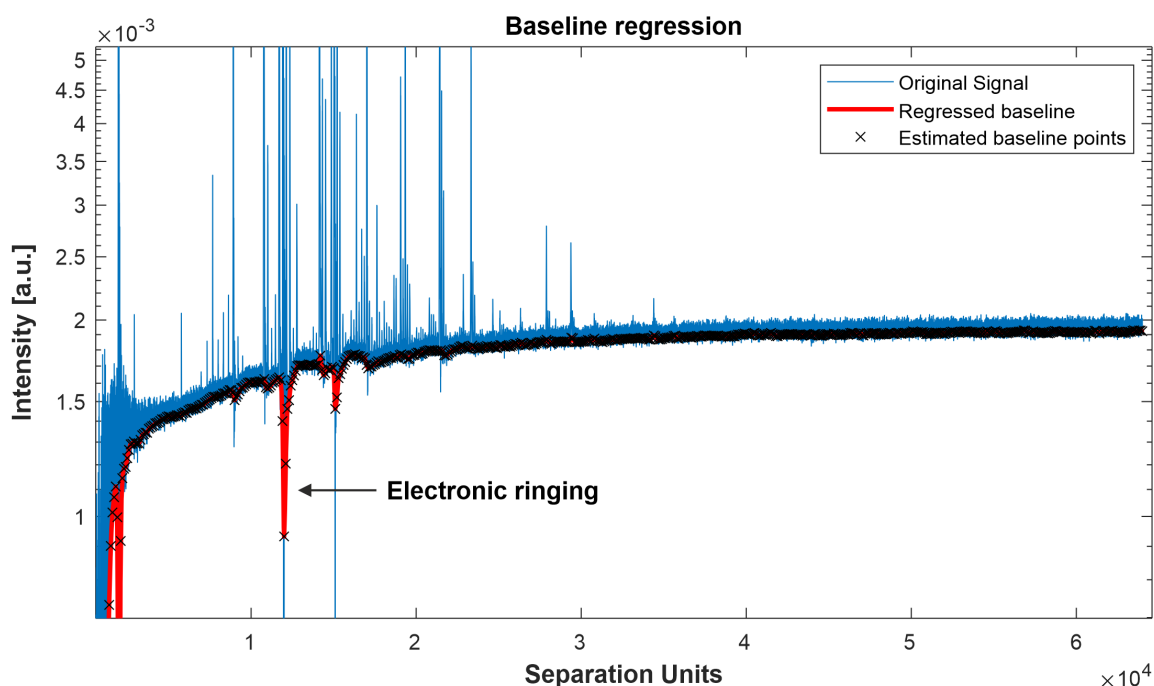
Figure 4. The varying baseline was estimated using the slider window with a step size of 100 samples.

The strategy for removal of low-frequency baseline follows the three-step algorithm: 1) Estimate the baseline within the filter function window. 2) Regress the baseline using spline interpolation. 3) Subtract the overall baseline from the mass spectra. As in many other smoothing algorithms, the baseline subtraction quality depends on the size of the filter function. If the window size is chosen too big, the approximate baseline will jump on the mass peaks, leading to the artefactual peak intensities. If too high, the baseline will follow the shape of the noisy baseline leading again to artefactual intensities. Therefore, it is important to choose the baseline approximation function parameters to satisfy the baseline diversity of the entire dataset. Mass spectrometry imaging of reasonable sizes usually contains tens of thousands of composite mass spectra, making the unsupervised baseline subtraction a vital part of spectral processing. However, using chain-linked baseline subtracting functions with differing filter windows allows achieving a reasonable quality on large datasets.

Figure 5 illustrates a composite mass spectrum (200 single laser shots) smoothed using the SG filter and removed baseline. The most prominent Si peak showcase the SNR around 10^4, and the noise band reduced to approximately half a decade. It is important to note that the mass spectra shown in Fig. 5 represent the single position

measurement (a single-pixel from the MSI). Therefore, additional signal improvement can be achieved by further accumulating waveforms. In general, with accurate noise modeling, it is possible to reach the SNR around 10^5 for major (most abundant) peaks, with simultaneous improvement on the range of noise band. However, noise can reach the low-frequency floor, which is hard to distinguish from actual mass spectrometric signals and, therefore, it is challenging to improve the signal further. In principle, introducing hardware integrated active noise-canceling lines in the acquisition system should further increase the achievable SNR's by few orders of magnitude. Although such developments are out of the scope of this thesis, the acquisition line can be further improved, at least theoretically.
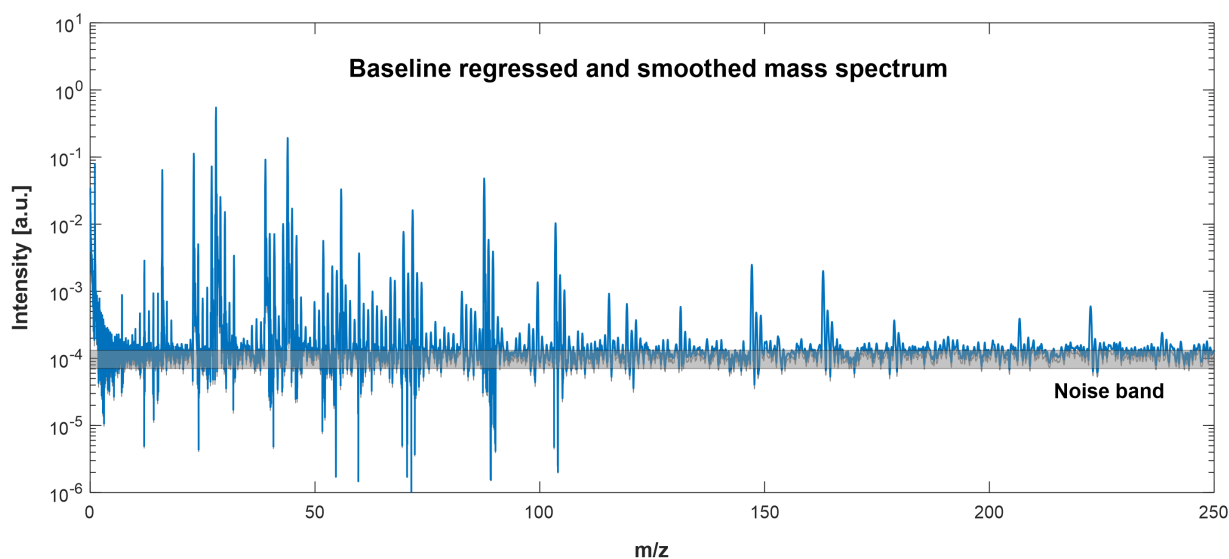
Figure 5. Baseline regressed and smoothed mass spectrum (an average of 200 single laser shot mass spectra).

2.1.4 Spectral alignment and normalization techniques

In general, spectral misalignment does not represent a big issue in the fs-LIMS characterization of solids. However, some minerals due to varying photon coupling and topography can be ionized with differences between ion yields, and the ion arrival times can slightly differ. The mismatch in the ion arrival times can lead to systemic misalignment of mass peaks, manifesting in the broadened and tailed peaks for histogrammed and later averaged in postprocessing mass spectra. For example, collecting significant numbers of mass spectra (>40k spectra for MSI) from heterogeneous samples can require correction procedures to be applied to the data. The change in the ion packet's arrival time can affect the averaged peak shapes, for example, transforming them from purely Gaussians to the tailed and broadened Gaussians. The change in the peak shapes leads to a decrease in the mass resolution and overall intensities of the investigated peaks. Therefore, misalignment of the ion arrival times is an issue that worsens the quality of measurements. The spectral alignment procedures can account for the shifts in the time-of-flight spectra and bring the observations to the expected arrival times for selected masses, thus improving the output signal quality.

The spectral alignment algorithm (Gil & Marco, 2007) constructs a synthetic spectrum with a user-defined number of peaks, represented by Gaussian kernels. The time-of-flight values of the synthetic peaks and actual observations are shifted and

scaled until the cross-correlation function cannot be further improved. When multiple spectra are aligned, the algorithm repetitively optimizes the cross-correlation function for a given observation. The strength of the warping function described in (Gil & Marco, 2007) is that it can learn the warp function from the raw data, whereas other solutions do not provide such an option. Another advantage of the algorithm is in the fact that it allows multi-peak alignment. The user can define the master vector with time-of-flight positions of peak centers that are further used in the alignment procedure. Thus, if some masses are missing, the alignment can be done on the remaining peaks. The algorithm works best with 5-8 peaks representing the most prominent masses present in the mass spectrum. Practically speaking, the alignment is almost always done on major isotopes of carbon, oxygen, sodium, and potassium.

Figure 6 illustrates result of the spectral alignment performed on the randomly subsampled data (65 composite mass spectra) from the imaging data set (30'000 mass spectra) recorded from the Gunflint chert. As it can be seen from the figure, the carbon peak alignment is performed on the nominal level. All peak centers are aligned in the tolerance window of a few ns. Every count in Fig. 6 corresponds to the 0.3125 ns of the flight time. The electronic ringing (oscillation of the signal in the acquisition line) observed on the right side of the peak is not confused with the actual peak position. Another important observation is that the noise band showcases varying intensity regions that form a narrow confidence interval. The shape of the pre and post-peak is determined by the SG filter and represents an artifact of the spectral smoothing using the polynomial filter.

Multi-peak aligned mass spectra
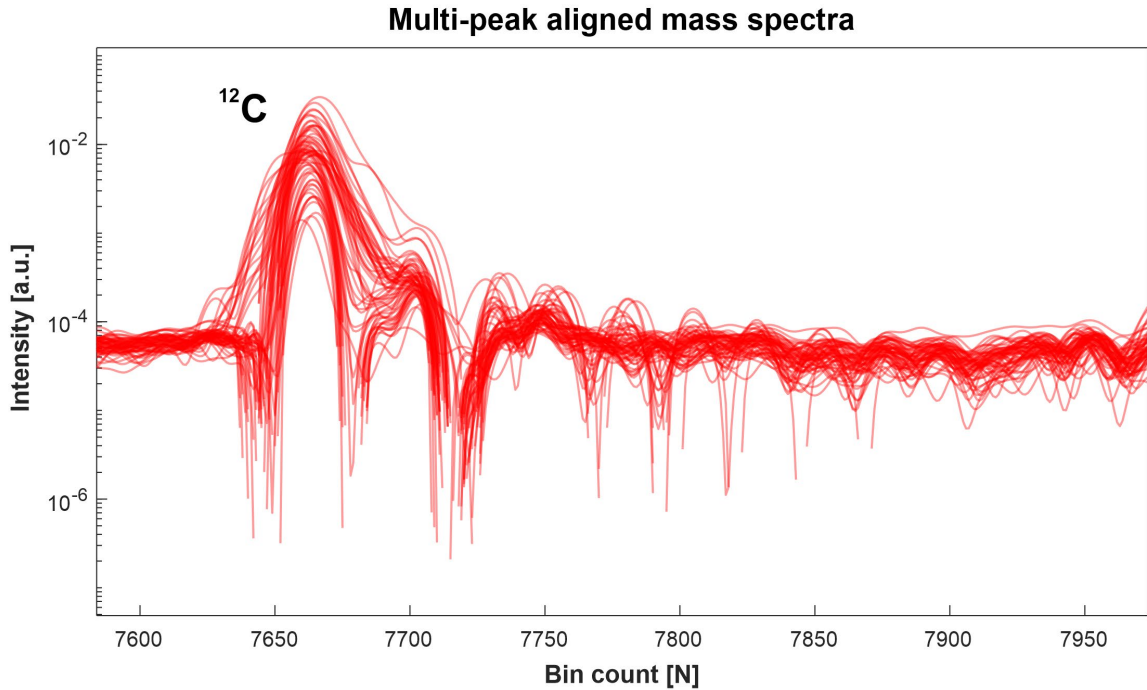
Figure 6. Multi-peak spectral alignment (5 peaks) of spectra registered from the Gunflint chert.

The broadened mass spectra can be seen on the left side of the central carbon peak. Typically, the peaks with the broad profile are shifted a few ns to fit the position of resolved carbon peaks. The results of the spectral alignment can improve the mass resolution of investigated peaks by 20-30 percent on the output histogrammed mass spectrum. This is of high importance for isotope ratio studies, where the resolution and the SNR of investigated isotopes are determinative of calculated ratios.

Repeated experiments or experiments with varying histogramming of laser shots will differ in the scale of ion yields, making the spectra incomparable. Various normalization techniques have to be utilized to bring the mass spectra to the same scale. One possible way is normalization to the noise level, which will bring the baseline intensity values of all mass spectra to the level of the noise band of the master spectrum. Another normalization technique provides normalization to the maximal peak intensity within a given intensity range. The intensity range from 0-1 or 0-100 is among the standard choices. Normalization to the maximum peak intensity is useful, for example, in LIMS imaging, where a variable drift of ion yields are observed, with systematic (depth-dependent) and random (local absorption variations). By bringing the most intense peaks to the same level, it is possible to improve the resulting contrast. Spectral normalization is the last step of spectral preprocessing and represents the spectrum's final form.

2.1.5 Single mass unit spectral decomposition.

After spectral preprocessing procedures are finished, the mass spectra can be further decomposed into vectors with single mass unit intensity values. Every mass spectrum contains 64000 digitized data points (which represents approximately 20 μs of the flight time and covers a mass range from 1 to 800 [a.m.u.]) and around 100 points for a single mass peak. The single mass unit peak in its complete shape encodes information on how many electrons were detected; therefore, we can integrate over the time-of-flight window of the mass peak to retrieve the mass peak area. The peak area integration is preferred relative to the measuring peak height (prominence) because the peak area encodes more information.
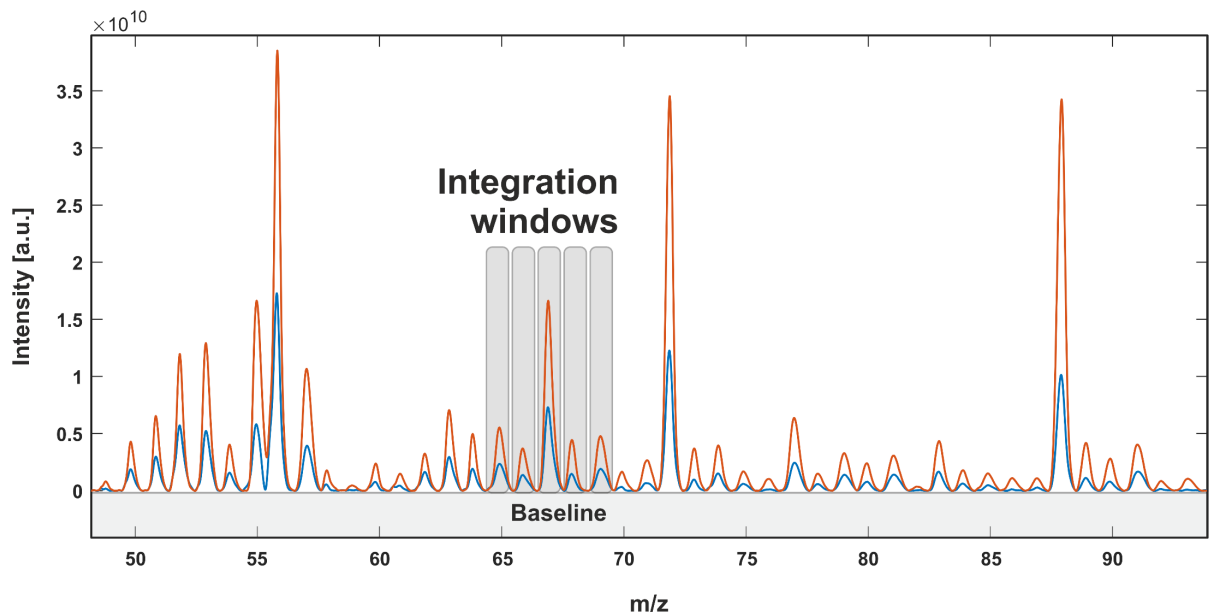
Figure 7. Exemplary composite mass spectra (200'000 single laser shot spectra each mass spectrum) after a full stack of spectral preprocessing procedures were applied. The preprocessing steps include denoising, smoothing, baseline correction, and spectral alignment. Grey bars illustrate the single unit masses (widths of the integration windows) chosen for peak integration.

The time-of-flight spectrum has to be linked with the corresponding mass scale to extract the peak intensities of the masses. The separation of the different masses occurs in the field free part of the mass analyzer when all extracted ions were equalized

to the same kinetic energy by initial acceleration. Consequently, the mass scale to the time-of-flight spectrum can be linked using the equation:

$$\frac{m}{z} = k_0 * (t - t_0)^2, \qquad (6)$$

Where the k_0 and t_0 are instrument-dependent constants. It is possible to assign a mass scale with accuracies better than 500 ppm using a set of isotopes that will cover a significant part of the mass range. If the isotopes used for mass calibration are all localized in the range of light masses, accuracies for heavier masses can experience loss of precision. Therefore, a wide stretch of isotope masses is required for accurate mass calibration.

For example, Fig. 7 illustrates the time-of-flight mass spectra with the assigned mass scale. The spectra are preprocessed using the full chain of preprocessing routines and reveal the uniform baseline with a clean signal that reflects the chemical change in the analyte materials. The two spectra compared with each other represent the cumulative mass spectra registered from two types of microfossils, measured from the Gunflint chert. Every spectrum represents the composite of 200'000 single laser shot mass spectra. As it could be seen from the figure, the quality of the spectral preprocessing is very high since no artifacts could be noted from the spectra.

Quantitative analysis of mass spectra deals with the abundances of given mass units. Therefore, we can calculate the number of electrons registered within the mass unit of interest using the equation:

$$A_e^{[a,b]} = \int_a^b N_e(TOF)\, dTOF - \int_i^j F_{bg}(TOF)\, dTOF, \qquad (7)$$

Where $A_e^{[a,b]}$ is a number of electrons of a particular isotope, $[a,b]$ is the time-of-flight interval where the mass peak is distributed, and $N_e(TOF)$ is the number of electrons per time increment. F_{bg} denotes the number of electrons registered from the background in the time-of-flight range $[i,j]$, which corresponds to the same width as $[a,b]$.

The peak list for integration can be derived from the peak detection algorithms. However, for multivariate analysis of the entire datasets, one can derive the peak area values for all single unit masses. Figure 7 denotes the widths of mass peaks that are chosen for the integration (shown with grey bars). In complete spectral decomposition, the widths of all single unit masses are considered. Practically speaking, convolution of the full mass spectrum includes an integration of the masses in the range from 1 to 300 a.m.u., which covers all stable isotopes and small molecular compounds. Therefore, the output reduced dataset contains a vector for every mass spectrum with intensity values of single mass units. This step dramatically reduces the data size to handle and allows the use of reduced data matrix directly in a follow-up multivariate analysis.

The data reduction using peak integration can be realized using different integration strategies. The peak integration using the Simpson 3/8 rule provides the best speed to quality ratio (Riedo *et al.*, 2013a) and retrieves the integration accuracy estimate. However, other integration approaches can be used, such as Monte Carlo (MC) integration, which will yield an accurate peak area assessment, although MC requires significantly more time. As it was mentioned before, peak intensities can be retrieved as the maximal value of peak prominence. It is one of the fastest approaches; however,

it will result in the loss of information about isotope abundances because maximum peak intensity represents only a proxy to the full peak area and because peak widths are changing with mass. Another way to assess peak area can be calculated by fitting the Gaussian kernel to the investigated peak and retrieving the area of the fitted Gaussian, though such an approach is neither fast nor accurate. Overall, Simpson's integration was chosen as the best alternative to all described integration strategies as a compromise solution between the speed of data acquisition and quality of the peak area extraction.

Another viable strategy that can be applied is spectral down-sampling that can reduce the size of the original mass spectrum down to the requested number of sample points in the spectrum. Although spectral down-sampling does not represent a type of spectral decomposition, such spectral transformations can be useful as a fast data preprocessing step, making the spectral data bank readily available for the downstream analysis and low-dimensional convolution. A high degree of the spectral down-sampling tends to lose large portions of the signal, thus making it an artefactual spectrum with modified intensity values (spectra can contain a high degree of distortions of the original peak intensities). Therefore, the down-sampling if applied, should be applied with care.

Another point that needs attention is the broadened mass peaks (peaks with ion overload or peaks of metastable molecules) can have a negative impact on the integrated peak intensity values. The broadened peaks can intrude into neighboring mass peak areas, thus, modifying their intensity values. Therefore, a high mass resolution of the analyzed mass spectra is of high importance in the spectral decomposition process. The mass spectra prior to the application of preprocessing routines can be preselected using spectral scoring approaches. For example, it is possible to take into analysis only spectra that are nicely resolved, by fitting the Gaussian kernel to the peaks of interest and measuring the resolution of the fitted Gaussian kernel (see (Wiesendanger et al. 2019), for more details on the spectral scoring approach). However, spectral quality and characteristics of the mass peaks can be addressed through modification of the ion source. The implementation of the UV-258 nm laser in contrast to the IR-775 nm laser will showcase that it was possible to improve the mass resolution almost by order of magnitude by using the more absorptive color of the fs laser radiation (see "Multiwavelength ablation/ionization and mass spectrometric analysis of 1.88 Ga Gunflint chert").

Overall, concluding this section, I can remind us that the mass spectrum prior to the decomposition needs to be accurately preprocessed using the LPF, SGF, denoising, normalization, baseline correction and spectral alignment, if necessary. Thus, mass peaks in the analyzed spectrum should be artifacts-free and represent only the chemical change in the analyte material. If dealt with large mass spectrometric observations, the randomly sampled mass spectra needs to be analyzed for the quality of the spectral preprocessing routines. Further, using the solution of the quadratic equation used for the mass scale assignment, one can recalculate the mass-time intervals for the peak integration procedures. Therefore, the required range of masses can be integrated, providing the discretized intensity values for investigated mass peaks. Although, the baseline (background) shapes can vary significantly for unprocessed spectra, the preprocessed data should reveal a significantly improved mass spectra.

2.2 Matrix factorization methods

2.2.1 Principal components analysis

The Laser Ionization Mass Spectrometry as a high throughput solid samples characterization technique provides very sensitive and spatially resolved data that are often difficult to interpret due to its size and complexity. A principal component analysis (PCA), also called Karhunen-Loève transform, Singular Value Decomposition (SVD), or eigenvector projection, is a data dimensionality reduction technique developed in 1901 by Karl Pearson (Pearson 1901). The PCA reduces the initial data and minimizes the loss of information and increases the overall interpretability. In its shortest description, the PCA does that by creating the uncorrelated new variables that successively maximize observed variance in the dataset. The PCA has been reinvented many times using different approaches; for example, the PCA solutions could be derived as eigenvectors of the covariance matrix or equivalently as the singular values of the centered data matrix. Furthermore, most recently, a contribution by (Gemp *et al.* 2020) provides a novel game-theoretic view on principal component analysis as a Nash equilibrium in a competitive game in which each eigenvector is approximated by a player, whose goal is to maximize their own utility function.

Principal components analysis provides a foundation for multivariate analysis. PCA, essentially, as a matrix factorization method, gives an approximation of a data matrix **X** as a set of new variables that capture essential patterns of original data in a linear fashion. The steps involved in the standard PCA algorithm could be summarized as follows:

1. Computation of the centered data matrix (normalization of the variables, so that the mean of any given variable is zero).
2. Construction of the covariance matrix from the normalized data matrix.
3. Finding the eigenvectors and eigenvalues of the covariance matrix.
4. Formation of a new set of features, that successively maximize variance observed in the original dataset (by multiplication of the eigenvectors with the original data matrix).

Below, we will explore the standard derivations of the algorithm, showing that the PCA could be solved as an eigenvalue problem or as a singular value decomposition problem. The main equations will be provided to outline the computational logic and equivalence between the eigendecomposition of the covariance matrix with the singular value decomposition of the data matrix. All notations and derivations used here provided as in (Jolliffe and Cadima 2016), unless is stated otherwise.

The standard representation of the method builds on a dataset with p numerical variables and n observations. Thus, data are defined as a set p (number of spectra) of $n-$ dimensional vectors $x_1, ..., x_p$ (intensities of given mass units) that form the matrix **X** of original mass spectrometric observations. The dimensionality reduction part of the algorithm stems from the fact that we want to find linear combinations of columns (single unit masses) in the matrix **X**, that have maximal variance, thus, preserving the information about major chemical composition and omitting dimensions with small variance. These linear combinations are given by $\sum_{j=1}^{p} a_j x_j = \mathbf{X}\boldsymbol{a}$, where $\boldsymbol{a}$ is a vector of constants $a_1, a_2, ..., a_p$. The variance of such linear combinations is given by $\mathrm{var}(\mathbf{X}\boldsymbol{a}) = \boldsymbol{a}'\mathbf{S}\boldsymbol{a}$, where **S** is the covariance matrix of the data matrix, and ' denotes transpose.

$$S = \begin{pmatrix} S_{1,1} & \cdots & S_{1,n} \\ \vdots & \ddots & \vdots \\ S_{n,1} & \cdots & S_{n,n} \end{pmatrix}$$

The identification of maximum variance can be formulated as obtaining a p – dimensional vector $\boldsymbol{a}$ which will maximize the $\boldsymbol{a}'\boldsymbol{Sa}$. To have a straightforward solution, we can impose a requirement, that $\boldsymbol{a}'\boldsymbol{a} = 1$. Thus, the solution will be equivalent to maximizing $\boldsymbol{a}'\boldsymbol{Sa} - \lambda(\boldsymbol{a}'\boldsymbol{a} - 1)$. Therefore, we can reformulate our equation as a matrix factorization problem of given covariance matrix:

$$\boldsymbol{Sa} - \lambda\boldsymbol{a} = 0 \Leftrightarrow \boldsymbol{Sa} = \lambda\boldsymbol{a}, \qquad (8)$$

Thus, $\boldsymbol{a}$ must be a (non-zero) eigenvector, and λ (λ_k, ($k = 1, \dots, p$)), the corresponding eigenvalues of the covariance matrix $\boldsymbol{S}$.

Any symmetric matrix (for example, our covariance matrix $\boldsymbol{S}$), has p real eigenvalues and together with its eigenvectors. one can define an orthonormal set of vectors. Hence, the complete set of eigenvectors of the covariance matrix $\boldsymbol{S}$ are the solutions to the problem of obtaining up to p new linear combinations $\boldsymbol{Xa_k} = \sum_{j=1}^{p} a_{jk} x_j$, which successively maximize variance and are uncorrelated with other pairs. Thus, these new linear combinations $\boldsymbol{Xa_k}$ are called the principal components of the dataset, and it is important to note that they have a physical interpretation (change of the major chemical composition in case of analysis of first PCs for LIMS data). In the standard PCA terminology, the elements of the eigenvectors $\boldsymbol{a_k}$ are called the PCA loadings, whereas the projections of linear combinations $\boldsymbol{Xa_k}$ are called the PCA scores. Overall, the largest eigenvalue, λ_1 and its corresponding eigenvector $\boldsymbol{a_1}$ retains the information about the first principal component, which represents the maximum variance in the dataset. Thus, the second principal component is computed in such a way that it reflects the second largest source of variation in the data.

It is also possible to find PCs as the linear combinations of the normalized variables x_j^*, with an element $x_{ij}^* = x_{ij} - \bar{x}_j$, where $\bar{x}_j$ denotes the mean values of the observation on variable j. The normalization does not change the solution other than just centering it around the mean, but normalized data matrix provides a connection to another approach of finding principal components using alternative matrix factorization method – singular value decomposition.

2.2.2 Singular value decomposition

Building on previously outlined notation and denoting by $\boldsymbol{X}^*$ the $n \times p$ matrix, where columns are represented by centered variables x_j^*, we can rewrite our equivalence equation:

$$(n - 1)\boldsymbol{S} = \boldsymbol{X}^{*'}\boldsymbol{X}^*, \qquad (9)$$

The equation 9 connects the eigendecomposition of the covariance matrix $\boldsymbol{S}$ with the singular value decomposition (SVD) of the centered data matrix $\boldsymbol{X}^*$. In fact, any matrix $\boldsymbol{Y}$ of size $n \times p$ and rank r (with some minor restrictions, see (Jolliffe and Cadima 2016)) can be written as:

$$\boldsymbol{Y} = \boldsymbol{ULA}', \qquad (10)$$

Where $\boldsymbol{U}$, $\boldsymbol{A}$ are $n \times r$ and $p \times r$ matrices with orthonormal columns ($\boldsymbol{UU}' = \boldsymbol{I}_r = \boldsymbol{A}'\boldsymbol{A}$, were $\boldsymbol{I}_r$ is an $r \times r$ identity matrix). Further, $\boldsymbol{L}$ is an $r \times r$ diagonal matrix. The columns of $\boldsymbol{A}$, are the right singular vectors of $\boldsymbol{Y}$ and the eigenvectors of the $p \times p$

matrix $\mathbf{Y'Y}$. The columns of $\mathbf{U}$ are the left singular vectors of $\mathbf{Y}$ and the eigenvectors of the $n \times n$ matrix $\mathbf{Y'Y}$. The diagonal elements of matrix $\mathbf{L}$ are called the singular values of $\mathbf{Y}$ and are the non-negative square roots of the non-zero eigenvalues of both matrix $\mathbf{Y'Y}$ and matrix $\mathbf{YY'}$ (Jolliffe and Cadima 2016).

Using the assumption that diagonal elements of $\mathbf{L}$ are in decreasing order, we can define order of the columns of $\mathbf{U}$ and $\mathbf{A}$. Therefore, substituting matrix $\mathbf{Y}$ with $\mathbf{X}^*$, the right singular vectors of the column-centered data matrix $\mathbf{X}^*$ are the vectors $\boldsymbol{a_k}$ of PC loadings. Equivalently:

$$(n-1)\mathbf{S} = \mathbf{X}^{*\prime}\mathbf{X}^* = (\mathbf{ULA'})'(\mathbf{ULA'}) = \mathbf{ALU'ULA'} = \mathbf{AL^2A'}, \quad (11)$$

Where $\mathbf{L^2}$ is the diagonal matrix with squared singular values (i.e., the eigenvalues of $(n-1)\mathbf{S}$). Thus, equation (11) provides the spectral decomposition or eigen-decomposition of matrix $(n-1)\mathbf{S}$, hence, the principal components can be obtained as SVD of the column-centered data matrix $\mathbf{X}^*$.

Here, we also have to outline that quality of the PCA or SVD reduction depends on the number of PCs retained for further analysis. Therefore, we have to define a cutoff threshold for the number of components. The standard measure of quality of the given principal component is a proportion of variance that it explains, thus, such measure can be recalculated as:

$$\pi_j = \frac{\lambda_j}{\sum_{j=1}^{p} \lambda_j} = \frac{\lambda_j}{tr(\mathbf{S})'}, \qquad (12)$$

Where $'$ denotes the transposed trace of $\mathbf{S}$. Accordingly, we can recalculate total percentage of the variance that we accounted for by using the defined set of PCs. It is a common practice in machine learning to account for 70% of the variance, assuming that the remaining part of the data variance is primarily represented by noise. However, in the analysis of PCA reduced LIMS data, the chemical composition of the trace elements often important; thus, the attention may be needed to the last few components that typically preserve such information (dimensions with small variance). At the same time, first PCs represent composition of the major chemistry present in the dataset. Another common approach to define number of PCs to retain is based on slope analysis of the scree plot - line plot of the eigenvalues that displays them in the downward curve. The rapid change in slope of the line indicates location of the "elbow" or breaking point of the graph, where internal variance changes maximally; thus, the PCs can be retained as significant (such test also called the "Kaiser rule"). However, more complex datasets with nonlinear variances are prone to have multiple breaking points, and this approach is often criticized.

Here, we also need to mention some of the limitations and assumptions of PCA:

1 PCA assumes that underlying data have correlated features and that correlated features are linearly dependent on each other
2 PCA is sensitive to the scale of the variance; therefore, it is not robust against outliers
3 Nonlinear dependencies will manifest themselves with many significant PCs.
4 Data loss is unavoidable, and the amount of approximated variance should be maximized. Thus, the analysis of only first PCs can be misleading.
5 Covariance matrices can be large, so initial preprocessing can be of high utility.

Modern machine learning libraries such as scikit-learn (Pedregosa et al., 2011) provide high-level language access to the standard algorithms such as PCA using library-wide application programming interface (API).

Overall, PCA in its standard form is widely used in various scientific areas; however, there are many different adaptations of the technique, such as Kernel PCA, Sparse PCA, Truncated SVD, Incremental PCA, etc., that are all might be of high utility for analysis of high-dimensional data.

2.2.3 Laplacian Eigenmaps

The Laplacian Eigenmaps (LE), also known as Spectral Embedding, is another matrix factorization method and a dimensionality reduction algorithm developed by (Belkin & Niyogi, 2003). The algorithm provides a computationally efficient method for nonlinear dimensionality reduction that has locality-preserving properties and a natural connection to clustering. The LE builds on a previously presented eigen decomposition, but in contrast to PCA, does that using affinity matrices (distance or similarity matrix, see further section "generic similarity measures"). The output of the algorithm is a set of new dimensions called "maps" that consecutively approximate the intrinsic dimensionality of the data. Thus, the resulting transformation of the original data points is given by the value of corresponding eigenvectors.

Let $p \times p$, be a symmetric matrix $\mathbf{V}$ (where p represents a number of n-dimensional observations) of pairwise affinities defined using metric $d(x, y)$. Now, let diagonal matrix $\mathbf{D}$ contain row sums of $\mathbf{V}$, or equivalently, $D_{ij} = \sum V_{ij}$. Now, we can introduce a new matrix $\mathbf{L}$, so that $\mathbf{L} = \mathbf{D} - \mathbf{V}$. Such matrix is known as graph Laplacian of the affinity matrix $\mathbf{V}$. Thus, to find sets of observations that share a significant similarity, we can formulate the problem as an eigen decomposition of the graph Laplacian. Only the main equations are provided, for details see (Belkin & Niyogi, 2003).

$$\mathbf{La} = \lambda \mathbf{Da}, \qquad (13)$$

Further, by multiplying each side of this equation by $\mathbf{D}^{-1}$, the decomposition can be reformulated as finding the set of eigenvectors of $\mathbf{D}^{-1}\mathbf{V}$.

$$\mathbf{D}^{-1}\mathbf{Va} = (1 - \lambda)\mathbf{a}, \qquad (14)$$

The matrix $\mathbf{D}^{-1}\mathbf{Va}$ is not symmetric and has rows normalized to 1. The physical interpretation of the matrix $\mathbf{D}^{-1}\mathbf{Va}$ could be viewed as a diffusion operator on the graph, making it equivalent to another nonlinear data reduction method called Diffusion maps (Coifman & Lafon, 2006).

Another way of solving the LE could be done by using the normalized graph Laplacian:

$$\mathbf{L}_{\text{norm}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}, \qquad (15)$$

And solve an eigenvector problem: $\mathbf{L}_{\text{norm}}\mathbf{b} = \lambda\mathbf{b}$, (4), where $b = \mathbf{D}^{-\frac{1}{2}}\mathbf{a}$

Thus, the Laplacian eigenmaps, similar to PCA, can produce a set of "PC's" that capture the structure of the local neighborhood. However, as with many other manifold learning methods, the quality of embeddings can heavily depend on the specific similarity or distance function. Therefore, if the similarity function applied to the dataset does not properly capture the structure of observations, resulting spectral embedding will not be very informative either. See (Deza & Deza, 2009) for an extensive

description of various distances and similarity measures. Need to mention that cosine similarity typically better preserves the information of structural heterogeneity in LIMS data analysis in comparison with Euclidean distances. Therefore, it can be considered a default choice of similarity function in the LE decomposition of mass spectrometry data. See further chapter "Generic similarity measures" for more in-depth discussion of various distances and similarity measures.

It is important to note that LE in comparison to other manifold learning methods, cannot embed into lower dimensional space the out-of-sample data points, thus, requiring a new LE decomposition as new points are acquired. Another important point refers to the fact that LE decomposes the similarity graph; therefore, the graph construction poses a critical constraint on the algorithm. A neighborhood graph from the given data can be constructed using the k-NN approach (see further section "Graph-theoretic measures and centralities" for a more in-depth description of k-NN networks and different types of networks) or using the ϵ-nearest-neighbor (ϵ-NN) graphs. The ϵ-NN stands for recovering the ϵ-neighbors (neighbors in the given ϵ radius) within a given metric. To produce the k-dimensional embedding, as it was mentioned before, LE allows doing that by using more eigenvectors of the normalized graph Laplacian. Although the LE deals with spectral graph theory and heavily relies on k-NN networks, it remains a matrix factorization method for dimensionality reduction. As it will be shown later in the section describing graph-based dimensionality reduction methods (i.e., Uniform Manifold Approximation and Projection - UMAP), the LE decomposition will be important in providing the dimensionality parameter on the output embedding graphs. Here we have also need to mention that LE naturally connects to the clustering through the important application of the Laplacian - spectral clustering (SC) that corresponds to a computationally tractable solution to the graph partitioning problem. The spectrum of the normalized graph Laplacian also provides a unifying view into the grouping of nodes in the network. The complete range of all eigenvalues will fall in the range between 0 and 2, enabling the comparison of networks with different sizes. The presence of large eigenvalues can indicate that network has a strong modular structure. The similarities between spectra of the normalized graph Laplacians can be used to classify networks. Therefore, the LE as a useful dimensionality reduction method also provides insight into a set of related with data additional matrices that are of high utility in analyzing high-dimensional observations.

Concluding, most of the dimensionality reduction (DR) methods require an injection of some degree of prior knowledge (by choosing k and number of the output dimensions). Thus, the unsupervised data characterization still requires careful assessment of the reduction results to achieve high classification and prediction accuracies. The interested reader can further explore other manifold learning methods such as Isomap, Diffusion maps, Locally-linear embedding. There is also a number of methods that use the neural networks to learn the latent embedding spaces, such as Kohonen's Self-organizing maps (SOM's) and Autoencoders. The autoencoder networks use the specific architecture that has a bottleneck in their structure, which forces a formation of compressed knowledge representation (by data pooling) of the original input, thus, representing another approach for dimensionality reduction. The autoencoders are a distinct type of rapidly developing DR methods, although they provide the least amount of interpretability.

2.3 Graph-based methods

2.3.1 Graph-theoretic measures and centralities

The first application of the network analysis can be traced back to Leonard Euler, who famously solved the Königsberg Bridge Problem using reduced network representation of the city and its bridges. Since Euler, graphs are widely used almost in all aspects of science. They naturally encode information from social networks (Borgatti *et al.*, 2009), computer networks in cybersecurity (Wu & Irwin, 2016), and visualize spatially resolved brain connectomes (Farahani *et al.*, 2019), to name a few. In recent years specific graph-oriented data libraries were developed, making graph-based data processing more available to the general scientific audience (Kumar Kaliyar, 2015; Pedregosa *et al.*, 2011).

In this subchapter, I will provide a summary of graph-theoretic measures commonly used in network science. I will also outline the importance of network-based statistics and network segregation methods in visualizing and classifying extensive mass spectrometric data. The relational data modeling provides the visualization of edge-node diagrams that encode the relationships between observations and reflect the structural changes, in which new, informative properties of investigated materials can emerge.

In general, the graph (network) consists of a number of vertices (nodes) connected with links (edges). As illustrated in Fig. 8, graphs can be drawn as a diagram of nodes connected with varying numbers of edges, where nodes and edges can encode domain-specific information. The networks can be binary (edge is encoded as 0 – no connection, and 1 – there is a connection) or weighted. The weighted graph edge encodes the strength of the connection. For example, in the normalized form, weights can be encoded in the range from 0 to 1, where 0 means no connection, and one denotes the strongest connection available.

Figure 8 illustrates the types of networks that can be utilized in mass spectrometry data analysis. The individual observations (mass spectra) can represent the individual nodes in the network, and edges can encode similarities between spectra defined on some metric space. If the similarity function is not symmetric (i.e., concave functions like logarithm) and does not qualify as a metric (i.e., the distance between spectrum a and b is not the same as a distance between b and a), the network can be constructed as a directed weighted network (network in the right, Fig.8) visualizing either one direction: $a \rightarrow b$, or another: $b \rightarrow a$. Therefore, the information on edges can be stored as directional weight that encodes the structure of spectral proximity (see (Deza & Deza, 2009) for the list of metric and non-metric distances).

If every mass spectrum is connected to its k nearest neighbors, we can form a k-neighborhood graph (k-nearest neighbors' network or k-NN) or spectral similarity network. The k-NN graph can be constructed as an undirected binary network, where edges are encoded in a binary fashion, or as a weighted undirected network, where the edges can encode actual similarity measures on its edges. Although it is not very informative, it is possible to compile a farthest neighbor graph (FNG) where nodes can be connected to the k-farthest neighbors encoding the most dissimilar spectra in the network, though large distances in high dimensions can be meaningless or misleading. As it was mentioned before it is also possible to construct ϵ-nearest-neighbor (ϵ-NN) graphs.

Another available approach is that nodes can represent the single unit masses, and edges between pairs of masses can encode the correlation values. In this case, we will get mass correlation networks constructed similarly to functional magnetic resonance imaging (fMRI) brain connectomes and gene co-expression networks (Horvath, 2011).
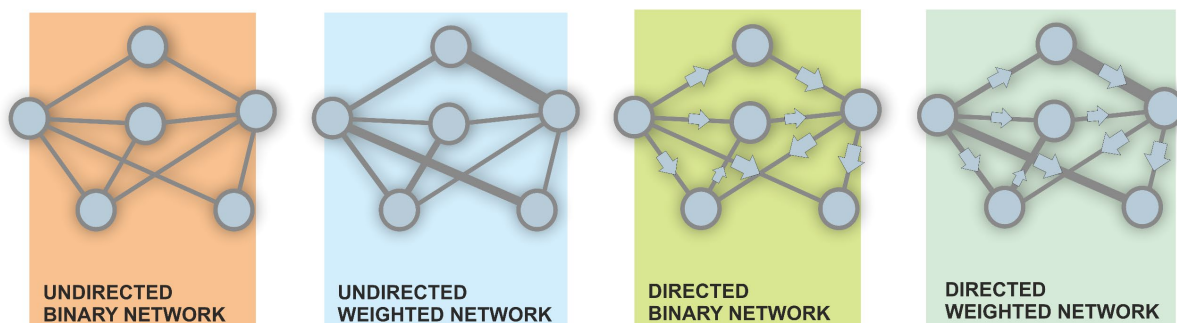
Figure 8. Types of networks that can be used in the analysis of mass spectrometry data. The figure is adapted from (Farahani et al., 2019). Note the thickness of edges

The spectral similarity networks (k-NN's) and correlation networks can be static or dynamic (static networks chained together). All previously described types of networks can be dynamic (i.e., changing with some increment value, defined on the dataset). Further in the thesis, some presented networks are dynamic (i.e., correlation networks from the depth profiles). However, they will be presented as a set of static networks. The directed binary networks can also be constructed, encoding the direction of causal effects.

As illustrated in Fig. 9, the networks can have an intricate structure, where a multitude of graph-theoretic measures can be applied. The figure demonstrates commonly used global graph measures. The figure on the left indicates the modularity structure of the network that can be uncovered using the greedy modularity optimization algorithm, i.e., Louvain (Blondel *et al.*, 2008) or the recently proposed Leiden algorithm (Traag *et al.*, 2019). Modularity score describes the density of links inside a community compared to the density of links outside communities. In fs-LIMS data analysis, the discovered modules correspond to the chemical entities present in the data set (i.e., minerals, organic remains of microfossils). The middle section of Fig. 9 illustrates the measure called the shortest path, or geodesic path. The measure encodes the path between nodes in the network that consists of a minimal number of edges. If the network is weighted, the shortest path represents the minimal sum of edge weights. The geodesic path length represents the distance defined on the network (it may not be a unique value). For example, the shortest path length can be used in measuring the distance from mass spectra registered from one mineral to the mass spectra registered from the microfossil. Thus, it represents the secondary metric that could be defined on the body of mass spectrometric observations registered from heterogeneous media. One can also define the diameter of the network, which characterizes the largest geodesic distance in the network. Geodesic distances can also be useful in characterization of the diameters of the subgraphs which are indicative of the cluster sizes. The consequence of application of graph-based distances is that it is always possible to define a central node which in mass spectrometric analysis represents the most representative spectrum from the given mineral or more generally it reveals a measure of the influence of a node in a network.
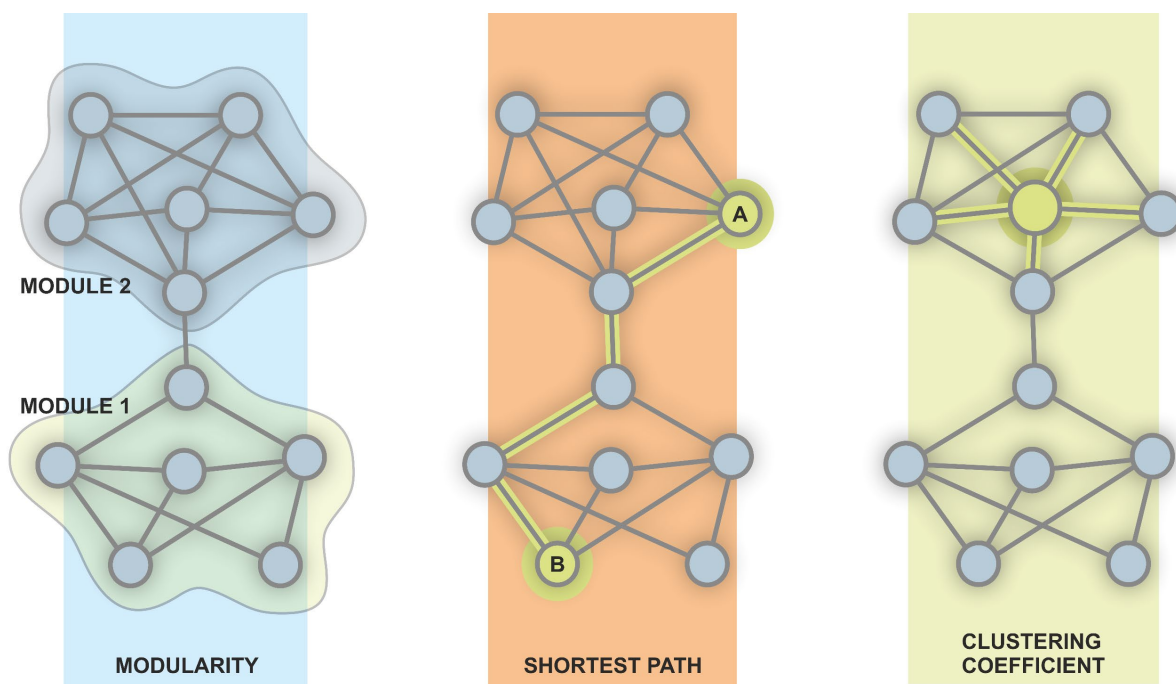
Figure 9. Illustration of graph-theoretic measures that emerge from the network structure. The figure is adapted from (Farahani et al., 2019).

The typical path lengths in the "small world" networks are scaling as $\log N$, where N is the number of nodes in the network. Thus, if the analyzed network is a k-NN graph of spectral similarities, with a number of observations $\sim$40'000, the expected shortest path length will be only 4.6, indicating that expected distances on graphs tend to remain small. Several efficient algorithms are known for calculating geodesic distances on graphs, such as Dijkstra's algorithm, Bellman-Ford algorithm, and the A* search algorithm, which uses the heuristic approaches to speed the search times.

The clustering coefficient represents the measure of the degree to which nodes in the network have a tendency to form a cluster of nodes. The coefficient can be used to color and segregate the network and potentially find the aggregations of nodes if they cannot be separated into communities right away. The degree centrality illustrated in Fig. 10, also called just degree, identifies the node's centrality in the network calculated from the number of edges that the node contains (or simply a number of edges per node). For example, the degree of the central yellow node in the figure is 6, whereas the degree of smaller singular nodes (sometimes called "leaves") is one. Degree centrality is directly linked to the average degree of the network and represents the easiest measure to compute. The betweenness centrality (BC) identified in fig. 10 (middle illustration) represents the measure of centrality between communities. For example, the BC can be used to identify mass spectra that belong to the transition structure between classes of mass spectra. Suppose the module on the left represents hypothetical mass spectra registered from the host mineral and the module on the right - organic inclusions. In that case, the BC measure will identify mass spectra characterized by an equal mixing ratio of host over organic material and outline the location of transitional spectra.

The final for this chapter centrality measure is shown on the right, in Fig. 10. The eigenvector centrality is a self-referential measure of centrality that takes into account the quality of the link. The measure does it in a way that if the node is connected to a more central node, it will increase its centrality (of a connected node) in return.

This effect is shown in the figure with the centrality of the yellow node, which is more central in comparison with the centrality of the node on the right, notwithstanding that degree of compared nodes is the same.

Many more other centrality measures can be of high utility in analyzing the mass spectrometry data. For example, measures such as Katz centrality, Harmonic centrality, Percolation centrality can be defined on the topology of the produced networks (Bonacich, 2007; Borgatti *et al.*, 2009; Farahani *et al.*, 2019). The centrality measure called PageRank is a modification of the eigenvector centrality developed by Google Search and used to determine the reputation and importance of the web page in the network of websites.

DEGREE CENTRALITY   BETWEENNESS CENTRALITY   EIGENVECTOR CENTRALITY
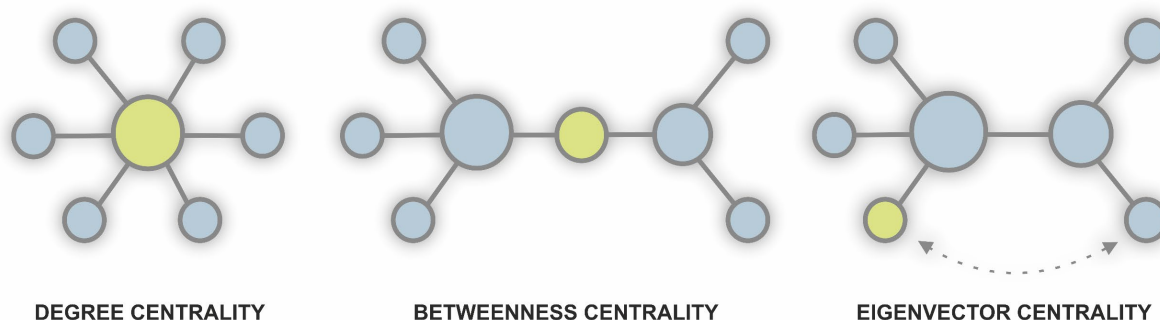
Figure 10. Basic concepts of network centralities. The most central node is identified with yellow color (except for eigenvector centrality, see the text). Sizes of nodes are scaled according to the nodal degree. The figure is adapted from (Farahani et al., 2019).

In general, the measures applied to the graph are divided into two categories - global and nodal. The global measures derive a single number for the graph, thus characterizing the network structure in its complete shape. The nodal characteristics derive specific values for the given node. In addition to the measures described above, we can mention the eccentricity, the measure that describes the maximal distance from the given node to all remaining nodes in the network. Global efficiency describes the average of the inverse shortest path length from a node to all other nodes. The efficiency of a network is a measure of how efficiently it exchanges information. In contrast to previously discussed measures, the efficiency can be applied as a global and local measure and represents another way of measuring the connectivity structure of the given graph. Although the efficiency is not directly utilized in the analysis of mass spectrometric networks in this thesis, it can represent an interesting measure describing the structure of the entire dataset. The measure called participation coefficient measures the distribution of a node's edges among the communities of a graph. A small-worldness (global) describes the measure of similarity of the given graph to the structure of the small-world graph. For example, the spectral similarity networks from the single mineral with low spectral dispersion of intensity values will grow as spheres (in 3D) and circles in 2D, thus, building the network that has a small-world organization. In other words, the small-world graph organization can be characterized by a high clustering coefficient and a short mean path between nodes. Another measure that can be defined on the given network is the assortativity coefficient which represents the Pearson correlation of the nodal degree between pairs of linked nodes. The positive correlation indicates that nodes tend to link to other nodes with a similar degree or strength.

2.3.2 Generic similarity measures

The exploration of large spectral datasets often requires understanding how one observation is different from another without looking into original spectra. This is an important issue in clustering, classification, outlier detection, and general explorative characterization of datasets. For example, spectra from the same mineral are expected to be similar to each other if laser powers and the output ion yields are held constant.

This section provides information on common similarity and distance measures that can be computed for sets of observations. The notations and descriptions of distances are provided as in (Han *et al.*, 2012)

Here, we can define that similarities between mass spectrometric observations can be expressed as a measure of distance and vice versa:

$$sim(i,j) = 1 - d(i,j), \qquad (16)$$

Where $sim(i,j)$ is the similarity between objects i and j, and $d(i,j)$ is the distance between objects i and j. Thus, the similarity is an inverse of the distance measured between two observations.

The most straightforward and commonly used distance is Euclidean distance (or the straight line between two points in 2D). Let $i = (x_{i1}, \dots, x_{ip})$ and $j = (x_{j1}, \dots, x_{jp})$ be two observations with p numeric attributes. The Euclidean distance between these observations could be defined as follows:

$$d(i,j) = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{ip} - x_{jp}\right)^2}, \qquad (17)$$

Another commonly used distance measure is Manhattan distance; it is defined as:

$$d(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \cdots + |x_{ip} - x_{jp}|, \qquad (18)$$

Both distances provide the following properties – the distance between two objects cannot be negative, and distance to the same objects is zero (distance to itself). Also, the functions are symmetric, meaning that $d(i,j) = d(j,i)$ and follow the triangle inequality. The measures that satisfy these properties are known as a metric.

The Minkowski distance is a generalization of the Euclidean and Manhattan distances:

$$d(i,j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \cdots + |x_{ip} - x_{jp}|^h}, \qquad (19)$$

Where h is a real number such that $h \geq 1$, the Minkowski distance is equal to Euclidean distance where h=2 and equals to Manhattan distance when h=1.

Another very popular measure is cosine similarity. It measures the cosine of the angle between two vectors and determines whether they are pointing in the same direction. It became a standard similarity measure in text mining, where text vectors produce sparse matrices. The cosine similarity function is defined as:

$$sim(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x}\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}, \qquad (20)$$

Where $\|x\|$ is the Euclidean norm of vector $x = (x_1, x_2, \ldots, x_p)$, and defined as $\sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}$, and similarly $\|y\|$ is the Euclidean norm of vector y.

The cosine similarity, thus, measures the cosine of the angle between two vectors. The value of the function close to 0, means that vectors are located at around 90 degrees to each other (or orthogonal/dissimilar). In contrast, values close to 1, identify very close proximity between vectors, whereas a value of exactly 1 means that vectors are identical. It is important to note that cosine similarity is not a proper metric because it does not obey the rules described above.

During the course of this thesis, we will use the Euclidean distances and Cosine similarity to measure the closeness of mass spectra to each other. In general, cosine similarity better captures the differences between various minerals; however, it does not reflect the change in magnitude. Thus, various notions of distances need to be assessed prior to the full analysis because there is no largely accepted notion of distance that works for all datasets. It all depends on the spareness of vectors and their magnitudes.

Overall, in this chapter, we covered only small portion of all available distances. A great variety of data-type dependent distances can be further explored. For example, distances on graphs (i.e., geodesic distances) or distances between distributions (Wasserstein distance) can be important for the characterization of different chemical entities measured with LIMS (Deza & Deza, 2009).

The mineral label assignment stabilities can also be assessed using the similarity score defined for clustering results. The Rand Index (RI) is a measure of similarity between two clustering results. The form of the Rand Index that is adjusted for the chance grouping of spectra (data points) is called the Adjusted Rand Index. The Rand index represents the frequency of occurrence of agreements over the total number of pairs. The RI score changes from 0 to 1, where 1 represents the complete overlap in the cluster assignment, and 0 indicates that there is no agreement between cluster assignments. Therefore, the RI can represent an analog to the p-value for the given clustering if robustness is estimated using the random subsamples.

With an increase in the data acquisition speeds, the output of the LIMS data analysis can represent multiple networks; therefore, it might be useful to compare the similarity of networks to each other. As it was mentioned before, the spectrum (set of the eigenvalues) of the normalized graph Laplacian can be used to compare networks with different sizes. Although it does not represent a proper distance measure, it reflects the structural similarity of investigated networks (see the spectral distances on graphs, i.e., Wasserstein distances between spectra of the graphs). The methods such as Euclidean, Jaccard, or DeltaCon can be applied to compare various networks with each other (Tantardini et al. 2019). If the analyzed networks are very complex and large, network coarsening approaches can be utilized. The Mapper-type algorithms can be applied to the node embeddings to further reduce graphs (see further section "Mapper" for more detailed characterization of the method). For example, the Mapper algorithm was initially developed with the purpose of comparing 3D shapes, which are essentially complex networks (Singh et al. 2007). Though networks can lose fine structural information, authors reported a high similarity score for analogous networks using the modification of the Gromov–Hausdorff distance; thus, the method is certainly applicable to shape comparison and matching tasks.

2.3.3 Louvain modularity

The Louvain method developed by (Blondel *et al.*, 2008) is a community detection algorithm used on networks. The algorithm represents the heuristic method of greedy modularity optimization. The Louvain modularity represents a fast and high-quality community detection method widely used in various scientific fields. Arguably, the Louvain method is one of the most popular algorithms for uncovering the community structure of networks on par with the Leiden algorithm (Traag *et al.*, 2019), although the latter algorithm claims an improvement in the quality of community detections.

In fs-LIMS data analysis, correlation networks or the low dimensional embeddings can yield densely packed communities of similar spectra representing mineralogical entities, or for example, they can represent the groups of spectra measured from microfossils. However, the structure of these groups is not known, and depending on the topological structure of the network, recovering the community structure might not be easy or self-evident. Thus, the Lovain modularity optimization algorithm, given the connectivity structure of the network, can infer the mineralogical composition of the analyzed set.

The best-known way to recover the modular structure of networks is so-called modularity (Newman & Girvan, 2004). The modularity score defined as a measure of the density of nodes inside the community compared to the density of nodes outside communities. However, optimization of the original modularity score was found to be computationally expensive (the algorithm estimates all possible iterations of the nodes into groups); thus, new methods of approximate calculation were needed.

The Louvain method calculates the modularity score by finding modularity on all nodes, then it aggerates nodes in the networks into smaller communities and repeats the calculation of modularity on the reduced graph.

Accordingly, the modularity of the weighted network is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i c_j), \qquad (21)$$

Where, A_{ij} represents the edge weight between nodes i and j. k_i and k_j are sums of the weights of the edges attached to nodes i and j. m is the sum of all the edge weights in the graph. c_i – is the community to which node i is assigned, and $\delta -$ is the delta function, such that $\delta(u, v)$ is 1 if $u = v$, and 0 otherwise (Blondel *et al.*, 2008).

The Louvain algorithm is divided into two phases that are repeated iteratively. In the first phase of the Louvain method, every single node in the graph is assigned with its own community. Then, nodes are aggregated into smaller communities, and the whole process is repeated until no further improvement can be achieved (maximization of the parameter Q). The modularity is optimized by allowing only local changes of communities. Where the algorithm found communities, nodes are aggregated further in order to build a new network of communities.

The gain in modularity (ΔQ) can be computed as follows:

$$\Delta Q = \frac{1}{2m} \sum_{ij} \left[\frac{\Sigma_{in} + 2k_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right], \qquad (22)$$

Where Σ_{in} is the sum of all weights of the links the community i is moving into. Σ_{tot} is the sum of all the weights of the links to nodes in the community i is moving into. k_i is the weighted degree of community i; $k_{i,in}$ is the sum of the weights of the links between community i and other nodes in the community that i is moving into, and m is the sum of the weights of links in the network.

The Louvain clustering works well in junction with the Mapper algorithm (see further section - Mapper). The experimental results can yield a highly complex distribution of points that are not always easy to partition into clusters. Modern methods of clustering, for example, DBSCAN, rely on the density distributions of the data to find reliable clusters. However, it is often the case that data can present clusters with varying densities and complex internal structures. However, such groups of data points that can be easily identified by eye are typically hard to partition into clusters using numeric approaches. The Mapper algorithm provides the connectivity structure of the data on the given scale, producing the network as an output. Consequently, this network can be segregated into communities using the Louvain algorithm (see contribution "On topological analysis of fs-LIMS data" for examples of graph partitioning using the Louvain method). As in all clustering algorithms, a user has to define hyper-parameters to achieve a desirable outcome. In the case of Louvain, there is only one parameter that controls the granularity of the graph partitioning. The parameter called "resolution" can be defined above 1, to find large-scale communities and below 1, to further partition the graph into a smaller set of connected groups of nodes.

In general, by using the junction of Mapper and Louvain, the user has to be less "lucky" to partition data into meaningful clusters in comparison with conventional methods, like, k-means, DBSCAN, spectral clustering or agglomerative clustering. During the course of this thesis, I will also present another clustering algorithm called HDBSCAN, which uses hierarchical representations of density distributions, thus allowing for clustering more complex data (Campello et al. 2013; McInnes et al. 2017). Even though, HDBSCAN works very well with most of the datasets, the pair of Louvain with Mapper provides more stability (see supplementary information for "On topological analysis of fs-LIMS data") and provides better tolerance to the noise in data. Thus, the data points that can be assigned as noise in most of the clustering algorithms can be retrieved within a proper cluster using topological methods and community detection on graphs.

The Louvain algorithm works well for large and small networks - as tested in the original contribution, the method can be applied to uncover modular structure of the network with millions of nodes. The results of partitioning of small networks can be seen in the contribution "Towards empirical biosignatures using LIMS", where correlation networks were segregated into communities, revealing elements from minerals and microfossils. The exemplary Lovain partitioning of the correlation network is illustrated in Fig. 12 where the community detection algorithm found three clusters that belong to microfossils and two minerals.

Overall, the Louvain method derives highly informative community structures from complex networks, therefore, allowing to uncover mineralogical and chemical entities from LIMS observations.

2.3.4 Weighted mass correlation networks

The weighted mass correlation analysis (WMCA) (or just correlation networks) is an analytical method that naturally arises from PCA, and in some sense, represents the alternative solution for eigendecomposition of centered covariance matrices. Originally, the method was developed by (Horvath, 2011) and named Weighted Gene Co-expression Network Analysis (WGCNA). Initially, the method was applied to the analysis of gene expression profiles; thus, it entered other fields through genetics and systems biology. At its core, the WMCA is a data visualization technique and method for the interpretation of high-dimensional linear observations. The analyzed data in correlation networks are encoded as a set of nodes and edges and further visualized using graph spatialization methods.

Here we will define the main computational parts of the algorithm and show an exemplary network, outlining the pros and cons of the method.

In fs-LIMS analysis, minerals and chemical entities on a relatively small scale of observations tend to form linear ion yield profiles. The variance of ion yields can be caused by unequal ionization of portions of the mineral due to the inhomogeneities present on the grain boundaries and due to the crater shape evolution that changes during ablation. Thus, if we have a single mineral, all the masses defined by the stoichiometric formula of mineral, in principle, should be linearly correlated. However, it is not always the case – some minerals can be on a microscale more hydrogenated, weathered etc., others are less. Also, the trace elements with impurities present within minerals can be not abundant enough to register a high-quality correlation factor between them. Thus, we can map the full scale of all linear variances and lay them out in the form of the network.

Let $\mathbf{X}$ be the data matrix that consists of n random variables ($\mathbf{X} = (X_1, \dots, X_n)$), then the correlation matrix of X, can be defined as follows:

$$corr\ (\mathbf{X}) = \ \mathbf{S} = \begin{bmatrix} 1 & \cdots & \dfrac{E[(X_1 - \mu_1)(X_n - \mu_n)]}{\sigma(X_1)\sigma(X_n)} \\ \vdots & \ddots & \vdots \\ \dfrac{E[(X_n - \mu_n)(X_1 - \mu_1)]}{\sigma(X_n)\sigma(X_1)} & \cdots & 1 \end{bmatrix}, \qquad (23)$$

Where, $E[(X_x - \mu_X)(X_y - \mu_y)]$ denotes the covariance of the given mass pairs - $cov(x, y)$ in $\mathbf{X}$, μ_y denotes the expected value, and σ denotes standard deviation.

Given the number of variables in the matrix X, we can define the number of unique correlation pairs:

$$\mathrm{R} = \left(\frac{n^2}{2}\right) - n, \qquad (24)$$

Using the correlation matrix $S = [s_{ij}]$, we can form an adjacency matrix $A = [a_{ij}]$, that can be used to define network $G = (V, E, w)$, where V are the vertexes in the graph, E pairwise edges between vertexes, weighted as s_{ij} and w denotes a vector of attributes (i.e., element and isotope abundances, centrality measures).

The network has to be set to the specific threshold (hard threshold), to remove the clutter of insignificant correlation factors from the graph. Such threshold splits the network into separate parts, where insignificant correlation pairs are removed. In contrast to the hard threshold, a soft threshold uses powered ($s_{ij}{}^b$) correlation

matrix, thus preserving more of the fine structure of linear and near-linear dependencies. Additionally, the negative correlation factors cannot be used in the spatialization of the network; thus, they should be separated and visualized as a separate network of negative correlation factors.
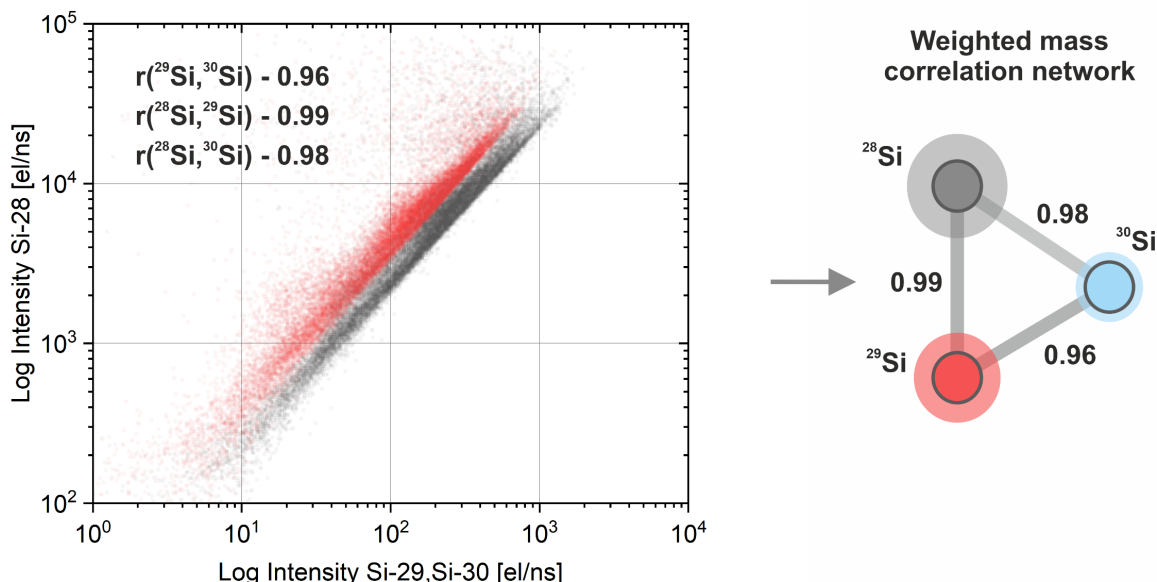
Figure 11. An illustration of the weighted mass correlation analysis (WMCA) workflow. Left – distribution of Si isotope intensities measured from the Gunflint chert. Right – correlation network of Si isotopes. The varying circles around nodes encode isotope abundances.

As shown in Fig. 11, correlation networks can encode a relatively large number of observations (4400 mass spectra are visualized in the figure) with linear structure into a concise and informative network that preserves the structure of co-dependencies present in the original dataset. The parameter w, defined in the network, is visualized as varying circles around every node. In principle, it can be any type of secondary metric associated with the network. For example, w can encode node centrality in the network or averaged abundance of the given mass. The network, after construction, requires spatialization, which can be done with algorithms, like FA2(Jacomy *et al.*, 2014), Kamada-Kawai (Kamada & Kawai, 1989), Fruchterman-Reingold (Fruchterman & Reingold, 1991), and many others.

Another important feature is that the Pearson correlation function is not robust against outliers; therefore, logarithmic or square root transforms might be useful in the construction of correlation networks. However, if one is interested in the full structure of mass correlations, an alternative to the correlation networks could be used. For example, a hierarchical biclustering algorithm recovers groups of correlated entities as well as spatialization algorithms discussed above.

Figure 12 shows a correlation network calculated from 500 fs-LIMS composite mass spectra (100'000 single laser shot mass spectra). The data was acquired from the depth profile, where a microfossil was detected. As one can see, the structure of the correlations clearly identifies chemical entities – quartz mineral, polymetallic inclusion, and a microfossil. The groups in the network were recovered using Louvain clustering (see chapter Louvain modularity). The topology of the subgraphs in the

network identifies specific elements that are central to the given cluster. For example, the 12C is among the central nodes in the subgraph of microfossils. Whereas 28Si and first SiO-44 amu, node are centrally located in the quartz cluster.

Figure 12. An exemplary correlation network measured from the Gunflint chert. Nodes in the network encode single unit masses. Edges between masses indicate strong correlations (higher than 0.5). The colors of nodes and edges are defined according to the Louvain modularity of the network (see further in the text). The length of edges indicates weights (correlation factors) defined on the given edge and optimized using the graph spatialization algorithm.

In the lower part of the quartz mineral, the presence of the bio-relevant element (^{31}P) in the cluster of quartz-related masses could be explained by the intrusion of neighboring mass ^{30}Si into the integration window of ^{31}P. Therefore, some of the masses can be placed in various clusters in an artefactual manner. Overall, mass correlation networks provide an exploratory tool that allows understanding the multi-dimensional relationships between many single unit masses. The method provides visually recognizable and meaningful clusters and makes the mineralogical inference less bothersome. Moreover, networks look visually very appealing and can be used as an interpretable guide of the collected data. The dynamic networks can be calculated on large datasets, where one network will be merging into another one. The time series calculated on secondary metrics on such data can be indicative of inclusions. For example, significant carbon centralities are appearing only within microfossils, whereas diluted inorganic carbon present in the host mineral will not be showing such metrics. The structure of the correlation networks also provides an insight into molecular clustering patterns, for example, the Si oxides are connected to the Si isotopes.

2.3.5 ForceAtlas 2

An original contribution by (Jacomy *et al.*, 2014) introduced a continuous graph drawing algorithm called ForceAtlas2. As many other force-directed graph layout algorithms, ForceAtlas2 was designed to visualize complex directed and undirected networks with weighted and unweighted edges. In general, ForceAtlas2 could be compared with other continuous network layouts such as Yifan Hu (Hu, 2005), Kamada-Kawai (Kamada & Kawai, 1989), and Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991).

The ForceAtlas2 (FA2) recently has gained attention in single-cell transcriptomics as a preferred tool for high dimensional data visualizations. It was shown that FA2 embeddings could capture the structure of gene expression profiles on par with t-SNE and UMAP (Tusi *et al.*, 2018; Wagner *et al.*, 2018; Weinreb *et al.*, 2018). Arguably, the FA2 preserves more continuity of the manifolds and provides embeddings of the k-NN graphs in an easily understandable way. However, the FA2 was designed as a general-purpose graph spatialization technique and therefore was implemented in many commercial and open-source graph drawing platforms. Among available options, the python implementation of the FA2 provides a node embedding option that can be used in addition to the visualization engine implemented in Gephi (an open-source graph drawing software) (Bastian *et al.*, 2009). Nevertheless, the node embedding option makes the FA2 also a data dimensionality reduction method so that FA2 scores can be used for classification, clustering, and predictive analytics.

As in many other spatialization methods, the FA2 has attractive and repulsive forces that iteratively act on the network, optimizing the positioning of the nodes in a stepwise manner until global force-balance is achieved. The physical intuition behind FA2 could be represented as an equilibration of freely floating electrical charges that converge to the balanced state after a number of pairwise interactions. Therefore, the FA2 essentially is an n-body simulation algorithm with strictly defined forces.

The attractive force F_a, utilized in the FA2, presents the linear attraction that controlled by the distance d, defined between two nodes n_1 and n_2:

$$F_a(n_1, n_2) = d(n_1, n_2), \qquad (25)$$

The repulsive force is defined as:

$$F_r(n_1, n_2) = k_r \frac{(deg(n_1) + 1)(deg(n_2) + 1)}{d(n_1, n_2)} , \qquad (26)$$

Where deg is a degree of the node, which is the number of connections that it has to other nodes and k_r is a parameter that can be set by the user. The repulsive force is designed to have additional visibility of "leaves" – nodes with degree one that could clutter the structure of the network.

The overall visualization quality could be improved with a number of additional settings. For example, the LinLog mode provides the logarithmic attraction force, offers better placement of nodes, and better readability of the network's structure.

$$LogF_a(n_1, n_2) = \log(1 + d(n_1, n_2)), \qquad (27)$$

Another parameter that could be controlled is gravity. The gravity presents an additional force that can prevent weakly connected nodes from drifting too far away from the central part of the network. The gravity is defined as:

$$F_g(n) = k_g(deg\,(n) + 1), \qquad (28)$$

Where k_g is a weighting factor.

If the spatialized graph has weights defined on edges, which can be defined as any similarity measure between nodes, then the attraction force is computed in the following manner:

$$F_a = w(e)^\delta d(n_1, n_2), \qquad (29)$$

Where $w(e)$ is the weight of the edge e.

The remaining set of finer parameters, affecting the spatialization in a minor way, can be found in an original publication by (Jacomy *et al.,* 2014).

Among the advantages of FA2, one can note that it spatializes the graph in a real-time drawing, which allows interactive exploration of structures of the given network by rapidly testing different scales, forces, and constraints. Thus, learning by trial and error can provide good intuition for the appropriate settings choices.

Figure 13 illustrates the exemplary k-NN (k=5) network spatialized using the FA2 algorithm. The network represents the cosine similarity structure of the imaging dataset measured from the Gunflint chert. The 20'000 fs-LIMS mass spectra were preprocessed using the single mass unit spectral decomposition (260 single mass units were extracted from the mass spectra), log-normalization, and SVD. The structure of the k-NN network indicates the presence of microfossils and the quartz mineral. The grey color shades assigned to different parts of the network are calculated from the network topology using the Louvain modularity. As it was mentioned before, the FA2 provides a more continuous layout in contrast to other dimensionality reduction methods. Among limitations of the FA2 is that node embeddings can be done only in 2D space.

In general, the FA2 provides good spatialization quality that comes close to methods like UMAP and t-SNE. However, the continuous character of the graph drawing slows the layout computation time for large graphs because it requires graph rendering on every step of the optimization. Another drawback that is partially resolved is that FA2 does not provide a circular constraint on the attraction field. This leads to the excessive repulsion of the disjoint subgraphs and visualization scale issues. The gravity force, which provides an additional attraction to the center of the graph spatialization, partially solves that; however, it can be challenging to properly visualize highly disconnected subgraphs using FA2.

Overall, the FA2 showcases that high-quality graph drawing can be achieved using simple forces and relatively few constraints and settings. The strength of the method is provided by its interpretability and engaging visualizations that are competitive with all commonly-used graph drawing algorithms (Hu, 2005; Kamada & Kawai, 1989). The FA2 provides the continuous layout which is a pro and con at the same time. Due to the complexity of the large graphs the FA2 can converge longer in comparison to the discrete layouts, however, the real-time drawing provides better understanding of the node positioning. If the node gets trapped in the local neighborhood, from drawing it will be cleat where that node belongs.
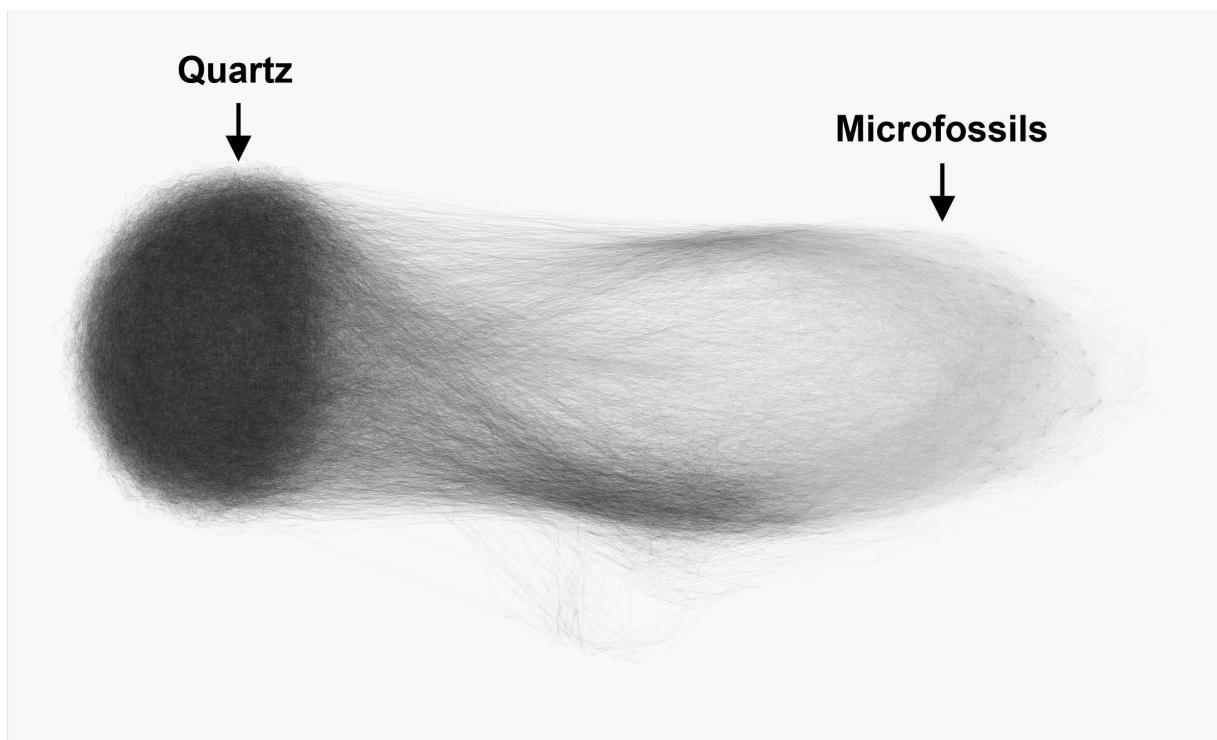
Figure 13. The FA2 visualization of the exemplary k-NN network, constructed on 20'000 fs-LIMS mass spectra measured from the Gunflint chert. The weights on edges are defined as cosine similarity on SVD precompressed mass spectra. The topology of the network illustrates smooth transition of the microfossil related spectra from the quartz, forming a loop-like topology.

The FA2 embedding of 20'000 LIMS observations is illustrated in Fig. 13. The structure of the tightly organized cluster on the left represents the observations from the quartz mineral. The nodes of the network are not visualized. The density of edges gradually changing to the right cluster represents the group of observations registered from the Gunflint microfossils. The edges between pairs of mass spectra are made translucent. As it could be seen from the figure, the similarity structure of the Gunflint microfossils clearly captures the transition of spectra from the host material (quartz). The density of edges indicates an overall number of connections between different pairs of mass spectrometric observations. On the lower part of the figure, a small number of mass spectra can be noted that represent the inclusions that are different from the main clusters. Altogether, the classification accuracy and quality of the retrieval of mass spectrometric classes is very high, and an embedding preserves a small distance between different pairs of observations.

In general, the quality of the FA2 embeddings is almost always very good, even in comparison to more complex methods, i.e., UMAP (see further in the text subsection UMAP). Although graph convergence times are different from those from UMAP, it takes significantly more time for large graphs to reach a force balance state than UMAP. However, the utility of the FA2 in analyzing complex LIMS data is high since it represents an easily interpretable graph spatialization method that can be used to visualize various k-NN graphs.

2.3.6 UMAP – Uniform Manifold Approximation and Projection

The UMAP (Uniform Manifold Approximation and Projection) is a novel manifold learning technique for dimensionality reduction developed by (McInnes *et al.*, 2018). The UMAP, in contrast, to the previously described matrix factorization methods, could be associated with the family of neighbor embedding methods. Such methods are specifically designed to visualize high-dimensional complex data using k-nearest neighborhood graphs (k-NNs) and typically use PCA as an initial data preprocessing step. The UMAP embeddings are constructed from a theoretical framework based on Riemannian geometry and algebraic topology (McInnes *et al.*, 2018). Although the mathematics behind the technique is complex, the UMAP represents a scalable and fast algorithm for generic data similarity search applicable to large and high-dimensional LIMS data.

The UMAP algorithm is competitive with t-distributed stochastic neighbors embedding (t-SNE) (Van der Maaten & Hinton, 2008), which was considered to be a standard data reduction method in biology and genetics for some years. The UMAP provides high visualization quality, increases interpretability, and arguably preserves more of the global structure (structure between clusters) with superior run time performance. The run time performance, recently, has been improved even more, with a CUDA accelerated GPU version of UMAP (Nolet *et al.*, 2020) (with up to 100x reported speedup). Furthermore, utilizing the strength of the Laplacian Eigenmaps (Belkin & Niyogi, 2003), UMAP has no restrictions on the number of output embedding dimensions, making it a convenient general-purpose dimension reduction technique for unsupervised and semi-supervised machine learning. The semi-supervised approach with UMAP can be utilized by re-embedding of the mass spectrometric observations with previously discovered mineralogical and chemical group assignments. In recent years, the UMAP algorithm became a popular tool in single-cell transcriptomics (mass cytometry and single-cell RNA sequencing) (Becht *et al.*, 2019), population genetics (Diaz-Papkovich *et al.*, 2020), and many other fields (Sainburg *et al.*, 2020).

It is argued that UMAP provides balanced embeddings, preserving more of the local and global structures, and provides an easier to interpret set of hyperparameters in comparison with other manifold learning techniques. For example, the t-SNE (Van der Maaten & Hinton, 2008), which is another dimensionality reduction algorithm, is an incredibly flexible method that can often find structure in the data, where other methods are failing. However, the same flexibility makes it harder to interpret (Wattenberg *et al.*, 2016) and find a balanced set of hyperparameters in unknown data.

The theoretical foundations for UMAP and associated with the algorithm mathematical proofs are published in the original publication (McInnes *et al.*, 2018). Here only the main equations will be provided to outline the mathematical logic and computational view of the algorithm. In practical terms, the UMAP represents the weighted graph algorithm that could be described in two main phases: 1) A weighted k-NN graph construction and 2) A low-dimensional graph layout optimization. In general, all neighborhood graph algorithms follow the same structure; the difference is in specific details on how the graph is constructed and how the layout is optimized.

The first part of the UMAP algorithm represents the construction of the weighted k-NN graph: Let $X = \{x_1, \dots, x_n\}$ be the input dataset with a metric d. Now, let us find

the k-nearest neighbors of x_i under the metric d. Given an input hyperparameter k (number of nearest neighbors) for each x_i, we can form a set of connections – $\{x_{i_1}, \dots, x_{i_j}\}$. In the original UMAP implementation, during the first stage of the graph construction, the search of nearest neighbors is accomplished using the approximate nearest neighbors descent algorithm (Dong *et al.*, 2011), which is a stochastic model that provides ~80%-100% accuracy rates for true neighbors recall. However, exact NN-search implementations exist. For example, the GPU accelerated UMAP provides a full neighbors search (Nolet *et al.*, 2020).

For each x_i we will define ρ_i and σ_i as:

$$\rho_i = \min\left\{d\left(x_i, x_{i_j}\right) \mid 1 \leq j \leq k, d(x_i, x_{i_j}) > 0\right\}, \qquad (30)$$

And σ_i results from:

$$\sum_{j=1}^{k} \exp\left(\frac{-\max\left(0, d\left(x_i, x_{i_j}\right) - \rho_i\right)}{\sigma_i}\right) = \log_2(k), \qquad (31)$$

The second stage of the graph construction, after finding closest neighbors, weights the network around each vertex in the k-nearest graph by using the adaptive exponential smoothing kernel (in a similar fashion to t-SNE). The kernel smooths the distances in each local neighborhood of the k-NN graph by finding the normalizer σ_i, such that equation (31) is satisfied and defining the Riemannian (differentiable) metric local to the point x_i. Accordingly, ρ_i will contain the distances to the closest non-zero neighbors and ensures that the vector x_i will be connected to at least one other data point with an edge weight 1. By doing that, the algorithm ensures that global connectivity will be preserved.

Now, we can define a weighted graph $G = (V, E, w)$, where G is the graph, V is the vertexes, and E is the set of directed edges. w denotes the weights on pairwise edges. The weight function can be defined as follows:

$$w\left(x_i, x_{i_j}\right) = \exp\left(\frac{-max\left(0, d\left(x_i, x_{i_j}\right) - \rho_i\right)}{\sigma_i}\right), \qquad (32)$$

Now, for a given set of points x_i, with a set of nearest neighbors x_{i_j}, we have a set of weight values w. In the output graphs, observations are connected in a probabilistic manner, meaning that the edge between two data points represents the likelihood (w) that two points are connected. Further, the graphs with the varying notion of distance (local metric is defined in terms of ρ_i and σ_i) are merged into a single graph using operators defined for fuzzy (i.e., probabilistic) sets (McInnes *et al.*, 2018). This step patches the incompatible metric spaces together to form a single unified graph.

Let A be the weighted adjacency matrix of G, and consider the symmetric matrix B:

$$B = (A + A') + (A \circ A'), \qquad (33)$$

Where A' denotes transpose of the adjacency matrix A, and ∘ is an element-wise (Hadamard) product. Thus, the values of A_{ij} represents the probabilities that directed edge from x_i to x_j exist, then B_{ij} is the probability that at least one of the two directed edges exists. The output UMAP graph G is then an undirected weighted graph, whose adjacency matrix is given by B.

The second part of the algorithm uses a force-direct layout algorithm to visualize the weighted graph G. Initial coordinates of the vertexes, prior to the optimization, are provided by Laplacian Eigenmaps (spectral decomposition of the k-neighborhood matrix), which ensures faster layout convergence and better preservation of the global structure(Kobak & Linderman, 2019). Therefore, it is important to note that the number of extracted components is a hyperparameter of Laplacian Eigenmaps. A force-directed layout algorithm utilizes a set of attractive and repulsive forces on the edges and nodes of the weighted UMAP graph G, up until the point when a force balance is achieved.

The UMAP exploits attractive force between vertices i and j at coordinates $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$, determined by:

$$F_{att} = \frac{-2ab||\boldsymbol{y_i} - \boldsymbol{y_j}||_2^{2(b-1)}}{1 + ||\boldsymbol{y_i} - \boldsymbol{y_j}||_2^2} w\left((x_i, x_j)\right)(\boldsymbol{y_i} - \boldsymbol{y_j}), \quad (34)$$

Where a and b are hyperparameters (in current UMAP implementation defaults are a=1.577 and b=0.895), the repulsive force is given by:

$$F_{rep} = \frac{2b}{\left(\epsilon + ||\boldsymbol{y_i} - \boldsymbol{y_j}||_2^2\right)\left(1 + a||\boldsymbol{y_i} - \boldsymbol{y_j}||_2^{2b}\right)}\left(1 - w\left((x_i, x_j)\right)\right)(\boldsymbol{y_i} - \boldsymbol{y_j}), (6)$$

Where ϵ is a heuristic number designed to prevent division by zero (and equals 0.001). The attractive and repulsive forces described above are derived from the gradients that optimize the cross-entropy function between weighted graph G and an equivalent graph H, constructed from low-dimensional points $\boldsymbol{y_i}$, where $i = \{1, ..., N\}$. The discussion on similarities of cross-entropy functions used by UMAP, t-SNE, and FA2 can be found in (Böhm *et al.*, 2020).

Consequently, the output graph H approximates the structure of the original data points $\boldsymbol{X}$ and graph G on a given metric space, and the difference is minimized as much as optimization allows that. In practical terms, optimization of the cross-entropy function leads to the preservation of the local neighbors. As it was shown in (Becht *et al.*, 2019), the UMAP embeddings can be characterized by a high correlation of low-dimensional and high-dimensional distances (averages to ~0.7 across different model datasets).

Figure 14 illustrates the exemplary graph H, calculated on 20'000 260-dimensional fs-LIMS mass spectra. The mass spectra are measured from the Precambrian Gunflint chert sample. The left side of the graph indicates the location of mass spectra registered from the microfossils, and the right side indicates the positioning of mass spectra recorded from the inorganic host mineral. Additionally, the presence of various small inclusions can be noted, which are aggregations of less connected parts of the graph in-between main bodies. Figure 14A illustrates the connectivity structure of the UMAP embedding with an additional bundling of the edges. The bundling algorithm assigns an additional force between neighboring edges and forms groups of nearby connections, which helps make the structure at a particular scale clearer (Bednar *et al.*, 2017). Figure 14B illustrates the connectivity structure of the data points using linear edges.

As it could be seen from the graphs provided in Fig. 14, the optimized location of the mass spectra in the low dimensional embedding is controlled by the forces applied to the pairwise edges. Thus, if an edge between two data points does not exist, there is no force defined in between, and co-occurrence in the close neighborhood of not connected data points can be artefactual.
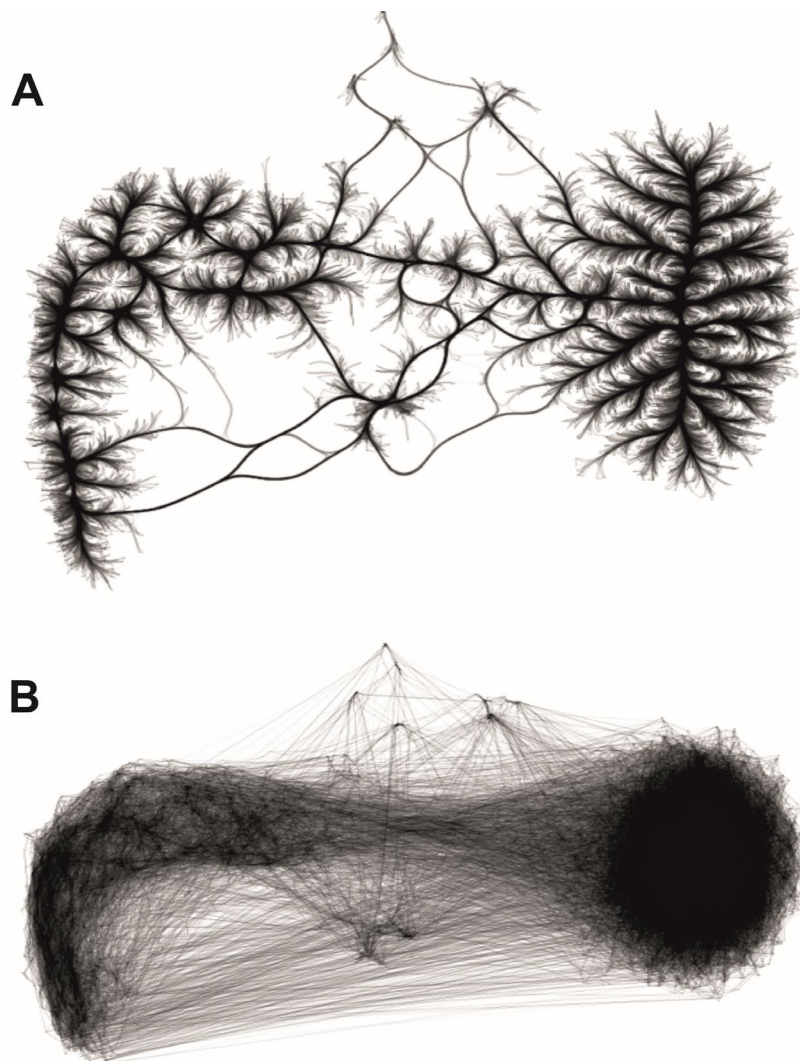
Figure 14. The exemplary UMAP embeddings of 20'000 fs-LIMS mass spectra (260-dimensional data) measured from the Gunflint chert. The edges between points represent the cosine similarity between individual mass spectra. Every mass spectrum is connected to k of its nearest neighbors. The structure of the connectivity between mass spectra identifies chemical entities. A – the UMAP embedding with pairwise edge bundling (with an additional attraction between edges). B – the UMAP embedding with linear connections between mass spectra.

While UMAP provides very helpful reduced data representations, interpretations of UMAP embedding results require some care.

It is worth noting that:

1. The output embeddings are hyperparameter-dependent. For example, the k in k-NN graphs could be used as an approximation of the global (larger structure between clusters) or local neighborhood (a structure within clusters). Typically, k=5 provides better preservation of the local neighborhood, and k=20 better approximates the global structure of the dataset.
2. The size of the clusters relative to each other can be meaningless.
3. The stochastic nature of the current UMAP implementation ensures that there will be differences between embeddings on different runs with the same

hyperparameter choices (a consequence of the stochastic gradient descent). Although, this problem was recently solved by implementing neural networks that learn the embedding space (Sainburg *et al.*, 2020).

For the review between mainstream algorithms such as *t* -distributed S tochastic Neighbors Embedding (t-SNE)(Van der Maaten, 2008), UMAP(McInnes *et al.*, 2018), ForceAtlas2 (Jacomy *et al.*, 2014), and Laplacian Eigenmaps (Belkin & Niyogi, 2003) one may consult the recent review by (Böhm *et al.*, 2020). It also provides an interesting perspective over potential equivalence between various nonlinear dimensionality reduction methods as a continuous attraction-repulsion spectrum, where methods with stronger attraction could be used to characterize continuous manifolds (i.e., Laplacian eigenmaps and ForceAtlas2) and stronger repulsion is more appropriate for characterization of manifolds with discrete, cluster-like strutures (i.e., UMAP and t-SNE).

Overall, UMAP is a powerful method that provides fast and meaningful embeddings of large datasets. Regarding the embeddings of fs-LIMS data, the method is capable of unsupervised retrieval of mineralogical and molecular groups present in the dataset. However, it is important to keep in mind inherent tradeoffs present within the algorithm for efficient interpretation of the data reduction results.

To summarize, UMAP performs the construction of the neighborhood graph and provides a subsequent embedding that preserves the structure of distances in low dimensions. The UMAP learns an embedding by minimizing the cross-entropy function that iterates over high-dimensional and low-dimensional distances. The positively weighted edges are attracted to each other, and randomly sampled points are repulsed by using the negative sampling. Thus, allowing minimization of the cross-entropy function through stochastic gradient descent. Altogether, this approach allows to recover similar data samples to be located in close proximity to each other, while dissimilar spectra are repulsed, thus being located farther in the embedding space.

2.3.7 The Mapper algorithm

Recovering the unknown structure from complex and noisy nonlinear data is among the fundamental problems in machine learning and statistical inference. The Mapper algorithm, originally presented by (Singh *et al.*, 2007) is a topological method developed for the analysis of high-dimensional datasets, recognition of 3D shapes, and analysis of large datasets using reduced topological representations that capture essential data patterns. The method transforms discrete point cloud data into simplicial complexes that contain far fewer points compared to original data. However, the structure of the output network, build from simplicial complexes, preserves the information about geometric coordination and closeness at the specified resolution. In topological terms, the Mapper algorithm represents an approximation of the Reeb graph (mathematical object associated with shapes), defined on a real-valued manifold (Biasotti *et al.*, 2008; Mohamed & Hamza, 2012). The Mapper algorithm is a fast and flexible data-skeletonization method that is of high utility in mass spectrometry. The fs-LIMS, being a fast analytical method (it takes ~20 µs per recorded mass spectrum), can produce large datasets in relatively short periods that can contain some level of noise and high levels of nonlinearity that can obscure the identification of minerals and chemical entities.

For a description of the algorithm, I'll refer to the original contribution by (Singh *et al.*, 2007). The illustrative explanation of the algorithm is provided in Fig. 15, where Mapper is applied to the fs-LIMS observations using two covering functions. The figure illustrates the distribution of 4400 ^{28}Si and ^{29}Si isotope intensity measurements registered from the quartz mineral, where no interferences with other minerals are detected. In an ideal situation, the correlation of two isotopes will be exactly unity; however, we see that the scatter of ^{28}Si and ^{29}Si adds some uncertainty to the ratio of isotopes. This uncertainty is increased by additional isobaric input from CO molecule contributing to the mass intensity of ^{28}Si. However, the essential data pattern that we see is that isotopes are linearly correlated, and this pattern needs to be preserved.

The data shown in Fig.15 represent the single mineral measurement obtained from the single location depth profile. However, the spatially resolved measurements such as mass spectrometric imaging or just a set of measurements from different minerals in heterogeneous media will form more complex shapes that often can have highly nonlinear structures. When complex data structures are present, Mapper networks can capture these structures in a simplified manner, making the interpretation of large datasets more intuitive and convenient.

To represent the Si isotope intensity distributions using Mapper, we have to design the filter functions that will overlap and cover the full space of isotope intensity variance. In the exemplary case, we can use the ^{28}Si and ^{29}Si intensity space as filter functions. The observations that appear in joint filter function windows (#1 and #2) will be aggregated into nodes, where an additional clustering of original observations will be applied. This is required to identify nonlinear and complex data shapes (i.e., clusters, flares etc.). The observations that appear in both overlapping filter function windows will be connected with edges, thus, preserving the locality and closeness notion present in the data. As simple it might seem, such an approach provides powerful data visualization capabilities. The mapper networks can be constructed with a given resolution (that determined with the size of overlapping filter functions), which in turn provides clear insights about structural inhomogeneity present in the datasets.

The clustering of data in the overlapping filter windows is designed to separate disjoint data points present within the given filter window. The clustering is not constrained to the specific algorithm. Any type of clustering can be applied, thus, changing the output topology of the network. Node colors and node sizes can also be used to encode additional information.

Typically, the Mapper networks are used to coarsen embedding results (i.e., UMAP embedding scores); thus, they require dimensionality reduction as a step of data preprocessing. However, the filter functions can represent any type of data transformation. Therefore, joint filter functions can be of high importance as well. For example, KDE (see further section "Kernel density estimates") values defined on the UMAP embeddings could be used as a coloring scale, providing crucial density information.

As it might seem odd and unintuitive, the Mapper coverings can be defined on high-dimensional sets as well. For example, Fig. 16 illustrates the Mapper network calculated on 4 UMAP dimensions. The networks represent a low dimensional structure of 20'000 mass spectra (260 original dimensions) registered from the Gunflint chert using the HR-LIMS-GT instrument. The network was constructed using the 20 filter functions overlapped by 10 percent. The spatialization of the network was achieved using the FA2 algorithm. It can be noted that the connectivity structure of the quartz group is relatively sparse in comparison to the connectivity of the "arm" of microfossils. This can be attributed to the inhomogeneity of the similarity structure present in the quartz group. The physical explanation for that could be found in the original mass spectra, where we can see that spectra from quartz are less intense and have fewer mass lines; thus, the variance of noise can contribute to the similarity structure in the cluster of spectra registered from quartz. It is important to mention that visualization of 4D spaces using conventional methods is impossible. Therefore, these granular similarity structures present in the UMAP embedding can be missed using standard data visualization techniques.

In a classical data visualization path, i.e., using scatter plots, description of the same amount of information would require assessment of 4 separate scatter plots. And what is interesting, higher dimensional covers can better preserve the distances among the neighbors due to the locality notion introduced by the size of the filter functions. The Mapper network does not distort distances present in the data. Instead, it aggregates data points into larger bins, allowing to compound high-dimensional spaces into graphs with complex topologies.

Overall, the technique provides a framework for constructing useful combinatorial representations of high-dimensional point-cloud data. It reasonably preserves the locality of the data and provides reduced networks that are of high utility in visualizing complex datasets. It allows an inference of chemical heterogeneity from the fs-LIMS observations and can be used in combination with the Louvain method in the creation of labels for identified compounds (i.e., minerals). In general, the Mapper algorithm being a topological method, provides mapping of the data-points in the close neighborhood (determined by the size of the filter windows); however, the structure of the large distances can be significantly distorted since they are not captured, and considered to be irrelevant (though it is possible to create Mapper networks that can capture large distances as well). The noise tolerance of the method comes from the size of the overlapping filter windows, thus, providing additional connectivity information on point-cloud data. For example, with Mapper networks, it is possible to cluster very sparse data that is hard to cluster with other methods with the same accuracy.
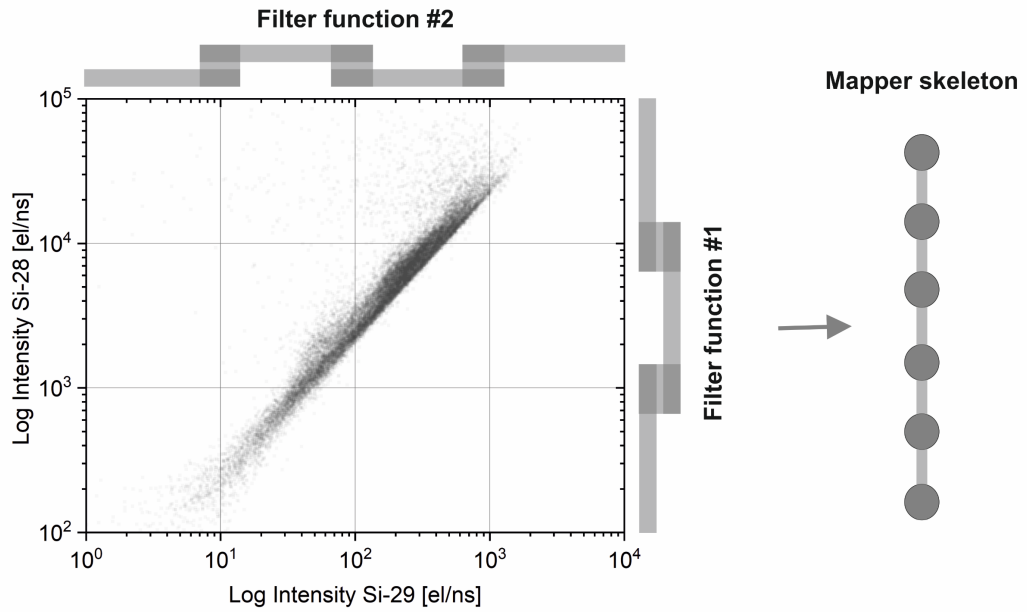
Figure 15. An illustration of the Mapper data skeletonization approach. Left – linearly distributed isotope intensities registered from the quartz (4400 measurements). Right – simplified data structure captured from Si isotope measurements.
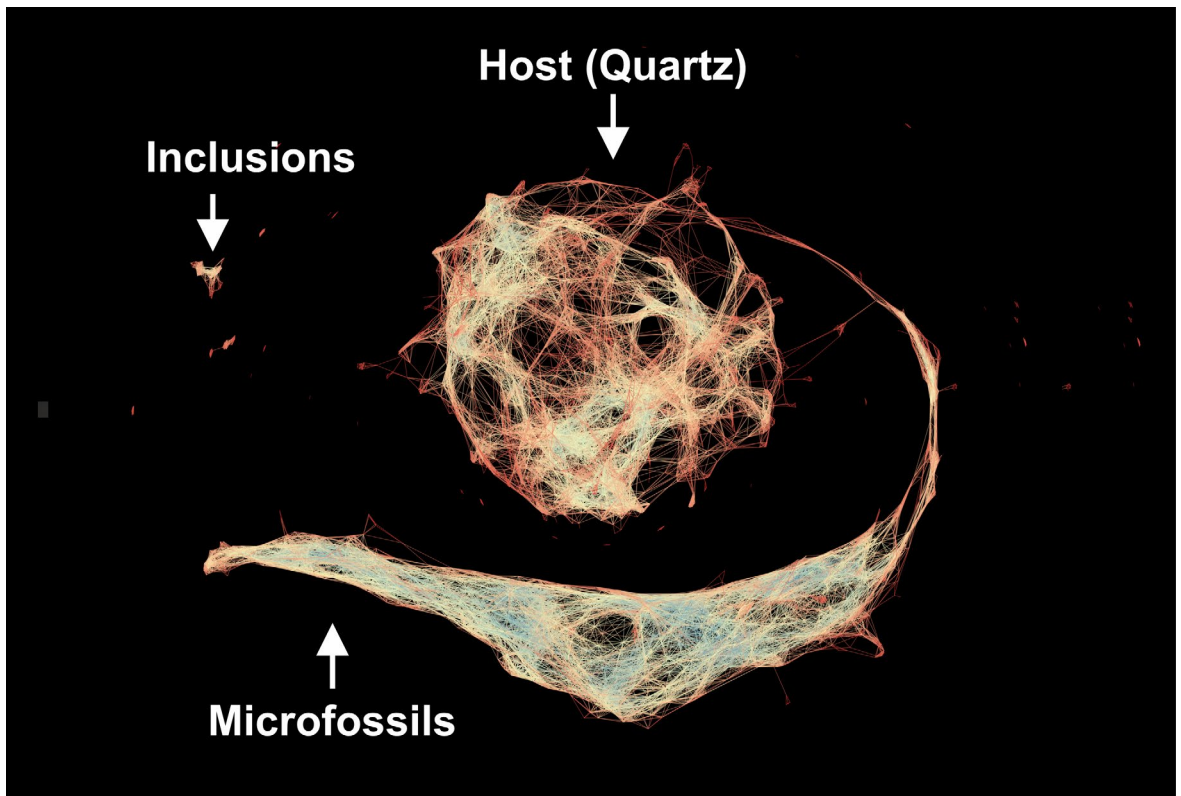

Figure 16. An exemplary Mapper network calculated on four UMAP dimensions. The structure of the network encodes the similarities of 20'000 mass spectra (defined as cosine similarity) and indicates the presence of microfossils, quartz, and a small set of inclusions. The network is colored according to the degree of nodes (red color indicates small degree; blue color indicates high degree (range of node degrees spans from 1 to 60)), illustrating the density of connections in the network.

2.4 Binary and multi-class classification models

Identification of spectra that are registered from a given compound is a statistical problem widely addressed in machine learning (ML) and chemometrics (Manrique-Martinez *et al.*, 2020). In specific areas where standardized spectrometric libraries exist, rapid developments were observed in the recent decade. For example, mass spectrometry based methods were developed for rapid identification and classification of bacteria and other microorganisms (Sauer *et al.*, 2008; Sauer & Kliem, 2010). With the development of deep learning models, significant progress was seen in the classification of tumors using imaging mass spectrometry data (Behrmann *et al.*, 2018). And recently, deep neural networks were used to predict 2,497 metabolite classes using tandem mass spectrometry data. (Dührkop *et al.*, 2021). Thus, in the field of supervised machine learning, necessary tools for the classification of highly complex mass spectrometric datasets already exist, and they are successfully applied across the board of different scientific disciplines. However, unsupervised learning places a harder problem as no labels (i.e., names of compounds) are known a priori; thus, every given observation from unknown samples contains some uncertainty and requires intense attention to interpret and assign labels correctly.

Here, combining all previously presented methods, it will be demonstrated that by using unsupervised data reduction and characterization techniques, it is possible to recover groups of spectra that share a significant level of similarity. Thus, a spectral signature of the given compound can be retrieved and assigned with an appropriate label.

Data colletion
Surface imaging
Volumetric imaging
Depth profiling

Spectral preprocessing
Binary conversion
Noise suppression
Baseline correction
Jitter correction

Feature extraction
Mass calibration
Single mass unit decomposition
Log transformation

Skeletonization and clustering
Mapper
Louvain
HDBSCAN

Dimensionality reduction
Linear (SVD or PCA) reduction
UMAP reduction

Interpretation
Weighted mass correlation networks
Mass spectrum analysis
Spatial visualization on maps
Creation of mineral labels

Machine learning
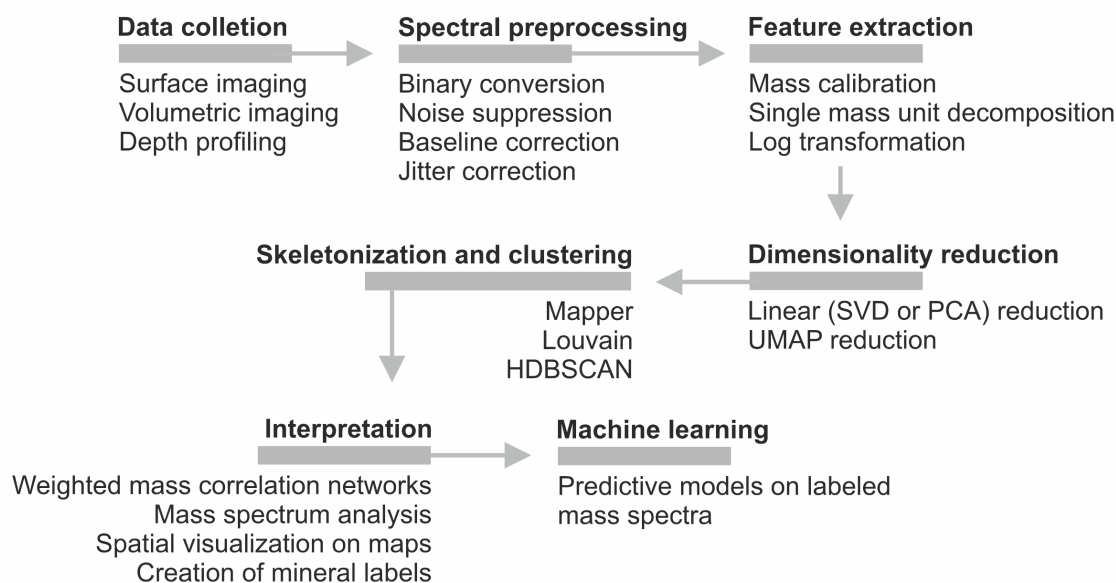Predictive models on labeled mass spectra

Figure 17. Full data processing pipeline for the construction of predictive models on given spectral sets. The diagram identifies seven steps required to build a predictive binary or multi-class classification model.

The data processing pipeline for unsupervised classification and identification of minerals and chemical entities using the fs-LIMS data is shown in Fig. 17. The diagram outlines seven analytical steps that are required to retrieve the structure of mineralogical and chemical diversity present in the acquired dataset. First, as shown in the diagram in fig. 17, the fs-LIMS being fast analytical method can

collect large-scale mass spectrometric information on any type of solids; thus, the method is not constrained by the sample types. Second, given the accurate spectral preprocessing steps are undertaken, the full spectral range can be decomposed into single unit mass intensities, and by that, finalizing the third step of the data reduction. Notwithstanding the fact that spectral intensity data in an original high-dimensional form already encode the necessary information about minerals and compounds of various origin, analysis of such data is not convenient and requires more time in comparison to the analysis of reduced representations. Therefore, the spectral sets can be further compressed using linear methods such as PCA and SVD. Typically, the level of PCA compression depends on the scientific task; however, 100 first PCs (or 75% of variance) are usually enough to represent the majority of the significant variance.

Further, linearly reduced mass spectra can be analyzed using manifold learning methods (i.e., UMAP, t-SNE, LE, FA2), which can capture essential data patterns based on their similarity. And the last fifth step of data transformations provides further coarsening and reduction of data complexity to enhance the separability of compounds using the Mapper algorithm. If Mapper networks reveal compelling structures, one can use Louvain clustering to statistically infer groups of observations that represent observations with specific chemistry.

Once numeric labels are acquired, an independent analysis of inclusions and main minerals is required (denoted as interpretation step in Fig. 17), using the raw spectrum analysis, spatialization of inclusions on chemical maps, and correlation networks. If from the given set of observations, it is clear that the chemical compound corresponds to a mineral and it could be identified with high fidelity from the bulk of measurements, the set of mass spectra could be assigned with a new label that will correspond to the mineral or chemical entity.

Thus, the inferred mineral name represents only a starting point in the last, seventh stage - construction of predictive models. Note that modern ML models can be directly deployed and run on exploratory rover missions and provide a real-time identification of the investigated materials. Therefore, such predictive models can be of high interest to future astrobiological missions and as well as to the field of Precambrian micropaleontology, where many examples of early life are problematic due to the chemical and morphological convergence of biosignatures.

Although machine learning is a complex field full of tailored data-specific models, modern community-supported libraries (such as scikit-learn) are adapted to the point where state-of-the-art ML architectures are directly accessible, providing broad access to the users from various fields. Therefore, latest models with high performance can be scored without too much problems. Although, binary classification of the bio-organic and inorganic chemistries was a primary target in this thesis, a multi-class classification of the spectral datasets represents a very similar problem, which can be addressed using the same classification tools. As it was mentioned previously, inference of the number of spectral types is a key moment that was addressed using graph-theoretic approaches. By measuring a series of attributes from the given compound, it was demonstrated in previous chapters that it is possible to characterize diversity of compounds with high fidelity. However, mineralogical inferences, for example, by using PCA, can be probabilistic due to the low separability of classes in linearly reduced representations. An illustration of the inferred cluster borderlines can be found in Figs. 18, 19 and 20. Various ML classification methods reveal different types of borderlines; however, all of them depend on the initial separability of classes.
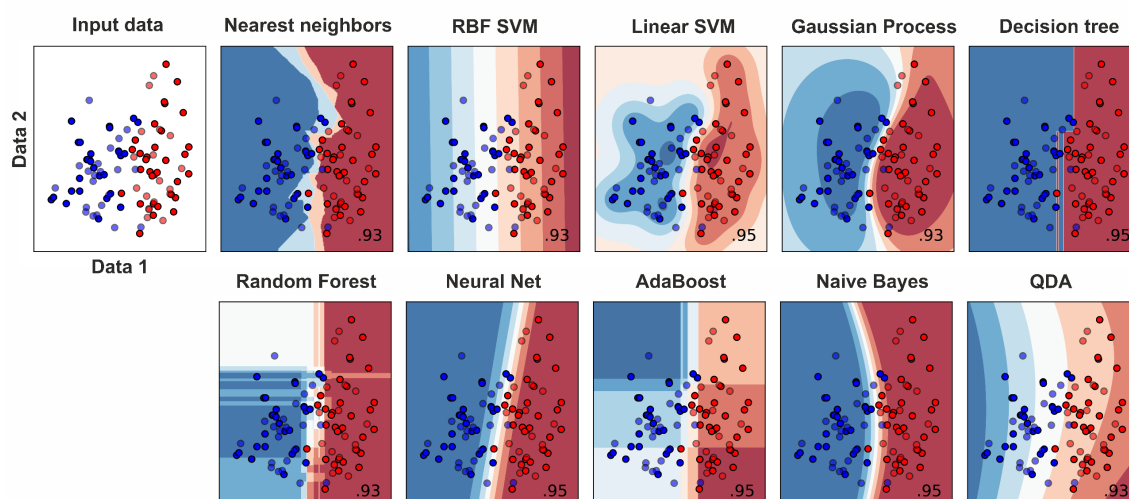
Figure 18. An illustration of the learned decision boundaries for various binary classification ML models on synthetic data. The classification accuracy is denoted in the lower right part of the figures. Color gradient communicates the uncertainty of the class assignment. Courtesy of the scikit-learn.
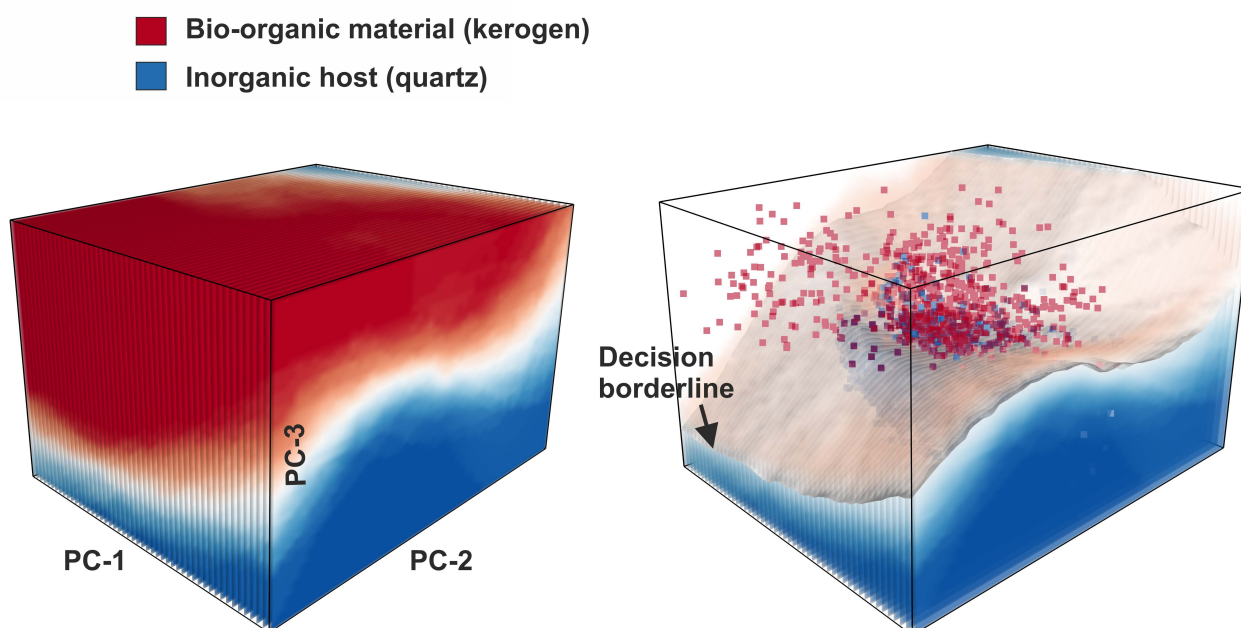

Figure 19. An illustration of the three-dimensional decision boundary identified for Precambrian Gunflint microfossils on PCA reduced fs-LIMS data. Red data points indicate the localization of spectra registered from microfossils. Blue data points (below the surface) indicate the positioning of spectra registered from Quartz mineral. The highly nonlinear character of the separation boundary can be noted. The color gradient communicates the uncertainty of the class assignment. Localization of the boundary between classes can be noted on the picture (right panel).
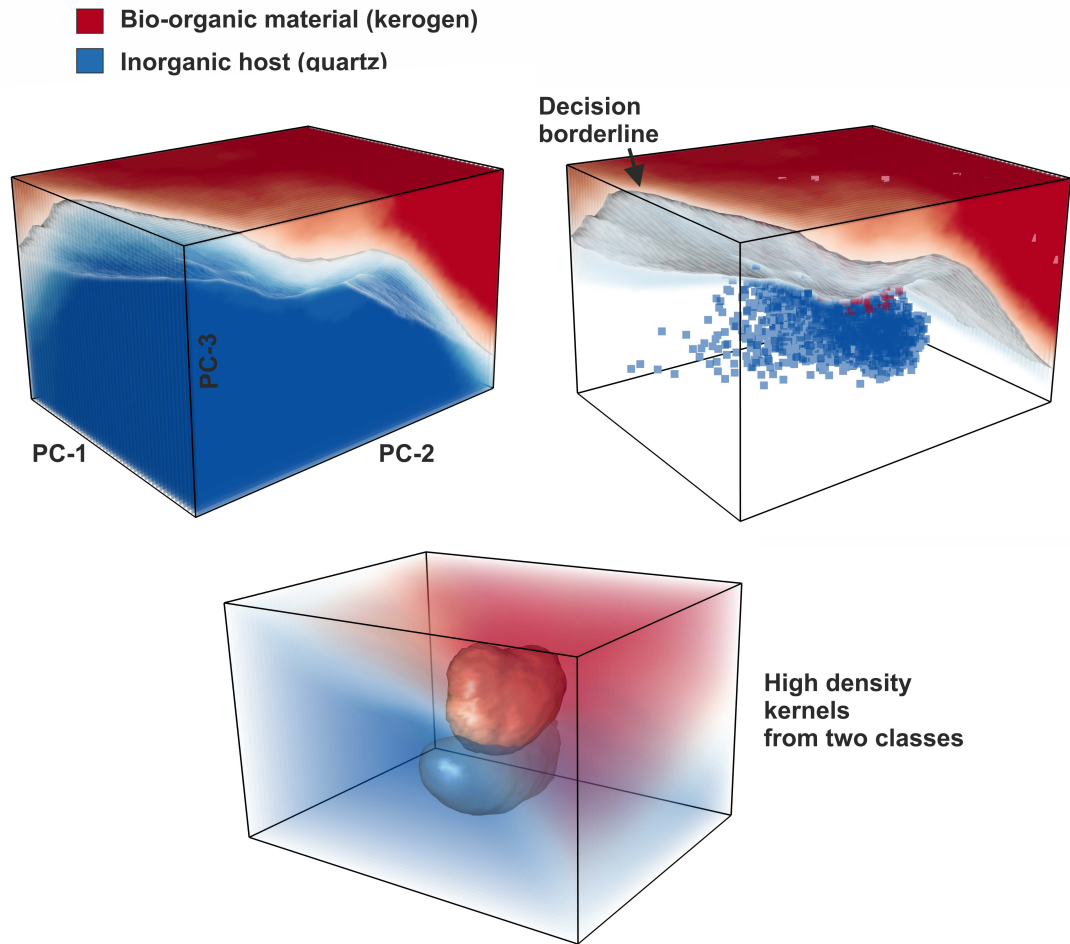
Figure 20. An illustration of the three-dimensional decision boundary identified for Pre-cambrian Gunflint microfossils on PCA reduced fs-LIMS data. Blue data points indicate the positioning of spectra registered from Quartz mineral. Red data points indicate the localization of spectra registered from microfossils (below the surface). The highly non-linear character of the separation boundary can be noted. The color gradient communi-cates the uncertainty of the class assignment. High density kernels can be seen on the bottom panel.

Current open-source ML libraries such as scikit-learn (Pedregosa *et al.*, 2011) or ML functionality provided with Matlab provide access to state-of-the-art models, and the accuracy of trained classifiers can be compared across the board of all available models. Such an approach rapidly increases the overall probability of training a good-performing ML classifier.

Figure 18 presents the illustration of the learned decision boundaries on a two-dimensional toy dataset, visualized using ten different classifiers. Though, in this thesis, more than 25 models were tested. Among the best-performing ones, repeatedly was identified the ensemble method - Adaptive Boosting (Ada Boost, lower-middle classifier in Fig. 18), which yielded on average a 99.7% separation rate between classes of the Gunflint dataset – bio-organic material (microfossils) and inorganic host material (silicified host - quartz), (see further contribution "Towards empirical biosignatures using LIMS" for detailed discussion of this figures). Interestingly, within binary classification models, the transition line between classes represents lower boundary of the high-dimensional intensity space for microfossils (see fig.18 and 20). Thus, the spectra that are not seen before (i.e., microfossils with unseen ionization profiles) can be successfully identified due to the smooth transition borderline between classes and complete coverage of the intensity space.

Figures 19 and 20 illustrate the three-dimensional borderline surface between classes - life (microfossils, red points) and inorganic host mineral (quartz, blue points), calculated on the first three PCA dimensions of the reduced data. Even though PCA performs better on highly linear datasets, we can see that interpolated borderline nicely approximates true separation surface in 3-dimensional PCA space. Despite the high accuracy of the tested models, deterministic assignment of the classes is still possible only to the high-intensity spectra (using linear data compression methods - PCA or SVD). However, the quality of the LIMS data allows to identify microfossils deterministically using nonlinear data reduction methods; thus, the labels presented in Figs. 19 and 20 (i.e., red and blue colors assigned to data points) can be further improved. Although recovery of the groups of spectra that share a significant level of similarity is possible using the nonlinear methods, some level of the class assignment uncertainty can be present as well (see the contribution "LIMS 3D imaging and manifold learning" for detailed information on nonlinear data transformation and spectral classification). However, the level of uncertainty is significantly smaller in comparison to linear data reduction methods.

The fs-LIMS being a precision analytics method (diameter of the analytical spot is ~5 μm for UV-258 nm laser and ~10 nm-level characterization on depth scale), can register fine chemical heterogeneity. Nonetheless, microfossils can be very finely distributed in the bulk of the sample. While it is not the case for surface imaging, in depth-profiling mode, it can be data-expensive to collect large-scale mass spectrometric observations from microfossils. On average, only 1 in 10 collected spectra can be attributed to the organic materials. Thus, the unequal sampling issue can be at play in analyzing such finely distributed entities. The kernel density estimates (KDE's, see section - kernel density estimates in the methods chapter) can be of high utility in drawing new synthetic samples from approximated distributions. While the biases present in the KDE will persist in the output synthetic data, there are additional approaches that can be used to compliment sparsely sampled entities.

In recent years, new approaches for data generation are gaining attention. For example, (Lindenbaum *et al.*, 2018) proposed a method for drawing new samples from the latent space of low dimensional embeddings. The approach proposes learning the geometry of the approximated manifold and generating points along the embedding structure. Need to mention that with the recent implementation of the Parametric UMAP (Sainburg *et al.*, 2020), the stochastic gradient descent was replaced with parametric optimization over neural network weights. Thus, allowing to learn stable UMAP embeddings and consequently generate new high-dimensional data points along any part of the approximated manifold. Although it is not fully explored in the thesis, the potential to generate new data points for sparsely sampled minerals (i.e., inclusions or rare fossils, or just fossils in general) is of high interest in fs-LIMS analysis, because it can help to generate statistics on sparse observations.

Another direction that caused significant attention in this field is the development of generative adversarial networks (Goodfellow *et al.*, 2014) and adversarial autoencoders (Makhzani *et al.*, 2015) that were shown to successfully generate new meaningful observations (pictures) from few original observations. Thus, the future direction of the fs-LIMS classification of minerals and compounds could be based on the acquisition of relatively few original observations followed by generative augmentation and training of the classifiers on augmented datasets.

2.5 Kernel density estimation

In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. In experimental sciences, KDE reveals the fundamental property of data – it's density function f, measured with some accuracy δ. Regarding the LIMS analysis, KDE plots are of high utility in visualizing the 2 and 3-dimensional densities and can be used to estimate the probability distribution function of any given mineral or chemical entity in narrowly defined intensity regions. Further in this thesis, I will show that KDE estimates of low dimensional embeddings identify minerals and compounds of specific chemistry in an interpretable and concise way. Moreover, since kernel densities approximate original density distribution f, it is possible to use KDE models for drawing new samples, which can be of high utility in machine learning for balancing the datasets with unequal sampling.

Let a random variable $X_1, \ldots, X_n$ be samples drawn from some univariate distribution with an unknown density f. To estimate the shape of the density function f, we can use the kernel density estimator, given by:

$$\breve{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(x - x_i\right) = \frac{1}{nh}\sum_{i=1}^{n} K\frac{(x - x_i)}{h}, \qquad (36)$$

Where K is a kernel function, and h is a smoothing parameter called bandwidth. There are many different kernel functions available and frequently used. However, only Gaussian kernels were utilized in the thesis.

The Gaussian kernel is given by:

$$K_u = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}, \qquad (37)$$

The construction of the kernel density estimates of univariate distributions requires selection of the kernel's bandwidth, which can be found using the rule-of-thumb, given by:

$$h = \frac{(4\breve{\sigma}^5)^{\frac{1}{5}}}{3n} \approx 1.06\breve{\sigma}n^{-\frac{1}{5}}, \qquad (38)$$

where $\breve{\sigma}$ is the standard deviation and n is the sample size.

Overall, the appropriate KDE bandwidth selection will result in a better approximation of the original density function. In contrast, erroneously chosen h could result in the formation of artefactual islands and cluster-like structures that are derived from the noise rather than the intrinsic density profile of the original variable. The KDE will also be of high importance in the clustering of observations, for example, using the HDBSCAN and DBSCAN algorithms (McInnes et al., 2017) as well as in the topological model constructions using Mapper (Carlsson, 2009; Singh et al., 2007). The 3-dimensional KDE estimates of the UMAP scores can be found in the contribution "On topological analysis of the fs-LIMS data". The 2D KDE's of single isotope intensity values are provided in "Chemical identification of microfossils from the 1.88 Ga Gunflint chert. Towards empirical biosignatures using LIMS" as well as in the last Gunflint contribution - "High-Resolution fs-LIMS 3D imaging and manifold learning reveal insight into chemical diversity of the Gunflint chert (1.88 Ga)". Thus, the KDE represent a density estimation tool that of high utility in fs-LIMS data analysis.
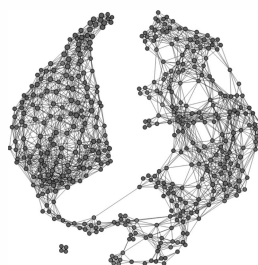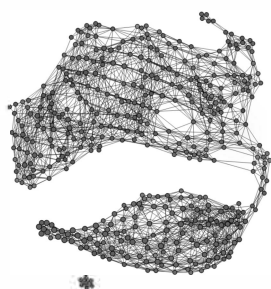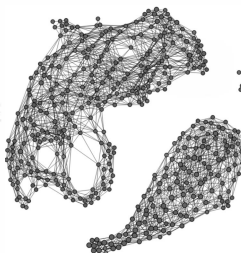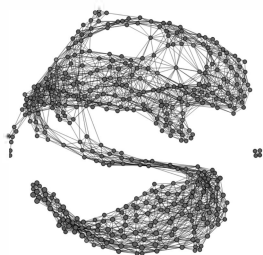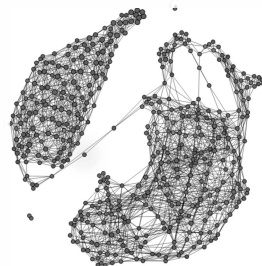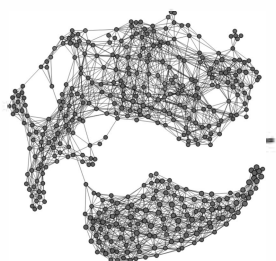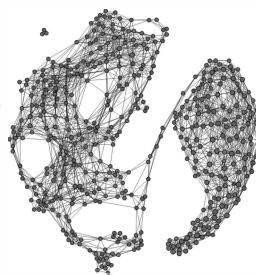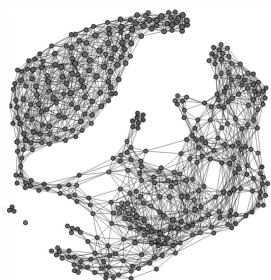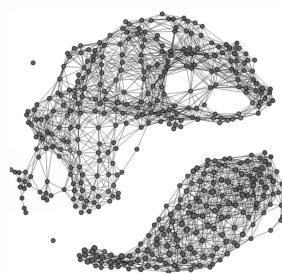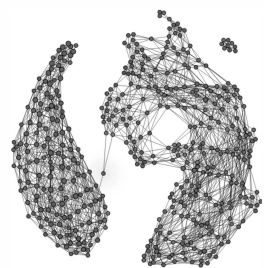
Results

Precis

Here, I will present the scientific results acquired during course of the PhD studies. The results chapter will present seven published (and submitted) contributions in peer-reviewed journals. Four contributions cover the work on Gunflint microfossils. Other contributions address specific aspects of ion-generation and accurate measurement of micro-inclusions with complex mineralogy. Lastly, I present the identification and characterization of amino acids with very low detection limits using the LDMS setup. Note, that figure numerations will be kept as in the publications.

3.1 Spatially localized analysis and depth profiling

3.1.1 Multiwavelength ablation/ionization and mass spectrometric analysis of 1.88 Ga Gunflint chert - (Astrobiology)

R. A. Lukmanov[1], M. Tulej[1], R. Wiesendanger[1], A. Riedo[1], V. Grimaudo[1], N.F.W. Ligterink[1], C. de Koning[1], A. Neubeck[2], D. Wacey[3] and P. Wurz[1]

[1]Space Research & Planetary Sciences (WP), Physics Institute, University of Bern, Bern, Switzerland;

[2]Department of Earth Sciences, Uppsala University, Uppsala, Sweden;

[3]Centre for Microscopy, Characterization & Analysis, The University of Western Australia, Perth, Australia;

Abstract

The investigation of chemical composition on planetary bodies without significant sample processing is of importance for nearly every mission aimed at robotic exploration. Moreover, it is a necessary tool to achieve the longstanding goal of finding evidence of life outside the Earth, for example, possibly preserved microbial remains within Martian sediments. The Laser Ablation Ionization Mass Spectrometer (LIMS) is a compact time-of-flight mass spectrometer intended to investigate the elemental, isotopic, and molecular composition of a wide range of solid samples including, e.g., low bulk density organic remains in microfossils. Here, we present an overview of the instrument and collected chemical spectrometric data at the micrometer level from a Precambrian chert sample (1.88 Ga Gunflint Formation, Ontario, Canada), which is considered to be a Martian analog. Data were collected from two distinct zones – a silicified host area and a carbon-bearing microfossil assemblage zone. We performed these measurements using an ultrafast pulsed laser system (pulse width of ~180 femtoseconds) with multiple wavelengths (IR-775 nm, UV-387 nm, UV-258 nm) and using a pulsed high voltage on the mass spectrometer to reveal small organic signals. We investigated: a) the chemical composition of the sample, and b) the different laser wavelengths' performance to provide chemical depth profiles in silicified media. Our key findings are 1) Microfossils from the Gunflint chert reveal a distinct chemical composition compared to the host mineralogy and we report the identification of 24 elements in the microfossils. 2) Detection of the pristine composition of microfossils and co-occurring fine chemistry (rare-earth elements) requires utilization of the depth profiling measurement protocol. 3) Our results show that for analysis of heterogeneous material from siliciclastic deposits, siliceous sinters and cherts, the most suitable wavelength for laser ablation/Ionization is UV-258 nm.

Introduction

In situ methods of chemical analysis on the surfaces of planetary bodies is of primary importance in current space research and promises to significantly increase the scientific return from these missions (Horneck *et al.*, 2016; Knight *et al.*, 2000; Wurz *et al.*, 2012). From the current perspective, Mars is a cold desert with an intense flux of ionizing radiation at the surface (Hassler *et al.*, 2014), high UV flux, chemical reactivity of soils (Carrier & Kounaves, 2015), and considered to be largely uninhabitable (Fairén *et al.*, 2010). However, scientific analysis and modeling suggest that life could survive in the subsurface and imply the possible existence of rock-hosted life in the deep subsurface (Stamenković *et al.*, 2019), where liquid water may be present (Orosei *et al.*, 2018). Considering impact cratering and early faulting, some of these formations may be accessible from the surface (Onstott *et al.*, 2019). Additionally, multiple lines of evidence exist for abundant liquid water on the surface, during the early stages of evolution of the planet, particularly in the Noachian (Ehlmann *et al.*, 2011). These lines of evidence include, for example, fluvial landforms, paleo-lakes, river deltas, and mineralogy indicative of water alteration and weathering (Arvidson *et al.*, 2014; Grotzinger *et al.*, 2014). Evidence of clement conditions on early Mars raises questions about the extent of the planet's possible habitability, for example, the scale and duration of this habitability. To answer such questions an in-situ chemical and mineralogical investigation of rocks from the Martian surface is required, using set of complementary instruments capable of recording high-quality chemical information. However, techniques that are traditionally used in space research, e.g., Pyrolysis–gas chromatography-mass spectrometry (Pyr-GCMS) or remote methods, e.g., Laser-induced breakdown spectroscopy (LIBS), might not be sufficiently sensitive to detect faint signatures of life from micrometer-sized and below organic remains (Navarro-González *et al.*, 2006).

Laser Ablation/Ionization Mass Spectrometry (LIMS) is a promising analytical technique (Azov *et al.*, 2020) capable of providing molecular (Ligterink *et al.*, 2020), elemental (Riedo *et al.*, 2020; Tulej *et al.*, 2015), and isotope (Riedo *et al.*, 2013; Tulej *et al.*, 2020) characterization of solid materials. LIMS provides fast measurements (20 μs per spectrum for a single laser shot) (Riedo *et al.*, 2019), spatially resolved (~10 μm) analysis with high detection sensitivity (ppm level), and excellent depth resolution (nm scale) (Grimaudo *et al.*, 2020). The laser ablation/ionization reflectron-type time-of-flight mass spectrometer used in this study is an instrument developed in the Space Research and Planetary Sciences division at the University of Bern and represents a real-sized prototype of a space instrument. The LIMS instrument incorporates a femtosecond laser ionization source, a miniature time-of-flight mass spectrometer (Ø 60 mm × 160 mm), making it suitable for space exploration, and an integrated microscopy system. However, the applicability of the LIMS systems to identify chemical signatures of life from microfossils that are billions of years old haven't been shown so far.

Among others, Precambrian cherts from banded iron formations were suggested as an astrobiologically relevant Mars analog site (Allen *et al.*, 2001) since iron oxides and silicates might precipitate as an aqueous mineral phase from a Fe and Si saturated water column, efficiently trapping any microbiota present. Moreover, terrestrial cherts offer an exceptional preservation level due to their capacity to encapsulate and preserve organic material (Alleon *et al.*, 2016). Gunflint microstromatolites preserved within the banded iron formation (outcrop on the North Shore of Lake Superior, Ontario, Canada) are well studied and considered among the best-known

Precambrian Fossil Lagerstätten. They represent a diverse microbial community that thrived in a shallow marine environment ~1.9 billion years ago (Barghoorn & Tyler, 1965). Moreover, Precambrian chert containing bona fide microfossils from the Gunflint formation represents an excellent and realistic testbed for in-situ space instrumentation.

To clarify the current state of the LIMS performance on realistic samples, we test the LIMS ability to identify ~1.9 Ga microfossils and potential chemical biosignatures in the Gunflint chert as a function of different fs-ablation wavelengths. The chemical identification and characterization of microfossils from cherts using space instrumentation is challenging due to the low bulk densities and small size of the microfossils, which are usually not larger than the first tens of micrometer in length, with the thickness of their organic cell walls in the order of hundreds of nm (Brasier *et al.*, 2014). The task is also complicated by the three-dimensional distribution of the microfossil bodies within the silicified matrix (only a fraction of microfossils might be exposed at the surface of the sample), which would require a layer-by-layer analysis. It is worth noting that the micrometer-scale investigation of putative structures can advance our ability to discriminate between the biotic and abiotic origin (pseudo-fossils) of micrometer-sized inclusions. For example, commonly occurring manganese dendrites resembling branching structures of biological origin will not pass the test for the presence of major biorelevant elements (CHNOPS), whereas organic remains of microfossils will.

In this contribution, we investigated the LIMS capabilities to detect faint element signatures of life by acquiring highly resolved spectrometric data from heterogeneous media present in the Gunflint chert. We will discuss the technical aspects of chemical identification of micrometer-sized inclusions and compare the ablation and ionization performance and quality of the depth profiling from two distinct zones (microfossil assemblage zone and clean host area) using femtosecond laser pulses at three different laser wavelengths. Lastly, we will present LIMS measurement results using a high-voltage (HV) pulse mode applied to the ion optical system of the mass spectrometer designed to enhance the detection sensitivity of species of interest.

Materials

All measurements were conducted on a doubly polished thin section (thickness ~ 30 μm) of a Precambrian chert sample from the Gunflint Formation (see Fig.1). Samples were collected from the Schreiber beach locality (north shore of Lake Superior, Canada). The Gunflint chip was attached to the glass holder and cut to an L shape with dimensions of 1.09 cm in height and 0.8 cm in width. The microfossils are localized within layered zones (laminae). For this study, material probing was mainly conducted within the pure host mineral area (quartz) and within the microfossil-rich laminae. No prior surface conditioning was conducted. The sample was kept in vacuum storage to avoid any surface contamination and handled only with gloves in a cleanroom environment (ISO 5). The sample was mounted on a 25 mm round steel holder with pre-milled cavities where a vacuum compatible copper tape fixed it, and the holder was positioned below the instrument on the XYZ translation stage.

Fig. 1 – Panoramic image of the Gunflint chert thin section mounted on a steel holder (L-shaped sample, left bottom).

Experimental setup and instrument overview

Material ablation, ion production, and subsequent mass spectrometric analysis, as illustrated in Fig. 2, is achieved by generating and guiding a pulsed femtosecond laser beam through three optical sections and focusing the laser beam on the surface of the sample inside a vacuum chamber. Femtosecond laser pulses at IR-775 nm (red line in Fig. 2), UV-387 nm (white line), and UV-258 nm (blue line) wavelengths were applied to ablate and ionize material from the microfossils and the surrounding host mineral. In the first section, the fundamental wavelength from a Ti: Sapphire Chirped Pulse Amplified Clark-MXR laser ($\lambda \sim$775 nm, t $\sim$180 fs, pulse repetition rate $\leq$ 1 kHz, maximum pulse energy: 1 mJ, and s-polarization) is directed to the remotely controlled variable power attenuator (Newport VA-BB-2-CONEX). The attenuated beam is guided further into the second section to generate second and third harmonics on the nonlinear beta barium borate crystals (BaB_2O_4). Second and third harmonic emission ($\lambda \sim$387.5 nm and $\lambda \sim$ 258.3 nm) generation is achieved by using the STORC harmonic generator. After generating additional frequencies, the laser beam is expanded to 35 mm and directed into the periscope system, containing a set of UV enhanced aluminum-coated mirrors, which guides the beam into the vacuum chamber. Inside the chamber, the vacuum is maintained at a pressure of $\sim 5 * 10^{-8}$ mbar by combining a molecular turbopump and an ion getter pump. A fused silica viewport with dedicated broadband anti-reflection coating was installed to reduce the beam transmission losses, which provides an enhanced beam transmittance of $\sim$99 % for the UV-258 nm wavelength. The beam is focused by a doublet lens and forms an ablation spot with a diameter of $\sim$12 μm for IR and $\sim$10 μm for UV-258. The focal point of the incident beam is positioned about $\sim$200–300 μm below the mass analyzer. The sample is positioned in the focal point of the laser beam by using a translation stage, which provides positioning accuracy $\sim$1 μm in the x, y, and z dimensions.

Calibration of the translation stage was achieved by using an internal microscopy system with a resolving power $\sim$1 μm (Wiesendanger *et al.*, 2018). Ions from the ablated plume are confined, accelerated, and focused by the dedicated ion optical system. Mass separation is achieved in the field-free area of the mass spectrometer. A chevron-type ring microchannel plate (MCP) detector is used for the measurements of arriving ions (Riedo *et al.*, 2017). From the initial ion cloud, only positively charged species enter the mass analyzer, as shown in Fig. 2. The time-of-flight spectrum is recorded by collection of the current from the atomic and molecular ions. The signal from the incoming ion flux is amplified by the MCP detector and recorded as a function of

the signal arrival time with a fast data acquisition system. The output current of the detector is measured as electrons*ns^{-1}. The time-of-flight spectrum is converted into a mass spectrum using a quadratic equation $m=a(t+b)^2$, where m is the mass/charge unit, t the arrival time of the charged species and a and b instrument dependent calibration constants. Data acquisition occurs by the simultaneous recording of the signal from two anode rings on the MCP detector. High-resolution digitizers from Agilent technologies with 3.2 GS*s^{-1} are used for data acquisition, providing sampling rates up to ~ 0.3 ns. Control of the ion-optical voltages, as well as translation stage positioning, are achieved with in-house made software. All collected data from the control PC are saved on the storage PC for post-processing and data analysis.
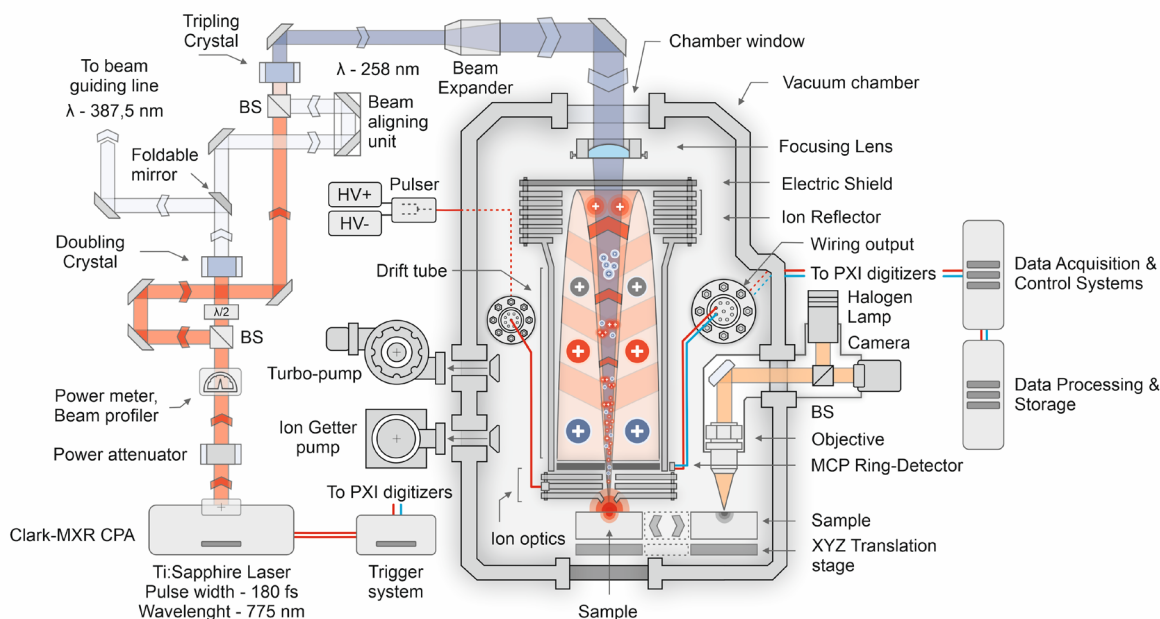

Fig. 2 Schematics of femtosecond laser light transmission and subsequent mass spectrometric analysis of ablated positive ions using LIMS. IR-775, UV-387, and UV-258 nm ablation occur by guiding appropriate wavelength to beam guiding line and periscope system. See text for more details.

Laser beam quality

The applied laser radiation for all used wavelengths is determined to be temporally and spatially Gaussian-shaped (Fig. 3). The spatial pulse profiles have been measured with a CCD camera and widths with an interferometer. The laser beam diameter and pulse width at FWHM were determined to be ~5 mm and 180 fs. The measurements of the laser pulse energy of our laser system (Clark-MXR, Inc.) at the maximum power output (1 mJ) were performed over 10'000 single laser shots and revealed a standard deviation of 0.28%, which corresponds to the nominal level of the laser power stability (Riedo *et al.*, 2013).

Due to the slight variations in the focal point position for different wavelengths, a set of mirror adjustments were applied to achieve the focal point positioning at ~0.3 mm below the extraction electrode of the mass spectrometer. Furthermore, to achieve maximum power transmission through the optical setup, installation started from dielectric mirrors with 99% of transmission for IR-775 nm, and later they have been exchanged to the aluminum-coated UV-enhanced 3-inch mirrors (R_{avg} >90% for

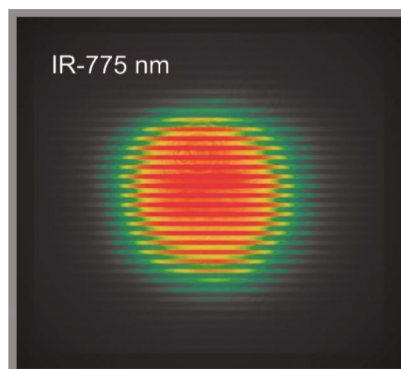250 – 450 nm), for better transmission of the UV-387 nm and UV-258 nm wavelengths.

Fig. 3 - Fundamental IR-775 nm laser beam profile measured with the CCD camera.

"HV-Pulser" mode

Our time-of-flight LIMS instrument with an MCP detector provides a record of all charged species (e.g., element isotopes and molecular species) present in the ablated plume. However, detector sensitivity for heavy species arriving later in the mass spectrum might decrease, which eventually will affect the ability to quantify their abundances, or they might not be registered at all. To avoid the detector saturation or detector gain loss caused by lighter elements of high abundance, we performed measurements with a short high-voltage pulse applied to the acceleration electrode of the ion-optical system. A short description will be given here, for more detailed information with wiring schemes and performance estimations can be found in our earlier publication (Wiesendanger *et al.*, 2017). The HV pulse used in this study was made with a high-speed switch (Behlke FSWP 51-02), with a rise and fall time of approximately ten nanoseconds. The HV pulse connects to the ion optics and forms a positive electric field, which repels positively charged ions from the optimal ion-optical path and switches back to the nominal voltage to allow the remaining ions reach the detector. Appropriate timing to efficiently eject major ions from the confined cloud is achieved by using a set of delay generators (DG535 and DG645, Stanford Research Systems).

Measurements

To probe the chemical composition of the Precambrian chert sample and compare abundances from different locations, we consecutively used three different laser wavelengths: IR-775 nm, UV-387 nm, and UV-258 nm. Also, a high voltage pulser with UV-258 nm laser was used to improve the detection of heavy ions (>150 amu). Each of these regimes was accompanied by applying a variable amount of laser shots. Depth profiles were recorded with a total laser shot count for IR-775 nm – 150,000 within a single spot and histogramming each 100 single laser shots spectra, resulting in 1,500 consecutive spectra. For UV-387 nm, the total laser shot count resulted in 300,000 laser shots from a single spot, histogramming mass spectra from every 200 single laser shots spectra (1,500 consecutive spectra). For UV-258 nm, 500,000 single laser shots were applied, histogramming mass spectra from every 200 single laser shot spectra (2,500 consecutive spectra). The increased number of collected data with short UV wavelengths was motivated by the improved mass spectral resolution and high ion signal levels. All applied laser pulse energies were above the ablation threshold of the investigated material and were set to the maximal achievable

spectral resolution by conducting laser power scans beforehand outside of the zone of interest.

Results

IR-775 measurements

The mass spectrometric studies with IR-775 nm radiation were conducted by applying ~10 TW cm^{-2} irradiances to the surface of the Gunflint sample. The depth profiles (150,000 applied laser shots binned into 1,500 spectra are referred to as "ablation layers") were measured within two locations – the dense assemblage of microfossils and the host mineral area (see Fig. 4A and 4B). Sample overview, acquired crater shapes, and locations are depicted in Figs. 4A, 5A-D. As can be seen in Fig. 5A, IR-775 radiation applied to the host area of the sample produced radial cracks and secondary craters (Fig. 5B) on the surface of the underlying steel holder. The high laser transmission rate through the quartz (SiO_2) explains the formation of the secondary craters, which led to the ablation of the material from the underlying steel holder. The presence of a dark 'halo' around each produced crater could be noted (Fig. 5A). This particular feature was identified as material ablated from the steel holder and deposited onto the bottom of the Gunflint thin section (see Fig.5D).

Moreover, radial cracks on the thin section were observed in the host and microfossil-rich area, as shown in Figs. 5A, and 5C. The formation of the cracks is likely caused by thermal stress occurring in the material during the ablation or by a thermodynamic expansion of the ablated plume originating from the steel holder.
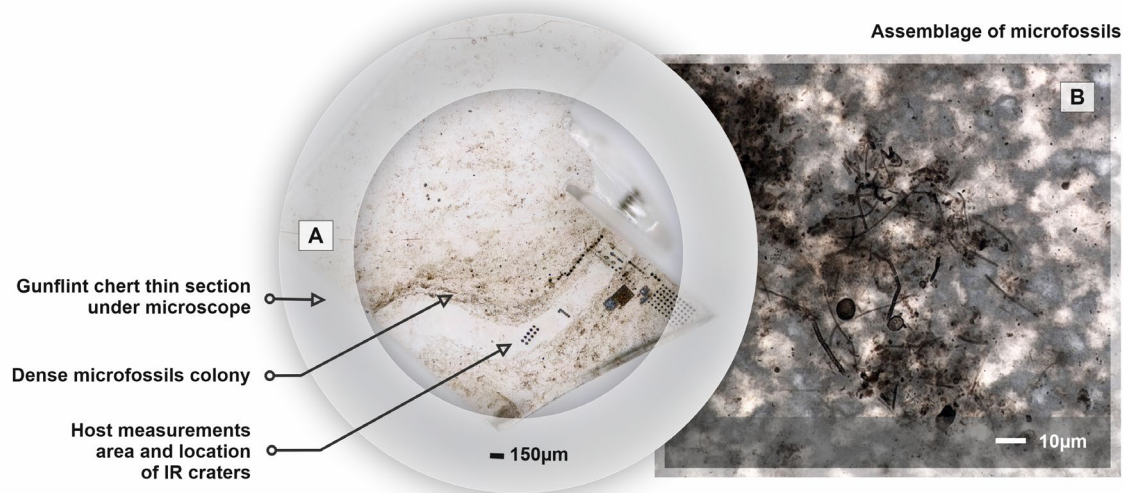
Fig. 4. Panoramic image of the Gunflint chert thin-section. A) Overview of the sample with the denoted location of the host area (white, meandering part in the middle of the sample), microfossil aggregation zone (dark patches), and panoramic view of the location of IR craters. B) Microscopic image of the individual microfossils embedded in the quartz matrix. Various states of microfossil decomposition could be noted.
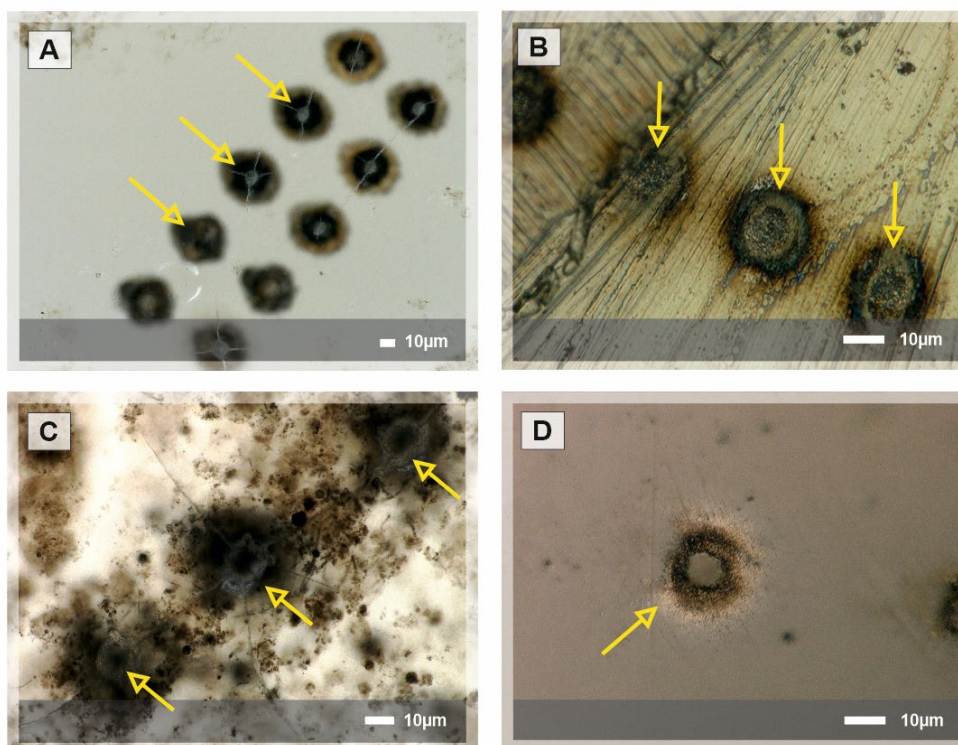

Fig. 5. A) Craters with radial cracks on the surface of the chert sample acquired with IR-775 nm laser within the host area (see Fig.4A). B) Secondary craters on the surface of the steel holder acquired from the same craters as in Fig. 4A. C) Craters with radial cracks produced within the microfossil-rich zone. D) Redeposited ablated material from the steel holder on the bottom of the thin-section.

Typical mass spectra (histogrammed over 150,000 individual laser shots) registered from the microfossil-rich spot and host mineral location obtained during the

campaign are depicted in Fig. 6A in upper and bottom panels, respectively. While investigating the quartz phase from the host location, the laser ablation stability can be maintained, mainly due to the uniform material properties. This results in the accumulation of mass spectra with high mass resolution. By applying the same conditions at the microfossil-rich locations, broadening of the mass peaks and decrease of the mass resolution were observed. The explosive appearance of the craters (presence of cracks and sharp edges) within the microfossil-rich zone depicted in Fig. 5C reveals preferential absorption of the laser energy by the dark features, e.g., the carbon-bearing cell walls of the microfossils. Changes in the ablation conditions at these locations result in changes in the plasma plume characteristics with surface and space charging effects, which may not be easily accommodated by the ion-optical setting of our miniature mass analyzer.

The depth profiles shown in Fig. 6B consist of consecutive signal intensities of ^{28}Si, ^{12}C, and ^{16}O determined by the Simpson integration of the investigated peaks from each spectrum (Meyer et al., 2017). Element isotope intensities are plotted along with the background signal (BG), measured within a time range free of any ion signal. The carbon depth profiles from both locations, depicted in Fig. 6B (thicker curves represent a lowess smoothed version of the data), reveal a nonuniform signal distribution. A significant portion of the carbon signal from the chert is located close to the surface (first hundreds of laser shots), which indicates presence of the surface contamination. Another question concerns the stability of the ion signal intensities from the depth profiles, as is shown in Fig. 6B, with fluctuations of raw silicon signal in the range of up to 10^3. The unstable ion yield is interpreted to be a result of the nonuniform laser ablation and ionization processes. Furthermore, re-deposition of the ablated material, namely, silicon oxides and kerogen from the microfossils, in the form of nanometer- and micrometer-sized particles, can be observed frequently on the surface of the sample (see further in the discussion). This redeposited material can further inter-mix and contaminate surface chemistry within close proximity to the previously analyzed spots.
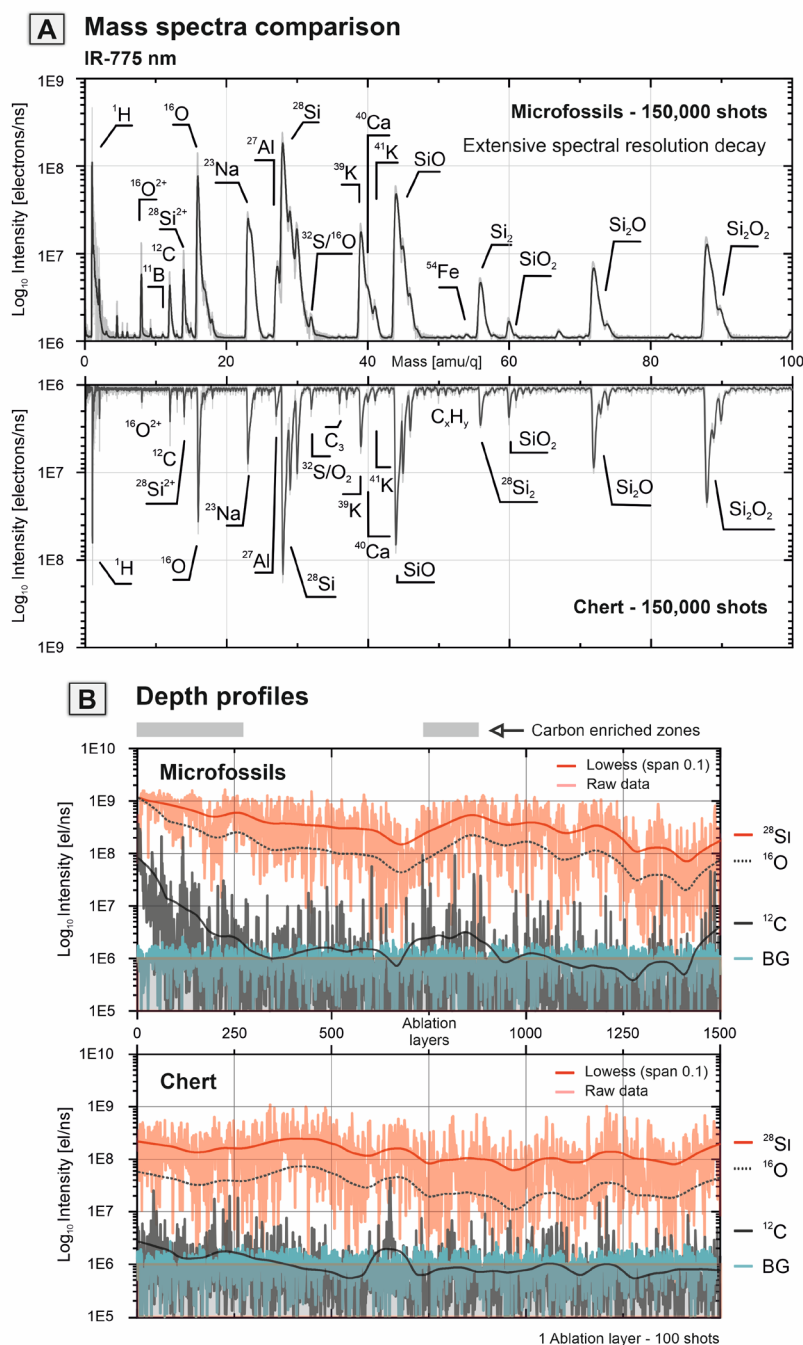
Fig. 6 - A) Mass spectra comparison of IR-775 nm laser ablation campaign in a micro-fossil-rich assemblage zone (top) and clear chert area (bottom). The light grey line in the spectra represents a raw spectrum, and the black line represents a smoothed spectrum. Extensive spectral resolution decay could be observed in the spectra acquired from the microfossil-rich location (top). B) Chemical depth profiles from a microfossil-containing spot (top) and chert spot (bottom) respectively (thick lines – smoothed data, BG – intensity of the background signal. Depth profiles of ^{28}Si (red line), ^{12}C (black line), and a background signal estimate (green) are shown. Dashed solid line represents a smoothed (lowess, span 0.1) ion yield line of ^{16}O.

Analysis of the depth profiles reveals an increased abundance of carbon at certain depths within a microfossil-rich spot and the chert location (indicated by gray boxes above the depth profile). The total carbon content in the microfossil-rich positions is higher than in the matrix (see Fig. 6A and Table 1). However, carbon is also present in the host location; an explanation for this might be found in the location of the matrix analysis spot. As shown in the panoramic picture of the Gunflint chert sample (Fig. 4A), there is a relatively clean matrix, which gradually turns into a dense assemblage of microfossils. Considering the relative proximity of the analyzed spot to the microfossil-rich zone, the presence of diluted organic and inorganic material within a matrix position is a possible explanation.

Elements identified inside the matrix (host area) and the microfossil-rich zone are presented in Table 1. The chemical composition of the microfossil-rich location shows the presence of ^{11}B, ^{10}B, ^{32}S, and ^{54}Fe, which is consistent with the previously reported presence of pyrite within the cell walls of some of the microfossils (Wacey *et al.*, 2013), as well as enhanced intensities of ^{12}C and ^{1}H and doubly charged $^{16}O^{2+}$. The chemical composition of the matrix from the host location reveals the presence of Si and O with contribution of Si oxides, Na, K, Ca, C, H, and hydrocarbon clusters. An increased amount of carbon localized in the close vicinity to the surface likely represents the source of the hydrocarbon clusters. Chert is found to be relatively hydrogen-rich and contains elevated amounts of alkali metals compared to other quartz minerals. Considering the particular chemical composition, quartz from the host area is interpreted as a dehydrated diagenetic product of precipitated silica gel (with relatively abundant species in the seawater - Na, K).

UV-387 measurements

The second set of measurements was performed with UV laser radiation (~387 nm), where the photon energy increases from 1.6 eV (IR-775) to 3.2 eV (UV-387), thus, increasing ionization efficiency. The laser ablation craters formed in these studies are illustrated in Fig. 7A. Mass spectra obtained from the chert and microfossils are shown in Fig. 8A. In comparison to the IR campaign, double the amount of laser shots (300,000 shots) were applied with increased histogramming of 200 shots per accumulated spectrum. Accordingly, the same number of spectra were collected from each position (1,500 in total). The increased number of applied laser shots with the UV-387 laser was motivated by an increased signal resolution and stability of the ion yield. In comparison to the IR-775 nm data, the UV-387 nm laser provides better spectrally resolved signals, e.g., at the ^{28}Si ion, and with reduced amplitude fluctuations in the depth profiles.

The mass spectra shown in Fig. 8A represent averaged chemical composition accumulated from the shot number 100,000 up to 150,000 (i.e., 50,000 shots in total, corresponding to 250 ablation layers) from the matrix (bottom) and microfossils (top), respectively. Fig. 8B depicts the depth profiles obtained from a microfossil and chert location. Data acquired from the surface show an enhancement of carbon signal within both positions, as it was previously observed, which is most probably caused by surface contamination. To avoid contribution from the surface, we analyzed the dataset localized within the depth profile. The most noticeable difference is the increased carbon and hydrogen content, as shown in Table 1 and Fig. 8A.

Moreover, small amounts of ^{51}V, ^{54}Fe, ^{55}Mn, and other isotopes are observed. However, an increased presence of hydrocarbons in the mass spectra makes the unambiguous assignment of elements challenging. On the other hand, an increased abundance of hydrocarbons indicates CH saturated inclusions within the investigated location, which points towards the presence of organic material in the depth profile. Other peaks present in the microfossil-rich location reveal the same Si_nO_m oxides distribution pattern, which means that on the scale of 50,000 shots, matrix signatures dominate the mass spectrum of the fossil location.

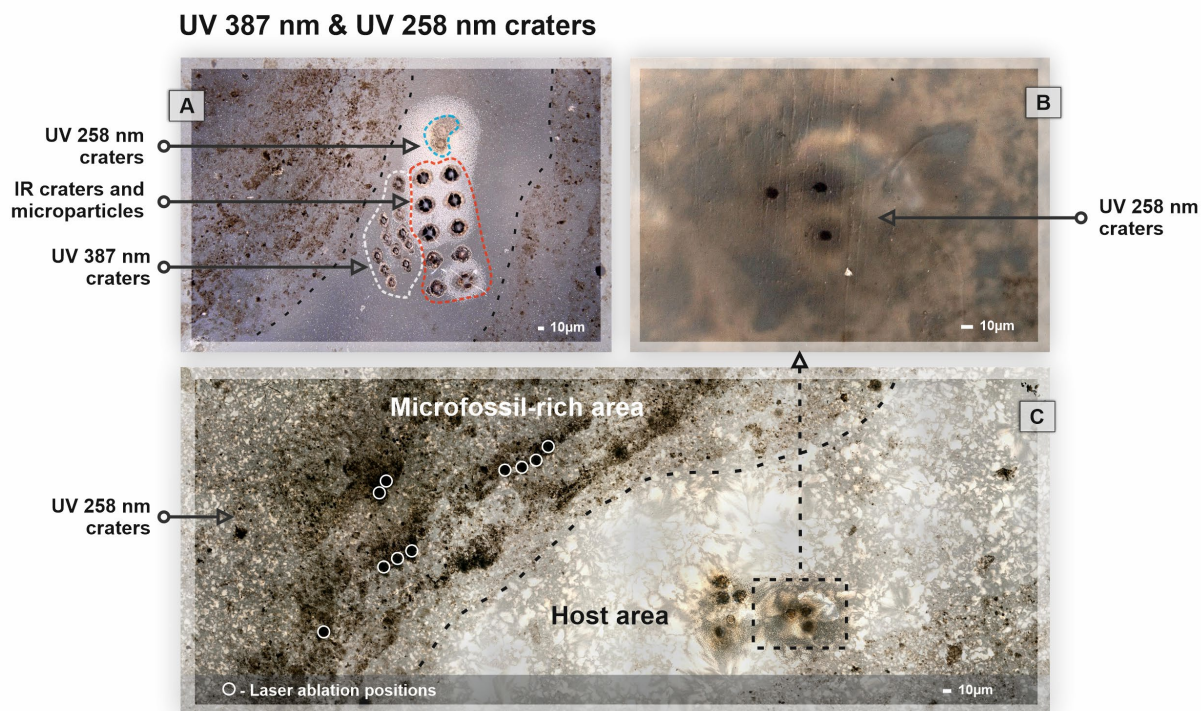**UV 387 nm & UV 258 nm craters**

Fig. 7 - A) Ablation craters produced with different wavelengths. Red encircled area – IR-775 nm ablation craters, white encircled area – UV-387 nm ablation craters, and blue encircled area – UV-258 nm ablation craters. Black dashed lines indicate location of the lamination area and separation of the host area from the microfossil-rich zone B) Close up image of the UV-258 nm craters present in Fig. 6C. C) Panoramic view of Gunflint chert sample with marked positions of the UV 258 nm craters and location of the micro-fossils-rich area and the host area (dashed black line).

Additionally, the spectral resolution compared to the IR results substantially increased (see section: Identification of rare earth elements (REE) and element composition of the analyzed spots), which allows the accurate analysis of the single mass spectra within the depth profiles. Investigation of the laser ablation craters produced with the UV-387 nm laser within the host area also shows an improvement in the crater formation processes. This observation could be attributed to the better absorption of applied laser radiation by the host mineral (Fig. 7A) and increased ionization efficiency. No secondary craters were observed on the backside of the sample, and no cracks formed in the thin section. Figure 9 depicts spectra registered from the microfossil-rich spot and reveals an enhanced ionization of the microfossils. As shown in Fig. 9, the spectrum contains mass peaks of triply charged nitrogen, carbon, oxygen, and silicon. All species were identified within the microfossil-rich zone by averaging 50 individual spectra with the most intense carbon signal, excluding surface data.

Since the main ^{14}N peak is interfering with CH_2 and with doubly charged ^{28}Si, identification of nitrogen is challenging; however, multiply charged species have distinct mass-to-charge ratios, and the presence of triply charged species allowed unambiguous identification of nitrogen $^{14}N^{3+}$ at mass 4.67 in the microfossil-rich zone (Fig. 9) along with $^{12}C^{3+}$, $^{16}O^{3+}$, and $^{28}Si^{3+}$.

Additionally, this observation shows that a high-power scan strategy (high laser irradiances) might be applied to identify elements with isobaric interferences, shifting the identification of elements into the lighter mass-over-charge range. Compared to IR-775 nm, using the UV-387 nm laser wavelength resulted in a considerable improvement in mass resolution, crater shape, and sensitivity.
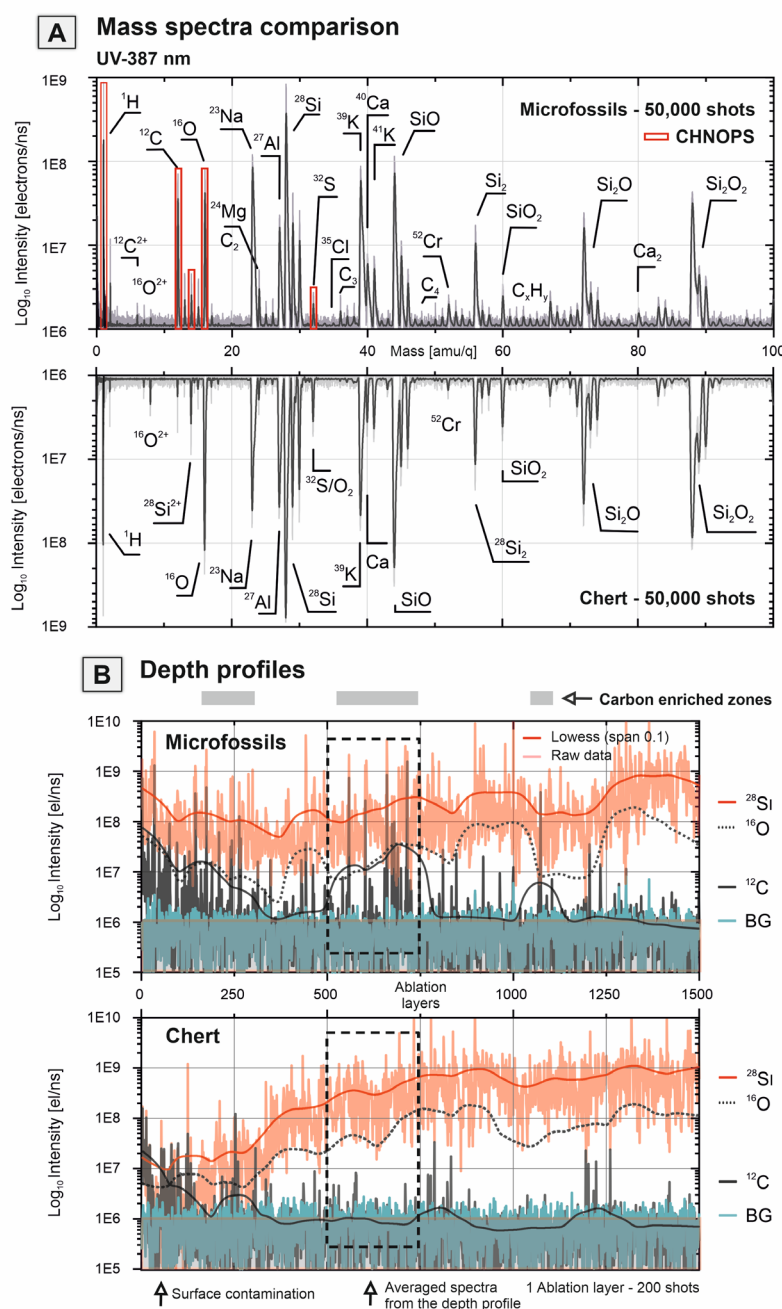

Fig. 8 - A) Mass spectra comparison of UV-387 nm laser ablation campaign in the microfossil-rich zone (top) and clear chert area (bottom). The light grey line in the spectra

represents a raw spectrum, and the black line represents a smoothed spectrum. Red bars show the peak intensities of major biorelevant elements – CHNOPS. B) Chemical depth profiles from a microfossil-containing spot (top) and chert spot (bottom) respectively (smoothed data are shown with thick lines, BG is the intensity of the background signal). Depth profiles of ^{28}Si (red line), ^{12}C (black line) and background signal estimate (green) are shown. The dashed solid line represents a smoothed (lowess, span 0.1) line of ^{16}O. The black dashed square represents the location in the depth profile in which spectra have been averaged.
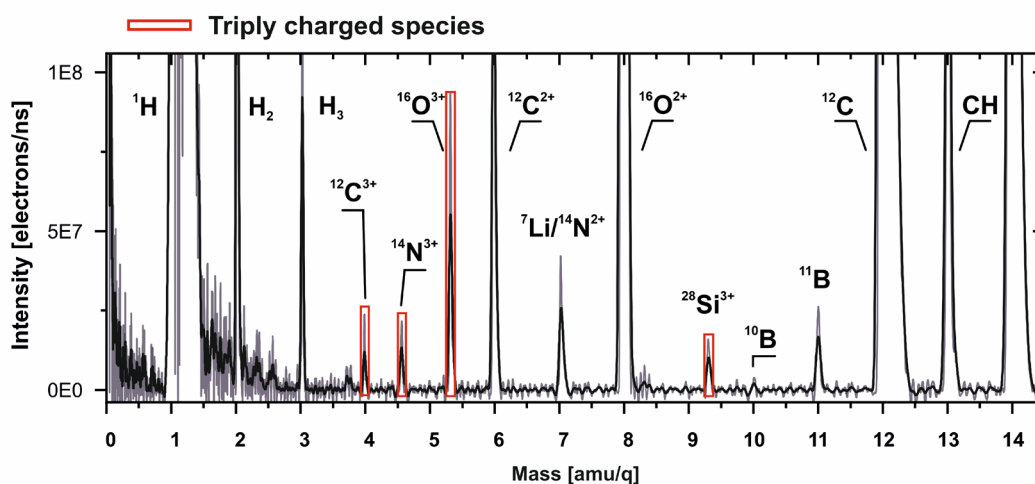

Fig. 9 - Mass spectrum registered from the microfossil-rich location using UV-387 nm laser. Red bars indicate the presence of triply charged nitrogen-14, carbon-12, oxygen-16, and silicon-28 peaks.

UV-258 measurements

The third measurement campaign was performed with the UV-258 nm laser radiation (photon energy reached 4.8 eV, further increasing the ionization efficiency). Single position analysis was performed by applying 500,000 laser shots and registering signals on two digitizing cards. In total, we registered 2,500 spectra from each card, where each of the accumulated spectra consists of 200 single laser shot spectra, making it 5000 spectra from a single depth profile. The experiments were held using a constant repetition rate of 1 kHz between batches of 200 laser shots. After finishing the batch of 200 laser shots, a one-second delay was implemented to ensure the spectrum transfer to the SSD. The length of the data recording for a single depth profile amounted to 50 minutes. Unlike the mass spectrometric studies with UV-387 nm and IR-775 nm laser radiation, where the mass spectra were measured from only two channels of the detector, the measurements with UV-258 nm were recorded on three anode channels (to mitigate signal saturation and fluctuations effects). This approach allows recording the ion signal collected from the MCP detector with a reduced probability of saturation. However, it requires a different ion-optical configuration (for an even distribution of ions on the MCP detector) and the collection of two separate data sets.

Figure 10 depicts mass spectra and depth profiles collected with the UV-258 nm laser radiation. In comparison to the studies with the IR-775 and UV-387 laser radiation, the mass peak intensity of ^{28}Si is observed to be less variable, allowing high-quality depth profiling analysis, both within the host and the microfossil-rich locations. The spectral resolution remains on the nominal level, even during the transition

between carbon-containing cell walls of the microfossils and pure silica matrix. Figure 10A depicts a comparison of two mass spectra histogrammed over 200 shots each, in two distinct zones – within the body of a microfossil and inside a clean host matrix. Selective low-scale averaging (200 shots) was applied here to emphasize the importance of single spectrum accumulation. The ^{12}C depth profile from the microfossil-rich location reveals multiple carbon sources located at different depths (see gray boxes in Fig. 10B). The signal from the microfossil-rich location differs from the chert and could be characterized by the presence of biologically relevant elements (CHNOPS) (see red bars in Fig. 10A).

Similar to the data obtained with the UV-387 nm laser, individual spectra along the depth profile reveal high-intensity mass peaks of ^{12}C and ^{1}H with the presence of mass peaks of carbon clusters but at much lower intensities. Another notable feature observed in the mass spectra of the microfossils is the presence of ^{31}P, ^{32}S, ^{34}S, and elevated values of transition metals with biological relevance ^{48}Ti, ^{51}V, ^{54}Fe, ^{55}Mn, ^{58}Ni, and ^{63}Cu, which are considered to be essential for the operation of oxygenic photosynthesis (Shcolnick & Keren, 2006). Elevated values of these elements compared to the abundances of other elements, like C and Si, result from more efficient ionization of metallic species. The host mass spectrum, depicted on the lower part of Fig. 10A, shows the elements expected from the host (quartz) mineral - Si and O with typical Si clusters and oxides distribution pattern, and additional contributions of H, Na, Ca, and K.

Moreover, the carbon mass peak intensities from the host location (quartz) are observed to be negligibly low with rare spikes in the depth profile, which may be interpreted as a contribution from the surface, reflecting the growing diameter of the ablation crater with increased depth. The presence of carbon on the surface of the matrix location was identified as contamination. As shown in Fig. 10B (see gray boxes), using carbon as a tracer of microfossils, we identified multiple locations of microfossil bodies, which are distributed within the SiO_2 matrix. Removal of several layers of the matrix is required to acquire a spectrum from the embedded microfossil. In this particular case, the microfossil spectrum was acquired at a depth of 70,000 shots.
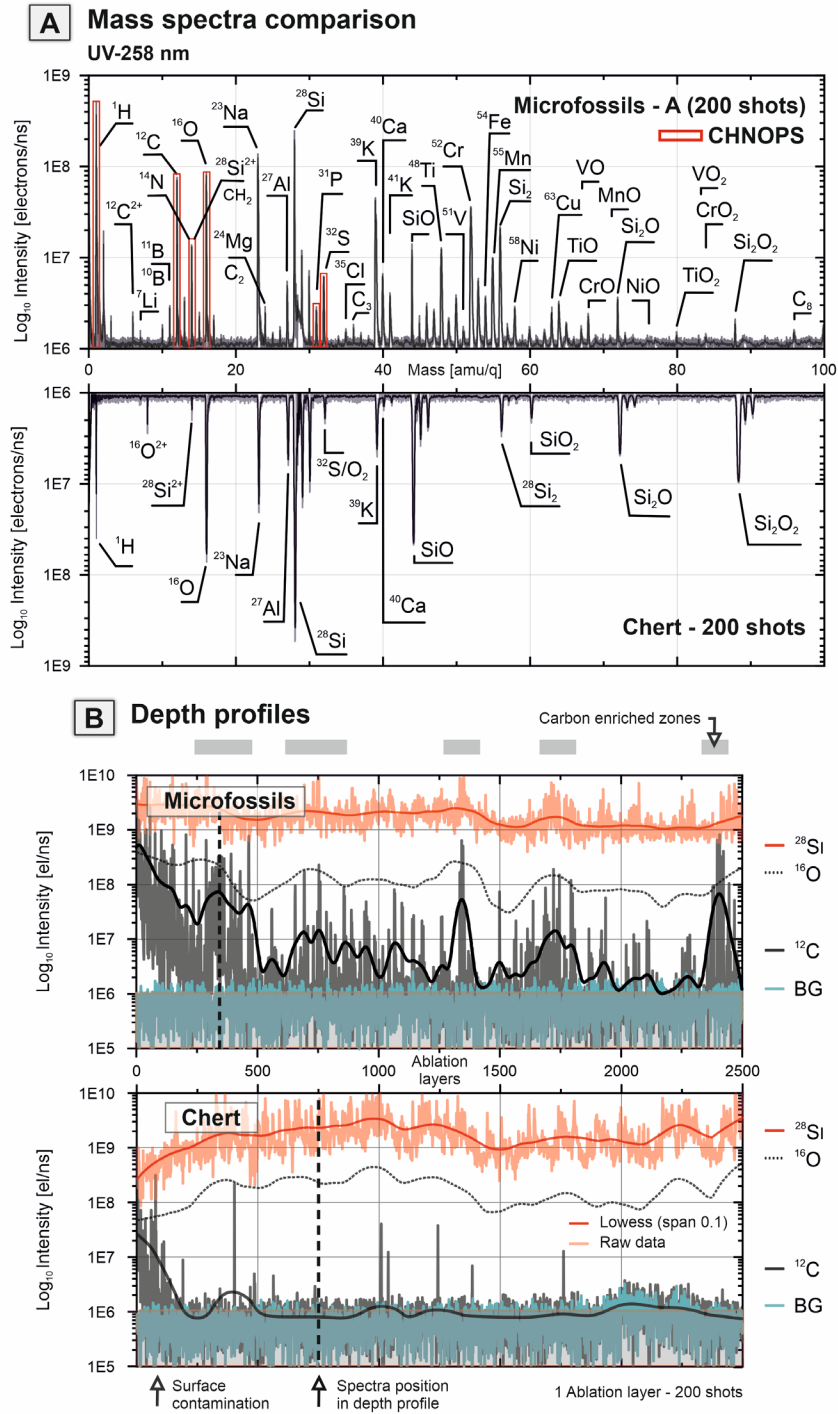
Fig. 10 - A) Mass spectra comparison of UV-258 nm laser ablation campaign in microfossil-rich assemblage zone (top) and clear chert area (bottom). The light grey line in the spectra represents a raw spectrum, and the black line represents a smoothed spectrum. Red bars show the peak intensities of major biorelevant elements – CHNOPS. B) Chemical depth profiles from a microfossil-containing spot (top) and chert spot (bottom) respectively (thick lines – smoothed data, BG – intensity of the background signal. Depth profiles of ^{28}Si (red line), ^{12}C (black line) and background signal estimate (green) are shown. The dashed solid line represents a smoothed (lowess, span 0.1) line of ^{16}O. The black dashed vertical line represents location of the spectra in the depth profile shown in Fig. 10A.

Identification of rare earth elements (REE) and elemental composition of analyzed spots

Figure 11A depicts a spectrum from a microfossil and reveals a similar element composition to previously described spectra but with the additional identification of La and Ce. The element composition of the analyzed locations is given in Table 1 (Column: Fossil-B). The distribution of La and Ce within the analyzed spot reveals a highly confined character, i.e., elements were identified in two distinct locations (see gray boxes, Fig. 11B). Each of these spectra shows the correlation with the elements previously identified in other microfossil-rich zones. The co-occurrence of P, O, La, and Ce can indicate the presence of an inclusion of a rare-earth phosphate mineral, i.e., monazite. The appearance of La and Ce together with carbon also raises questions about potential preservation of parts of the microfossils by monazite, which has been previously reported for microfossils from 1 billion-year-old Torridon phosphates (Wacey *et al.*, 2019). Monazite was suggested as the first mineral precipitate to form after cell death in the Torridon samples, forming an insoluble, stable phase, thus preserving fine structural information. However, a limited number of detections from the Gunflint microfossils do not allow us to infer a definitive correlation between monazite preservation and microfossils.

The abundance of REE from the Gunflint chert was reported previously (Shimizu & Masuda, 1977), with bulk values below one ppm: La – 0.0785 ppm, Ce – 0.1946 ppm. Other REE concentrations were estimated to be significantly lower (measurements were performed using a stable isotope dilution technique) (Shimizu & Masuda, 1977). Our observation also shows that heavier REEs' abundances are below the detection limit, which points towards the necessity of data acquisition with an even smaller number of laser shots to be binned (histogramming of 10–50 single laser shot mass spectra) to improve the detection limits. Enriched Ce and La points towards the presence of oxidizing conditions within Gunflint waters and supports a cyanobacterial affinity of at least some microfossils, which might indicate microbially mediated oxygenation of the photic zone. Another piece of evidence comes from the identification of Mn and Fe, two relatively abundant elements within the Gunflint sample. Mn has a similar redox profile as Ce (Mofett, 1994), and it is known that large numbers of microorganisms enzymatically oxidize Mn (Lozano & Rossi, 2012). Bulk rock geochemistry from the Gunflint formation (Fralick *et al.*, 2017) suggests that La and Ce concentrations vary within the formation in the range from single-digit ppms to hundreds of ppms (including data from calcite stromatolites, conglomerates, ankerite, and calcite grainstones). These lithologies represent variations within the stratigraphic column, reflecting changing input from different sources - detrital and fluvial to hydrothermal REE's and transition metals.

The total atomic composition of the analyzed locations is given in Table 1. More elements can be identified due to the increased sensitivity and improved ionization efficiency using shorter UV wavelengths (see Table 1. Columns: Fossil-A and Fossil-B). Figure 12 reveals histograms of ^{28}Si peak mass resolutions (left) and ^{28}Si signal intensities (right) measured in different depth profiles and using different wavelengths. The width of the distributions indicates the incidence of different mass resolutions and peak intensities (width is normalized to the same height between laser wavelengths and samples). In comparison to the IR measurements, spectral

resolution within highly absorptive microfossil bodies remains at the nominal level, without any significant distortions of the spectral features (see Fig. 12).
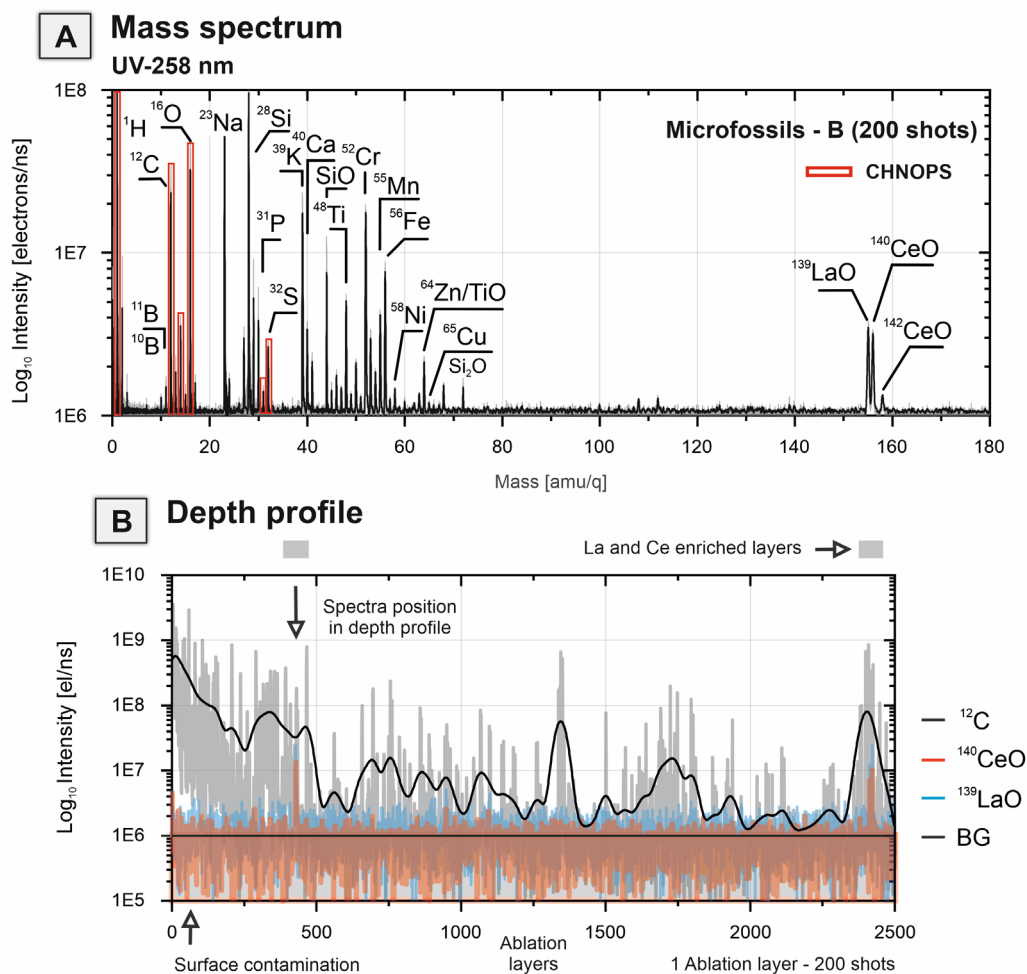

Fig. 11 - A) Mass spectrum from UV-258 nm laser ablation campaign from a microfossil-rich zone with detection of La and Ce oxides. Red bars show the peak intensities of major biorelevant elements – CHNOPS. B) Chemical depth profile in microfossil-rich zone outlining the scarce distribution of REE (thick lines –smoothed data). The black arrow indicates the location of the spectra in the depth profile shown in Fig. 11A. The depth profile of ^{12}C is shown with the light grey line. The depth profile of LaO is shown with the blue line. The depth profile of CeO is shown with the red line. Gray boxes on top of the figure indicate the location of the layers enriched in La and Ce. A black horizontal line indicates location of the background signal.

By collecting data in situ, within the microfossil bodies, a gradual increase of C and H with shorter laser wavelengths was detected (Table 1. Fossil-A). The atomic fraction of H increased from ~2% in the chert (host area) to ~24% within the microfossil body. A similar increase was detected for C ranging from non-detection (n.d.) to ~7%. A negative correlation was detected for the host elements on the same location: Si concentration decreases from ~77% to ~15% in the spectra acquired from the microfossil. A comparable decrease was observed for O as well: from ~ 15% to ~8%. We inferred the presence of pyrite microscopic-inclusions from the high values of Fe – an increase from n.d. in the host area to ~6% within a microfossil. Similarly, S increased from ~0.2% to ~0.7%, which is in agreement with previously reported data

(Wacey *et al.*, 2013). An almost identical distribution of element composition was detected in location B (see Table 1. Columns: Fossil-B), with the additional presence of REE elements (La and Ce). Application of the relative sensitivity coefficients (RSC's determined for matrix-matched quartz latite standard sample (USGS QLO-1)) (Neuland *et al.*, 2016) reveals accurate stoichiometric ratios for Quartz matrix [Si/O] ~1/2.
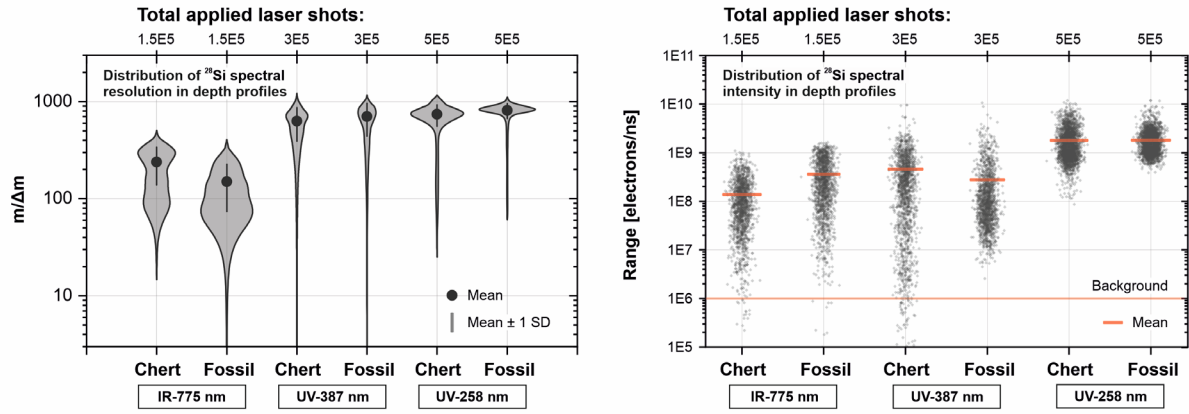

Fig. 12. Left – The distribution of the ^{28}Si spectral resolution in the depth profiles for each applied wavelength within two distinct zones. Distributions are calculated using kernel density estimates and have been normalized to the mode values. Mean value location and 1SD variation indicated with a dot and a vertical line. Right – The distribution of the ^{28}Si intensities in the depth profiles for each applied wavelength within two distinct locations (microfossil-rich zone and host mineral area) is shown as a scatter plot. The red thick horizontal line indicates location of the mean peak intensity registered from the depth profile. The red horizontal line indicates location of the background signal. The numbers at the top of the figures indicate the total laser count applied to the given locations.

Atomic fractions of elements measured with three femtosecond laser wavelenghts in microfossils and chert.

№	Elements	IR-775 nm, Eγ=1.6 eV		UV-387 nm, Eγ=3.2 eV		UV-258 nm, Eγ=4.8 eV		
		Chert (± *int. err.*)	Fossil (± *int. err.*)	Chert (± *int. err.*)	Fossil (± *int. err.*)	Chert (± *int. err.*)	Fossil - A (± *int. err.*)	Fossil - B (± *int. err.*)
Laser shots averaged:		150,000	150,000	50,000	50,000	200	200	200
1	H	15.11 (5.55)	16.18 (1.06)	7.49 (5.55)	17.9 (0.55)	1.811 (0.03)	23.98 (1.76)	28.26 (0.19)
2	Li	n.d	n.d.	n.d.	n.d.	n.d.	0.02 (0.01)	0.01 (0.01)
3	B	n.d	0.02 (0.01)	n.d.	n.d.	n.d.	0.20 (0.01)	0.11 (0.01)
4	C	0.50 (0.14)	0.75 (0.01)	0.04 (0.01)	4.02 (0.02)	n.d.	6.97 (0.03)	5.03 (0.04)
5	O	14.16 (1.93)	15.58 (0.26)	10.00 (0.03)	4.72 (0.02)	15.34 (0.06)	8.25 (0.05)	7.31 (0.05)
6	Na	3.95 (0.19)	9.19 (0.02)	4.33 (0.02)	12.55 (0.02)	3.39 (0.02)	16.91 (0.07)	15.8 (0.05)
7	Mg	n.d.	0.18 (0.03)	0.07 (0.02)	0.25 (0.01)	n.d.	0.16 (0.01)	0.18 (0.01)
8	Al	0.61 (0.04)	0.96 (0.01)	4.09 (0.02)	2.09 (0.02)	1.09 (0.02)	0.49 (0.01)	0.54 (0.01)
9	Si	63.44 (3.35)	50.03 (0.11)	68.47 (0.03)	48.67 (0.25)	77.32 (0.17)	15.34 (0.15)	19.97 (0.05)
10	P	n.d.	n.d.	n.d.	n.d.	n.d.	0.20 (0.01)	0.08 (0.01)
11	S	0.28* (0.02)	0.12* (0.01)	0.22* (0.01)	0.09 (0.01)	0.18* (0.01)	0.74 (0.01)	0.42 (0.01)
12	Cl	n.d.	n.d.	n.d.	0.09 (0.01)	n.d.	0.07 (0.01)	0.02 (0.01)
13	K	1.87 (0.04)	6.63 (0.02)	5.04 (0.02)	8.3 (0.02)	0.81 (0.01)	7.51 (0.03)	5.91 (0.05)
14	Ca	0.08 (0.01)	0.11 (0.01)	0.22 (0.01)	0.66 (0.55)	0.06 (0.01)	0.85 (0.02)	0.73 (0.02)
15	Ti	n.d.	n.d.	n.d.	n.d.	n.d.	2.62* (0.01)	1.72 (0.02)
16	V	n.d.	n.d.	n.d.	0.02* (0.01)	n.d.	0.07* (0.01)	0.05 (0.01)
17	Cr	n.d.	0.01 (0.01)	0.03 (0.01)	0.09* (0.01)	n.d.	7.84* (0.03)	7.08 (0.02)
18	Mn	n.d.	n.d.	n.d.	n.d.	n.d.	1.61* (0.02)	1.01 (0.02)
19	Fe	n.d.	0.26 (0.01)	n.d.	0.54* (0.01)	n.d.	5.60* (0.01)	3.79 (0.02)
20	Ni	n.d.	n.d.	n.d.	n.d.	n.d.	0.34* (0.01)	0.11 (0.02)
21	Cu	n.d.	n.d.	n.d.	0.03* (0.01)	n.d.	0.27* (0.02)	0.11* (0.01)
22	La	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.94 (0.02)
23	Ce	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.86 (0.03)

* values affected with isobaric interference

Table 1 – Atomic fractions of elements measured with three wavelengths of the femtosecond laser in microfossils and chert. Measurements are from different locations and represent uncalibrated abundances measured from spectra shown in figs. 6,8,10,11.

UV-258 nm measurements with HV-pulser

As mentioned previously, time-of-flight mass spectrometers can register positively charged species present in the ablation plume. However, highly abundant light ions, arriving early to the detector, with ion flux values exceeding 10^7 ions*s^{-1} might lead to the saturation of the detector (caused by the dead time of the triggered MCP pores) (Riedo et al., 2017; Wiesendanger et al., 2017), which will result in the reduced sensitivity of the detector for heavy elements and molecules. To suppress saturation effects, we applied two technical solutions: 1) Enlargement of the active area of the detector and implementation of the impedance matched multi-anode. 2) Implementation of a short high-voltage (HV) pulse applied to the ion optical system of the mass analyzer to repel the flux of the major light ions from the nominal ion-optical path. A mass spectrum measured with the HV pulser and UV-258 nm laser radiation is depicted in Fig. 13 and represents an average composition (200,000 laser shots) of the quartz mineral from the host location. In comparison to the conventional spectrum acquisition mode, the mass range of the detected species was substantially enlarged (by factor 4). The spectrum depicted in Fig. 13 illustrates the distribution of various silicon oxide clusters and reveals a good agreement with theoretical calculations of stability ranges of silicon oxides (Lepeshkin *et al.*, 2019). The identified species span from relatively small oxides - Si_3O_3, Si_3O_4, with masses 132 and 148 amu, respectively, up to the largest identified silicon oxide - $Si_{11}O_{22}$ (660 amu). Measurements conducted within the microfossils revealed an identical cluster distribution profile (on the same averaging scale), which supports the conclusion that signal from the microfossils is very narrowly localized within specific depths and quickly decays to the matrix values and requires the depth profile analysis with an averaging scales between 50 to 200 single laser shots.
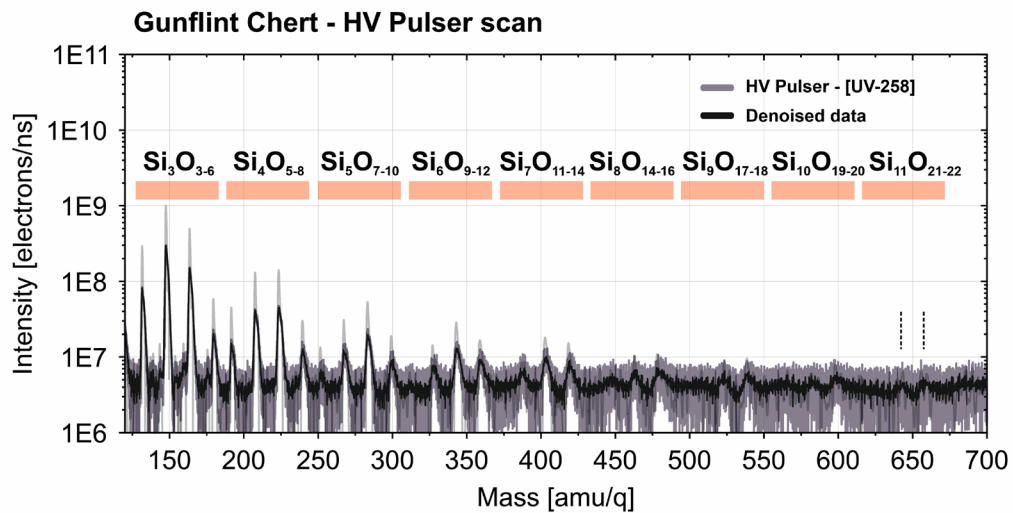
Fig. 13. Mass spectrum of Quartz mineral from chert in acquired in the HV pulse mode. Red bars on top of the spectrum indicate the location of the specific Si_xO_y chains. The last identified molecules are shown with two dashed horizontal lines and represent large molecules of $Si_{11}O_{21}$ and $Si_{11}O_{22}$ (m/q = 644 and 660). The light grey line indicates the raw spectrum, and the black line represents denoised mass spectrum.

Discussion

The main scope of this study was the investigation of the applicability of the integrated LIMS system (mass spectrometer and a microscope) for the detection and characterization of microscopic organic inclusions from astropaleontologically relevant samples, such as cherts, siliceous sinters, and siliciclastic deposits (McMahon *et al.*, 2018). One of the main advantages of the LIMS system is that it can provide localized and spatially resolved *in situ* element analysis (Riedo *et al.*, 2016) combined with molecular information (Ligterink *et al.*, 2020; Moreno-García *et al.*, 2016). However, performance of the mass analyzer coupled to a laser ionization source inherently depends on the physical parameters of the ablation/ionization. The implementation of femtosecond lasers instead of nanosecond lasers was shown to reduce various fractionation processes and matrix effects (Zhang *et al.*, 2013), and the corresponding relative sensitivity coefficients (RSC) tended towards one for polymetallic samples (Riedo *et al.*, 2013). Nevertheless, despite the LIMS applicability to an extensive range of samples, the quality of gathered data might vary. Optically transparent, dielectric samples with heterogeneous, highly absorptive inclusions, e.g., microfossils, are challenging targets for analysis, as shown in the results section for the longer wavelengths. For example, bandgap values reported for α-quartz are determined to be around 6 eV (Calabrese & Fowler, 1978) and 9 eV (Chelikowsky & Schlüter, 1977), which means that for an effective release of charge carriers from quartz, the absorption of a higher amount of photons is required, compared to the absorptive kerogen embedded in a silicate matrix. Within the experimental framework, analysis of such different materials with LIMS using an IR laser requires tuning the energy of the laser for each investigated material to achieve well-resolved spectral data (which is hard to accomplish within a single position). The issue of low spectral quality of data obtained from geological samples with the IR laser was addressed by filtering acquired spectra (Wiesendanger *et al.*, 2019). This method applies spectral quality scoring and selects only resolved peaks for the analysis, while discarding spectra affected by broadening. Method revealed significant improvement in the determination of isotope ratios, however, a substantial amount of data might be discarded, and thus information about the chemical depth profile may be lost. Unequal ionization of different minerals can also affect the mass spectrometric imaging quality, where potentially more sets of minerals can be present, which consequently has a potential to introduce artefactual imaging results.

Some of the processes observed during the ablation of the thin section from the Gunflint chert are shown in the sketch in Fig. 14. Panel A describes the formation of the ablation crater after a number of applied IR laser shots. Since the sample and the glass holder are nearly transparent to the applied wavelength, part of the IR beam passes through the transparent media and ablates material on the surface of the metal holder, forming a secondary crater. This process can be accompanied by the formation of radial cracks and possibly by contamination of the surface with the steel ablation products.

Figure 14B depicts the ablation of embedded microfossils due to their higher absorptivity of the IR radiation and subsequent formation of the secondary craters. The figure illustrates the quick removal of the absorptive microfossils from the analyzed spot. In contrast, ablation and ionization with UV-258 nm radiation provide a uniform absorption of photons by the host mineral and the microfossil inclusions, as illustrated in Fig. 14C. This conclusion is verified by the absence of the secondary craters on the underlying steel holder. Moreover, the distribution of the spectral resolution

and intensity of ^{28}Si, depicted in Fig. 12, indicates an improvement of the registered signal quality by using shorter laser wavelengths. Figures 15A and 15B show the close-up images of the obtained craters and demonstrate a gradual improvement of the crater morphologies by using shorter wavelengths. Figure 15C shows the AFM crater morphology measurement obtained from the host mineral area using a UV-258 nm laser (see the location of the same crater in Fig. 15B (white arrow) and the depth profile registered from the same crater in Fig. 10B). The figure indicates that a 10 μm wide crater quickly decays to half of its diameter at depth ∼ 4 μm and forms a conical crater with a sharp ending. The crater most likely goes as deep as ∼ 25–28 μm; however, the AFM depth range was not sufficient to measure such deep craters. Therefore, an accurate estimation of ablation rates could not be achieved. However, the precise ablation rate for the same instrument was previously determined for NiCr standard material, where a mean ablation rate of ~3 nm/pulse (~72 nJ/pulse) was observed using the double-pulse UV-258 nm fs-laser radiation to produce the optimal mass-spectrometric signal quality (Grimaudo *et al.*, 2020). In our case, ~360 nJ/pulse for UV-258 nm was needed to produce the optimal mass-spectrometric signal quality. Accordingly, the more absorptive NiCr standard sample requires less energy for ablation in contrast to the less absorptive quartz mineral. As a first-order approximation, we can assume that the mean ablation rate is in the same order as in (Grimaudo *et al.*, 2020), which gradually decreases during the formation of craters and likely ends with thermal ionization from side walls with even lower ablation rates.
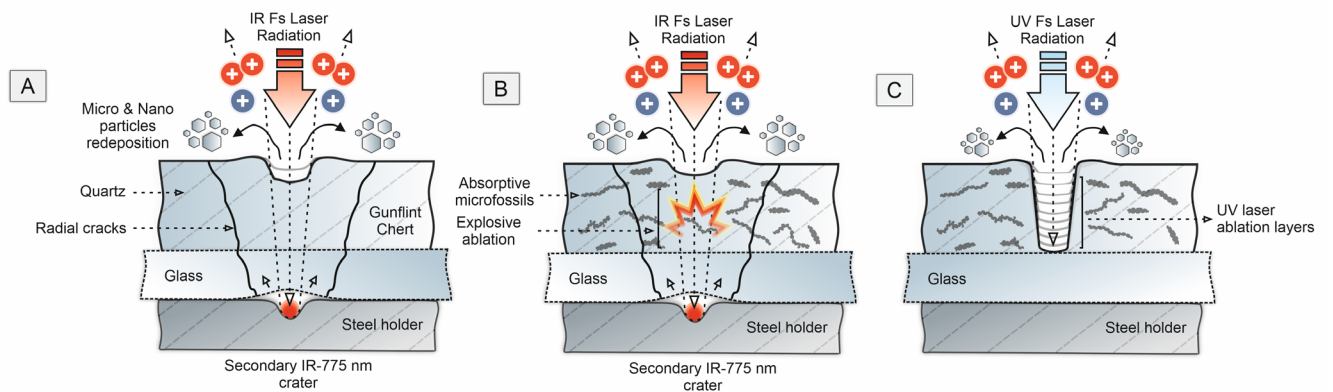

Fig. 14 - Schematics of the femtosecond laser ablation of the Gunflint chert sample. A) Ablation of the host area with IR-775 nm laser and formation of the secondary craters by residual laser energy not absorbed by the quartz mineral (see Fig.5). B) Ablation of the microfossil-containing area with IR-775 nm laser with formation of the secondary craters, radial cracks, and preferential absorption of the laser energy by the microfossils (see Fig. 5). C) Ablation of the microfossil-containing area with UV-258 nm laser with more uniform ionization of microfossils and surrounding chert. The absence of secondary craters and cracks could be noted (see Fig.7).
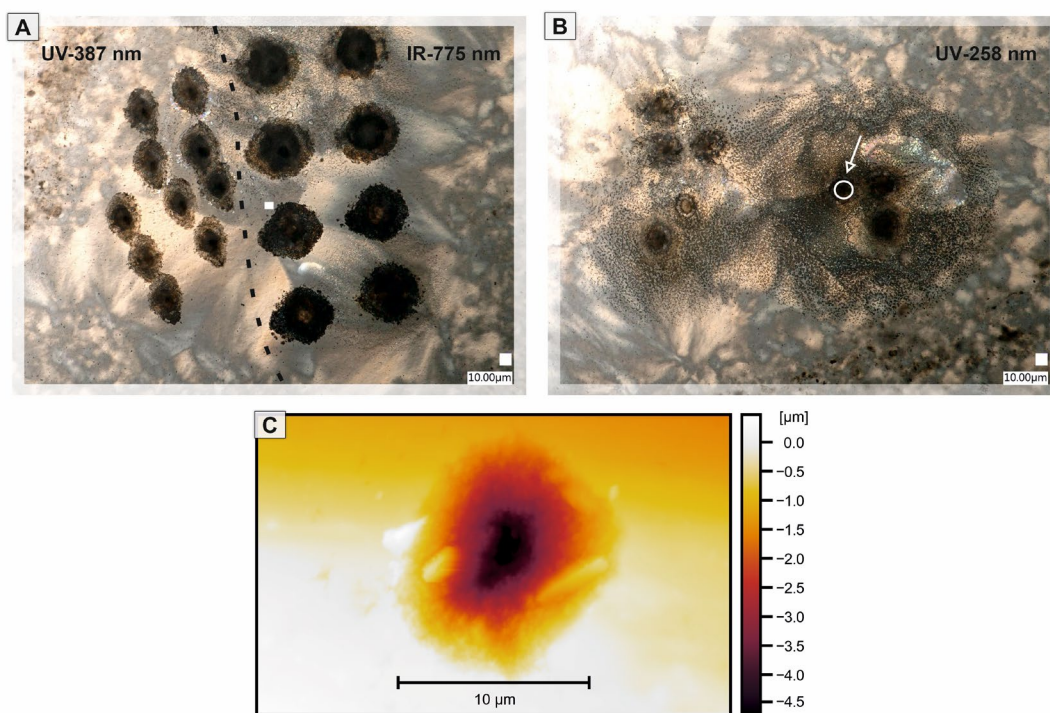
Fig. 15 - A) Microscopic image of the IR-775 and UV-387 nm ablation craters from the host area. Black dashed line indicates separation boundary between two different wavelength craters B) Microscopic image of the UV-258 nm craters from the host area. Scale of the image is equal to the scale of image in Fig. 13A. White encircled crater with an arrow indicates location of the crater shown in Fig. 15C. C) Result of the atomic force microscopy (Z-scan) of the UV-258 nm crater acquired from the host area (noted with white arrow in figure 15B). Depth profile recoded from this location shown in Fig. 8B.

Measurements of the mass resolution of the ^{28}Si peaks obtained with the IR-775 nm laser are shown in Fig. 12, left panel, with a distribution of m/Δm values ranging from 10 to 500, with the better resolution within the matrix than in the microfossil-rich location. UV-387 nm and UV-258 nm laser depth profiles reveal distributions with a higher mean mass resolution. For the UV-387 nm laser, data reveal highly stable m/Δm values around 800 at both matrix and microfossil locations. However, the extended tailing of m/Δm distribution towards lower values (see Fig. 12, left panel) is a definite drawback. In contrast, the UV-258 nm data reveals highly stable m/Δm values around 800 in both matrix and microfossil locations, without any significant tailing. Considering that the total applied laser shot count is much higher, 500,000 versus 300,000 and 150,000 for UV-258, UV-387, and IR-775, respectively, shorter UV wavelengths have a clear advantage in the depth profiling and large-scale spectral data collection (i.e., mass spectrometric imaging). Figure 12 (right panel) depicts the distribution of the ^{28}Si peak intensities and again demonstrates an improvement of the signal stability when using shorter laser wavelengths. Distribution of the intensities could also be observed in the depth profiles; however, stacked representation helps to recognize that a UV-258 nm laser provides more stable measurements with an improved mean signal-to-noise ratio $\sim 10^4$.

However, despite clearly improved ablation conditions, there are still some limitations: the spot size and accessible depth of the analysis are limited. Craters deeper than tens of μm are prone to admixing of signal from the crater walls since the laser

beam is of Gaussian shape. Additionally, the decay of signal intensities could be observed when the crater gets sufficiently deep (Cedeño López *et al.*, 2018; Grimaudo *et al.*, 2019). This aspect can be improved with an appropriate beam shaping (π-shape) and sample positioning so that the focal position of the incident laser beam tracks the surface of the analyzed material.

Moreover, in-situ characterization of the molecular composition of the microfossils would be considered beneficial. The analysis can be made with local direct laser desorption studies, as has been demonstrated before (Ligterink *et al.*, 2020). Even though the desorption of molecules from the microfossils is out of the scope of this study, our preliminary results show that an IR wavelength might provide better performance compared to UV wavelengths. More energetic photons from the deep UV laser are expected to increase the fragmentation of the molecules and therefore decrease the probability of detection of a parent peak within already substantially decomposed organic matter. However, more detailed studies are required to understand the performance of the different lasers in the direct desorption of kerogen molecules from the highly silicified media.

Overall, the mass spectrometric analysis conducted on the 1.88 Ga Gunflint chert sample with LIMS instrument and integrated microscope system using three different laser wavelengths clearly demonstrated the advantage of the UV-258 nm laser wavelength over longer wavelengths for elemental analysis and depth profiling of heterogeneous samples.

Conclusions

In this article, we have investigated the capability of the LIMS microprobe system to provide chemical depth profiles from a heterogeneous Precambrian chert sample and collected information about the chemical composition of microfossils embedded in the quartz matrix. We have discussed the generation and performance of three different femtosecond laser wavelengths: IR-775 nm, UV-387 nm, and UV-258 nm, and their relative advantages and drawbacks in the analysis of silicified geologic samples. Additionally, analysis of the host media was performed with the HV pulser mode. By combined usage of optical microscopy, in-situ mass spectrometry, and depth profiling, the chemical composition of microscopic inclusions within the chert could be characterized at precisely defined locations and depth positions. Chemical depth profiling is aimed to enhance the scientific output from a single position measurement and provides a pathway to highly resolved three-dimensional mapping. By tracking the ^{28}Si mass peak resolution, intensities, and crater shapes, we identified a significant increase in the analytical performance by using shorter laser wavelengths. Ablation with a UV-258 nm femtosecond laser revealed geometrically better-defined crater morphologies, higher and more stable spectrometric mass resolution (up to 800), and almost uniform ionization of heterogeneous materials present in the depth profiles. Identification of the microfossils and the quartz matrix was performed by tracking the carbon, silicon, and oxygen ion yield in the depth profiles. Concentrations of different elements within two distinct zones and their relative change is summarised for each wavelength as atomic fractions. The main elements identified within the silica matrix are Si, O, Na, K, H, and Ca. Analysis of the chemical composition of microfossils revealed the presence of the same elements as found in the matrix plus contributions from a suite of elements specific to the fossils, namely: Li, B, N, Mg, Cl, P, S, Ti, V, Mn, Fe, Cu, Ni, La, Ce, with strongly enhanced C and H.

References

Allen C. C., Westall F., and Schelble R. T. (2001) Importance of a martian hematite site for astrobiology. *Astrobiology*, 1: 111-123.

Alleon J., Bernard S., Le Guillou C., Daval D., Skouri-Panet F., Pont S., Delbes L., and Robert F. (2016) Early entombment within silica minimizes the molecular degradation of microorganisms during advanced diagenesis. *Chemical Geology*, 437: 98-108.

Arvidson R. E., Squyres S. W., Bell J. F., Catalano J. G., Clark B. C., Crumpler L. S., De Souza P. A., Fairén A. G., Farrand W. H., Fox V. K. and others. (2014) Ancient aqueous environments at endeavour crater, mars. *Science*, 343: 1248097-1248097.

Azov V. A., Mueller L., and Makarov A. A. (2020) Laser ionization mass spectrometry at 55: Quo Vadis? *Mass spectrometry reviews*.

Barghoorn E. S., and Tyler S. A. (1965) Microorganisms from the gunflint chert. *Science*, 147: 563-575.

Brasier M. D., Antcliffe J., Saunders M., Wacey D., and Jablonski D. (2014) Changing the picture of Earth's earliest fossils (3.5–1.9 Ga) with new approaches and new discoveries. *PNAS*, 112: 4859-4864.

Calabrese E., and Fowler W. B. (1978) Electronic energy-band structure of α quartz. *Physical Review B*, 18: 2888-2896.

Carrier B. L., and Kounaves S. P. (2015) The origins of perchlorate in the Martian soil. *Geophysical Research Letters*, 42: 3739-3745.

Cedeño López A., Grimaudo V., Moreno-García P., Riedo A., Tulej M., Wiesendanger R., Wurz P., and Broekmann P. (2018) Towards femtosecond laser ablation ionization mass spectrometric approaches for chemical depth-profiling analysis of lead-free Sn solder bumps with minimized side-wall contributions. *Journal of Analytical Atomic Spectrometry*, 33: 283-293.

Chelikowsky J. R., and Schlüter M. (1977) Electron states in α-quartz: A self-consistent pseudopotential calculation. *Physical Review B*, 15: 4020-4029.

Ehlmann B. L., Mustard J. F., Murchie S. L., Bibring J. P., Meunier A., Fraeman A. A., and Langevin Y. (2011) Subsurface water and clay mineral formation during the early history of Mars. Nature Publishing Group, pp 53-60.

Fairén A. G., Davila A. F., Lim D., Bramall N., Bonaccorsi R., Zavaleta J., Uceda E. R., Stoker C., Wierzchos J., Dohm J. M. and others. (2010) Astrobiology through the ages of Mars: The study of terrestrial analogues to understand the habitability of Mars. *Astrobiology*, 10: 821-843.

Fralick P., Planavsky N., Burton J., Jarvis I., Addison W. D. D., Barrett T. J. J., and Brumpton G. R. R. (2017) Geochemistry of Paleoproterozoic Gunflint Formation carbonate: Implications for hydrosphere-atmosphere evolution. *Precambrian Research*, 290: 126-146.

Grimaudo V., Moreno-García P., Riedo A., López A. C., Tulej M., Wiesendanger R., Wurz P., and Broekmann P. (2019) Review—Laser Ablation Ionization Mass Spectrometry (LIMS) for Analysis of Electrodeposited Cu Interconnects. *Journal of The Electrochemical Society*, 166: D3190-D3199.

Grimaudo V., Tulej M., Riedo A., Lukmanov R., Ligterink N. F. W., de Koning C., and Wurz P. (2020) UV post-ionization laser ablation ionization mass spectrometry for improved nm-depth profiling resolution on Cr/Ni reference standard. *Rapid Communications in Mass Spectrometry*, 34: e8803.

Grotzinger J. P., Sumner D. Y., Kah L. C., Stack K., Gupta S., Edgar L., Rubin D., Lewis K., Schieber J., Mangold N. and others. (2014) A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale crater, Mars. *Science*, 343: 1242777-1242777.

Hassler D. M., Zeitlin C., Wimmer-Schweingruber R. F., Ehresmann B., Rafkin S., Eigenbrode J. L., Brinza D. E., Weigle G., Böttcher S., Böhm E. and others. (2014) Mars' Surface Radiation Environment Measured with the Mars Science Laboratory's Curiosity Rover. *Science*, 343: 1244797-1244797.

Horneck G., Walter N., Westall F., Grenfell J. L., Martin W. F., Gomez F., Leuko S., Lee N., Onofri S., Tsiganis K. and others. (2016) AstRoMap European Astrobiology Roadmap. *Astrobiology*, 16: 201-43.

Knight A. K., Cremers D. A., Ferris M. J., and Scherbarth N. L. (2000) Characterization of Laser-Induced Breakdown Spectroscopy (LIBS) for Application to Space Exploration. *Applied Spectroscopy*, 54: 331-340.

Lepeshkin S. V., Baturin V. S., Uspenskii Y. A., and Oganov A. R. (2019) Method for Simultaneous Prediction of Atomic Structure and Stability of Nanoclusters in a Wide Area of Compositions. *The Journal of Physical Chemistry Letters*, 10: 102-106.

Ligterink N. F. W., Grimaudo V., Moreno-García P., Lukmanov R., Tulej M., Leya I., Lindner R., Wurz P., Cockell C. S., Ehrenfreund P. and others. (2020) ORIGIN: a novel and compact Laser Desorption – Mass Spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces. *Scientific Reports*, 10: 9641.

Lozano R. P., and Rossi C. (2012) Exceptional preservation of Mn-oxidizing microbes in cave stromatolites. *Sedimentary Geology*, 255-256: 42-55.

McMahon S., Bosak T., Grotzinger J. P., Milliken R. E., Summons R. E., Daye M., Newman S. A., Fraeman A., Williford K. H., and Briggs D. E. G. (2018) A Field Guide to Finding Fossils on Mars. John Wiley & Sons, Ltd, pp 1012-1040.

Mofett J. W. (1994) The relationship between cerium and manganese oxidation in the marine environment. *Limnology and Oceanography*, 39: 1309-1318.

Moreno-García P., Grimaudo V., Riedo A., Tulej M., Wurz P., and Broekmann P. (2016) Towards matrix-free femtosecond-laser desorption mass spectrometry for in situ space research. *Rapid Communications in Mass Spectrometry*, 30: 1031-1036.

Navarro-González R., Navarro K. F., de la Rosa J., Iñiguez E., Molina P., Miranda L. D., Morales P., Cienfuegos E., Coll P., Raulin F. and others. (2006) The limitations on organic detection in Mars-like soils by thermal volatilization-gas chromatography-MS and their implications for the Viking results. *Proceedings of the National Academy of Sciences of the United States of America*, 103: 16089-16094.

Neuland M. B., Grimaudo V., Mezger K., Moreno-García P., Riedo A., Tulej M., and Wurz P. (2016) Quantitative measurement of the chemical composition of geological standards with a miniature laser ablation/ionization mass spectrometer designed for in situ application in space research. *Measurement Science and Technology*, 27: 035904-035904.

Onstott T. C., Ehlmann B. L., Sapers H., Coleman M., Ivarsson M., Marlow J. J., Neubeck A., and Niles P. (2019) Paleo-Rock-Hosted Life on Earth and the Search on Mars: A Review and Strategy for Exploration. *Astrobiology*, 19: 1230-1262.

Orosei R., Lauro S. E., Pettinelli E., Cicchetti A., Coradini M., Cosciotti B., Di Paolo F., Flamini E., Mattei E., Pajola M. and others. (2018) Radar evidence of subglacial liquid water on Mars. *Science*, 361: 490-493.

Riedo A., Grimaudo V., Cedeño López A., Tulej M., Wurz P., and Broekmann P. (2019) Novel 2D binning approach for advanced LIMS depth profiling analysis. *Journal of Analytical Atomic Spectrometry*, 34: 1564-1570.

Riedo A., Grimaudo V., Moreno-García P., Neuland M. B., Tulej M., Broekmann P., and Wurz P. (2016) Laser ablation/ionisation mass spectrometry: Sensitive and quantitative chemical depth profiling of solid materials. *Chimia*, 70: 268-273.

Riedo A., Koning C. d., Stevens A., Cockell C., McDonald A., López A. C., Grimaudo V., Tulej M., Wurz P., and Ehrenfreund P. (2020) The Detection of Elemental Signatures of Microbes in Martian Mudstone Analogs Using High Spatial Resolution Laser Ablation Ionization Mass Spectrometry. *Astrobiology*.

Riedo A., Neuland M., Meyer S., Tulej M., and Wurz P. (2013) Coupling of LMS with a fs-laser ablation ion source: Elemental and isotope composition measurements. *Journal of Analytical Atomic Spectrometry*, 28: 1256-1269.

Riedo A., Tulej M., Rohner U., and Wurz P. (2017) High-speed microstrip multi-anode multichannel plate detector system. *Review of Scientific Instruments*, 88: 045114-045114.

Shcolnick S., and Keren N. (2006) Metal homeostasis in cyanobacteria and chloroplasts. Balancing benefits and risks to the photosynthetic apparatus. American Society of Plant Biologists, pp 805-810.

Shimizu H., and Masuda A. (1977) Cerium in chert as an indication of marine environment of its formation. *Nature*, 266: 346-348.

Stamenković V., Beegle L. W., Zacny K., Arumugam D. D., Baglioni P., Barba N., Baross J., Bell M. S., Bhartia R., Blank J. G. and others. (2019) The next frontier for planetary and human exploration. *Nature Astronomy*, 3: 116-120.

Tulej M., Neubeck A., Ivarsson M., Riedo A., Neuland M. B., Meyer S., and Wurz P. (2015) Chemical Composition of Micrometer-Sized Filaments in an Aragonite Host by a Miniature Laser Ablation/Ionization Mass Spectrometer. *Astrobiology*, 15: 669-682.

Tulej M., Neubeck A., Riedo A., Lukmanov R., Grimaudo V., Ligterink N. F. W., Ivarsson M., Bach W., de Koning C., and Wurz P. (2020) Isotope abundance ratio measurements using femtosecond laser ablation ionization mass spectrometry. *Journal of Mass Spectrometry*, 55: e4660.

Wacey D., McLoughlin N., Kilburn M. R., Saunders M., Cliff J. B., Kong C., Barley M. E., and Brasier M. D. (2013) Nanoscale analysis of pyritized microfossils reveals differential heterotrophic consumption in the ~1.9-Ga Gunflint chert. *PNAS*, 110: 1053-1060.

Wacey D., Sirantoine E., Saunders M., and Strother P. (2019) 1 billion-year-old cell contents preserved in monazite and xenotime. *Scientific Reports*, 9: 9068-9068.

Wiesendanger R., Tulej M., Grimaudo V., Cedeño-López A., Lukmanov R., Riedo A., and Wurz P. (2019) A method for improvement of mass resolution and isotope accuracy for laser ablation time-of-flight mass spectrometers. *Journal of Chemometrics*, 33: 1-10.

Wiesendanger R., Tulej M., Riedo A., Frey S., Shea H., and Wurz P. (2017) Improved detection sensitivity for heavy trace elements using a miniature laser ablation ionisation mass spectrometer. *Journal of Analytical Atomic Spectrometry*, 32: 2182-2188.

Wiesendanger R., Wacey D., Tulej M., Neubeck A., Ivarsson M., Grimaudo V., Moreno-García P., Cedeño-López A., Riedo A., Wurz P. and others. (2018) Chemical and Optical Identification of Micrometer-Sized 1.9 Billion-Year-Old Fossils by Combining a Miniature Laser Ablation Ionization Mass Spectrometry System with an Optical Microscope. *Astrobiology*, 18: 1071-1080.

Wurz P., Abplanalp D., Tulej M., Iakovleva M., Fernandes V. A., Chumikov A., and Managadze G. G. (2012) Mass spectrometric analysis in planetary science: Investigation of the surface and the atmosphere. *Solar System Research*, 46: 408-422.

Zhang B., He M., Hang W., and Huang B. (2013) Minimizing Matrix Effect by Femtosecond Laser Ablation and Ionization in Elemental Determination. *Analytical Chemistry*, 85: 4507-4511.

3.1.2 Determination of the microscopic mineralogy of inclusion in an amygdaloidal pillow basalt by fs-LIMS - (JAAS)

Marek Tulej[1], Rustam Lukmanov[1], Valentine Grimaudo[1], Andreas Riedo[1], Coenrad de Koning[1], Niels F. W. Ligterink[1], Anna Neubeck[2], c Magnus Ivarsson[2], Sean McMahon[3] and Peter Wurz[1]

[1]Space Research & Planetary Sciences (WP), Physics Institute, University of Bern, Bern, Switzerland;
[2]Department of Earth Sciences, Uppsala University, Uppsala, Sweden;
[3]School of Physics and Astronomy, University of Edinburgh, Edinburgh, UK;

JAAS

Journal of Analytical Atomic Spectrometry

ROYAL SOCIETY
OF CHEMISTRY
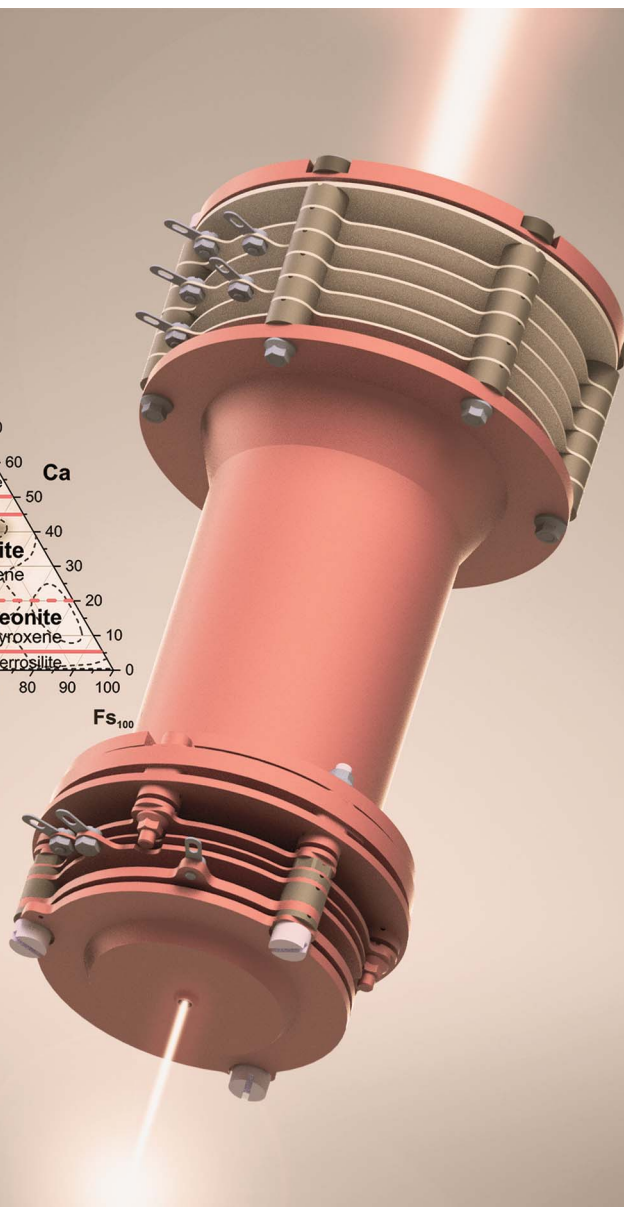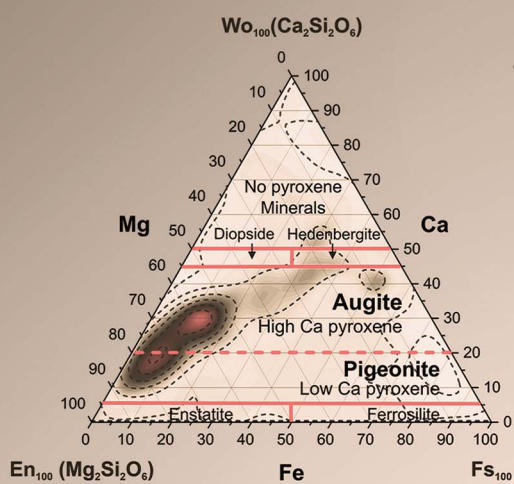
PAPER
Marek Tulej *et al.*
Determination of the microscopic mineralogy of inclusion
in an amygdaloidal pillow basalt by fs-LIMS

93

Check for updates

Determination of the microscopic mineralogy of inclusion in an amygdaloidal pillow basalt by fs-LIMS†

Marek Tulej, [ID] *[a] Rustam Lukmanov,[a] Valentine Grimaudo,[a] Andreas Riedo, [ID] [a]
Coenrad de Koning, [ID] [a] Niels F. W. Ligterink,[b] Anna Neubeck, [ID] [c] Magnus Ivarsson,[d]
Sean McMahon [ID] [e] and Peter Wurz [ID] [a]

We present chemical depth profiling studies on mineralogical inclusions embedded in amygdale calcium carbonate by our Laser Ablation Ionisation Mass Spectrometer designed for *in situ* space research. An IR femtosecond laser ablation is employed to generate ions that are recorded by a miniature time-of-flight mass spectrometer. The mass spectra were measured at several locations on the sample surface and yield chemical depth profiles along the depth length of about 30 μm. The presence of oxides and sulphides within inclusion material allows us to derive elemental abundance calibration factors (relative sensitivity coefficients, RSCs) for major and minor elements. These are obtained from the atomic intensity correlations performed on the depth profiling data. With the RSCs corrections the quantitative analysis of more complex mineralogical phases within the inclusion is conducted by correlating atomic abundance fractions in ternary diagrams, typically used in geology. The spatial resolution of the depth profiles was sufficient to study chemically distinct micrometre-sized objects, such as mineralogical grains and thin layers of minerals including micrometre-sized filamentous structures. The method presented here is well-suited for the quantitative chemical analyses of highly heterogeneous materials where the ablation condition can vary locally with the material composition making the application of standard reference materials less accurate. The presented method is developed to distinguish between abiotic and biological material while searching for micrometre-sized extinct or extent life forms on the surfaces of Solar System bodies.

aSpace Research and Planetary Sciences, Physics Institute, University of Bern, Switzerland. E-mail: marek.tulej@space.unibe.ch

bCentre for Space and Habitability, University Bern, Switzerland

cDepartment of Geological Sciences, Uppsala University, Sweden

dDepartment of Paleobiology, Swedish Museum of Natural History, Sweden

eSchool of Physics and Astronomy, University of Edinburgh, Room 1508, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, EH9 3FE, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0ja00390e

Introduction

One of the major objectives of current space research and planetology is the search for signatures of extant and extinct life on other planetary bodies. With the assumption that life on other planets should be based on the same physical and chemical rules as life we know from Earth and evolved along similar pathways, these searches should follow known terrestrial practice. Ancient extinct life can be identified by finding remains of organic material, stromatolitic layers, and individual microfossils within planetary rocks and soils.[1] Identification of metabolic products together with the deciphering of

fossilization mechanisms can be essential in such searches. Among available analytical instrumentation, only those instruments, which can deliver selective and sensitive chemical measurements of microscopic structures can be employed for this task.[2,3]

For conducting analyses on planetary bodies only small, light, low-powered and sufficiently robust instrumentation can be used. Several space-borne instruments including Laser-Induced Breakdown Spectroscopy (LIBS), X-Ray Fluorescence (XRF), and Laser Raman Spectroscopy have proved their capabilities in delivering composition measurements of the sample surface and important biosignatures in numerous performance studies. However, their measurement sensitivity might not be sufficient to investigate the chemical composition of microscopic structures of sparse life.[2,4] An example of the instrument combining Raman, LIBS and IR techniques is the SuperCam instrument designed on the NASA Mars2020 with the performance figures of each of applied technique described in recent publication and references therein.[5,6] Compared to these techniques, laser-based mass spectrometry adds complementary information. With current LIMS spatial resolution (lateral and vertical) and sensitivity, detailed analyses of grain-size objects

80 | *J. Anal. At. Spectrom.*, 2021, **36**, 80–91

This journal is © The Royal Society of Chemistry 2021

can be conducted with the determination of context mineralogy and detection of micro-sized fossils.[7–9]

Laser ablation ionisation mass spectrometry (LIMS) based on time-of-flight mass analysers is a well-known technique with its origin in the mid-1970s. In the beginning, the technique suffered from various drawbacks, which mostly were affecting the quantification and the measurement procedure. With the continuous advances in electronics, vacuum, and laser systems, this technology has re-emerged for laboratory and field application for chemical analysis of solid-state samples.[10–12] Over the years, a high degree of miniaturisation of the LIMS instrument has been achieved with the objective to become a space-borne instrument.[8,13–15] Robustness of the time-of-flight mass analyser, flexible operation of the laser ablation ion source, the control over the ion optics settings and detector gain allow direct and sensitive chemical analysis of raw samples providing efficient ion production and nearly 100% ion transmission from the measurement spot to ion detector.[15,16] For the analysis of micrometre-sized objects such as grains or microfossils, the integration of a microscope camera system into the system helps in an initial characterisation of micro-sized objects and improving overlap of the laser spot with the object of interest.[8,17]

LIMS with an fs laser ablation ion source shows an increasingly higher performance. It delivers improved quantitative chemical composition information with considerably reduced element fractionation.[18–20] Several recent studies show that LIMS can be considered as a standardless technique delivering semi-quantitative results.[18–22] Further improvements to the quantitative analysis can be achieved by adding to atomic abundance derived from the single-charged atomic mass peaks also multiple ionised atomic mass peaks.[23] An improved efficiency of the atomic ion production and quantification of atomic ions can be achieved applying a double pulse laser ablation/ionisation ion source.[24] In double pulse system, the first pulse is ablating the sample surface and the second, shortly delayed pulse interacts with freshly formed plasma plume resulting in additional plasma heating, atomisation and ionisation. These processes influence also post-plasma chemistry efficiency. The presence of isobaric multiple-charged atomic ions and clusters in ion beam can introduce uncertainty in the quantification of the atomic signal measurement; hence, by applying double pulse laser ion source, the isobaric interferences can be minimized. It has been demonstrating also that the isobaric interferences due to clusters could be reduced by implementing collisional cell which allows partial temporal separation of the cluster and atomic ion distributions.[25] High-resolution time-of-flight mass spectrometric measurements of atomic ions produced in laser ablation are typically difficult due to wide kinetic energy distributions of produced ions. Only, recently a high-resolution ($M/\Delta M \sim 10^4$) laboratory LIMS system with a double reflectron system was developed to resolve most of mass peaks of cluster ions, hence, the isobaric interferences resulting from polyatomic species can be avoided.[26] In miniature LIMS systems designed for *in situ* applications not all these implementations can be made. Here, we propose alternative method of the abundance calibration. The applied method uses depth-profiling data and requires presence of simple chemical compounds such as oxides and sulphides within the sample material.

In recent years, fs-LIMS proved its capabilities for chemical analysis with high spatial (lateral, vertical) resolution. High lateral resolution is achieved by focusing laser radiation onto the surface to a spot size of a few micrometres and by tuning over the pulse energy, ablation rate and thickness of the ablated layer can be well controlled so that the measurements can be conducted with the depth resolution down to a few nanometres.[27–32] In our recent studies on the Ni/Cr NIST standard, we have accomplished a depth resolution of about 30 nm applying UV femtosecond laser radiation arranged in a double pulse configuration.[24,32] 3D chemical mapping analysis with LIMS were recently presented.[33,34]

There are other laboratory analytical techniques, which offer high spatial resolution depth profiling capabilities such as Glow Discharge Mass Spectrometry (GD-MS), Laser Ablation Inductively Coupled Mass Spectrometry (LA-ICP-MS) and Secondary Ion Mass Spectrometry (SIMS).[35–40] These instruments in current state of development are less suitable for application to space research but can be useful in applications in terrestrial environments. Current LIMS systems fulfil the requirements of size, weight or power consumption necessary available aboard a spacecraft, either a rover or a lander.[7]

In the current contribution, we introduce a procedure for the *in situ* atomic abundance calibration using micrometre-sized inclusion embedded in a calcium carbonate host mineral. The procedure uses the depth profiling data *e.g.* the measurements of atomic intensities in a function of ablation layer. By correlating the intensities of atomic species measured at specific ablation layers and observation of linear correlation, we can deduce the presence of the molecular compounds and determine corrections to the measured atomic intensities. These corrections are obtained by comparison of the measured slopes of the correlations and the expected values from the actual chemical formula. From the measurements of various oxides and sulphides we obtain the correction coefficients for most of major and minor elements. Determination of these corrections on simple compounds allows for a more detailed analysis of complex mineralogical phases present in the sample at microscopic scales. Having the obtained corrected elemental abundances, atomic abundance fractions are plotted in ternary diagrams allowing the investigation of more complex mineralogical phases. A similar procedure which is known as LRS-linear regression slope method was applied in earlier MC-ICP-MS studies to improve accuracy and precision of isotope ratios.[41–43] This data reduction protocol used the simultaneous responses of all isotopes measured. In the current study, the simultaneous response of elements is expected while ablating specific molecular compound.

Similar materials to that studied here are found on Mars and can be considered as a distinct astropaleontological target. The calcium carbonate phase in basalts can host veins and voids, which are recognised as habitats for endolithic microorganisms and form secondary infilling mineralisation in volcanic environments.[44–47] Thus, extended characterisation of these materials and understanding their heterogeneities are important to account for their biological or abiotic origin.

Experimental

Materials

The sample of an amygdaloidal pillow basalt from Kinghorn, Fife, Scotland[48] was processed to a 30 μm thick thin-section, mounted on a stainless-steel sample holder and introduced into the vacuum chamber for its mass spectrometric analysis (Fig. 1). The left panel of Fig. 1 shows the thin-section of basalt with an inclusion of calcium carbonate. The right panel shows a close-up image of the dark inclusion, which contains a number of needle-shaped microscopic structures. The enclosure into the calcium carbonate host happened some 360–320 Ma ago.[48]

The LIMS instrument

The mass spectrometric investigations were conducted with our miniature laser ablation ionisation mass spectrometer designed for *in situ* space research applications. The instrument combines a femtosecond-laser ablation ion source with a miniature time-of-flight mass analyser.[22] Current studies are conducted with a NIR-fs-laser radiation (pulse width, Δt ~190 fs, λ = 775 nm, laser pulse repetition rate of 1 kHz) focused to a circular spot of about 10 μm in diameter. In all measurements, a pulse energy of about 8 μJ on the sample surface was applied. The resulting craters were inspected with an optical microscope.[17] The detailed crater morphology analyses can be found in our previous publications.[31] The measurements were computer-controlled using our custom-designed operating system.[15] The sample was attached to the *xyz*-micro-translation stages. The stage allows for about 2 μm positioning accuracy of the location of interest on the sample by using prior measurements size calibration with an *in situ* microscope camera system.[17]

Mass spectrometric measurements

Mass spectra were collected at ten locations on the sample, starting from the pure calcium carbonate mineral, across the interface of the calcium carbonate host all the way into the dark

Fig. 1 (a) The sample prior to the introduction into the vacuum chamber. The sample is placed within milled depression in the sample holder made of stainless steel and is attached by a copper tape. (b) Close up microscope image of the amygdale calcium carbonate sample with a dark inclusion in the centre. A network of needle-like structures spreading out from a denser, darker central part of the inclusion can be readily identified. The ten black spots with numbers indicate the location at which laser ablation mass spectrometric analyses were conducted.

inclusion, and on the inclusion's surface seen as a dark area in the centre of Fig. 1b. The dark inclusion consists of needle-shaped micrometre-sized structures that are embedded in complex mineralogical phases other than calcium carbonate. For each location 300 mass spectra were acquired corresponding to 300 ablated layers. The individual layer spectrum is measured by accumulation mass spectra of 200 laser shots on board of the acquisition system. Additionally, mass spectra on the calcium carbonate host were collected by drilling through the 30 μm thin-section, which corresponds to about 366 layers. The mass calibration of the spectra, mass peak integration and principles of depth profiling analysis were discussed in detail in our previous publications.[49,50]

Relative sensitivity coefficients

The Relative Sensitivity Coefficient (RSC) for an element in the sample is obtained by dividing measured element concentration, X_m by the expected concentration, X:

$$RSC_X = \frac{X_m}{X} \tag{1}$$

The LIMS instrument delivers measurement of atomic intensities. These intensities are calculated by integrating mass peaks of this element in the mass spectrum. Thus, the atomic concentrations are presented as the atomic fractions. To derive the RSC of an element, one compares the measured and expected atomic fractions:

$$RSC_X = \frac{A_X/A_{tot}}{P_X} \tag{2}$$

where A_X stands for the atomic ion intensity obtained from the mass spectrum and A_{tot} is the sum of all atomic intensities measured in the sample. Hence, A_X/A_{tot} is the measured fraction of given element in the sample. P_X denotes the stated fraction of the X atoms to all atoms in the sample. When the sample is a pure mineralogical compound such as calcium carbonate, $CaCO_3$, the atomic fractions of P_{Ca}, P_C, P_O are 0.2, 0.2 and 0.6, respectively.

Knowing concentration ratio between elements X_i and reference element Y in the sample, we can derive another formula for calculating RSCs:

$$RSC_{X_i} = RSC_Y \frac{X_{mi}}{Y_m} \frac{Y}{X_i} \tag{3}$$

where RSC_Y is the RSC of reference element Y, X_{mi}/Y_m is the ratio of the measured atomic concentrations (slope in the linear atomic intensity correlation), Y/X is the stated ratio of element Y to element X_i (see ESI, Section A for more details†).

Using depth profiling data and the signal intensity correlation, we can identify elements, which form simple chemical compounds even they are distributed within sample volume irregularly. In addition to their intensity relationship, the linear intensity correlation of the measuring signals also yields the location where the signals are measured in the sample *i.e.*, where these minerals are present in the sample. Hence, we can identify these compounds at specific ablation layers. The determination
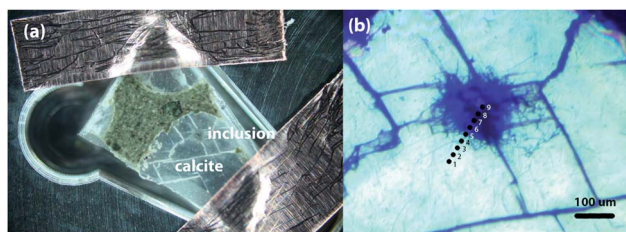
of RSCs from the signal intensity correlation between species in a simple chemical compound can be made this way, *e.g.*, two or three-element compounds such as oxides or sulphides present in the sample as layers or grains can be made this way. For the correlation, we select the atomic signals of a compound, which originate from various ablation layers. Atomic signals emerging from the same compound will always create a linear correlation, since the element ratio will be constant in all different layers, independently of the amount of the compound. The slope of a linearly correlated data set is then the effective abundance ratio of the two correlating atomic species.

Results and discussion

Mass spectra of the calcium carbonate phase and the dark inclusion

Fig. 2a shows the mass spectrum measured on the pure calcium carbonate with C, O, and Ca as the major elements that are characteristic for the calcium carbonate mineral composition, $CaCO_3$. From purely mass spectrometric perspective, it is hard to identify polymorphic modifications. The calcium carbonate could be calcite, aragonite, or vaterite, but earlier it was determined to be calcite.[7] Additionally, we also observe mass peaks of the elements Na and K. The dynamic range of the spectrum is close to 10^4, which allowed for simultaneous detection of the major elements and trace elements with a bulk concentration of at least 100 ppm (atomic fractions). In addition to C, O and Ca

also Li, B, F, Mg, S, Cl, and Sr elements are identified in the spectrum. These elements are known impurities in the calcium carbonate minerals. Other mass peaks within the low atomic mass range are double- and triple-charged ions of C, O, and Ca. Non-equilibrium chemistry in the expanding and cooling ablation plume leads to the formation of hydrocarbons, C_xH_y, and oxides, Ca_xO_y (where $x = 1–3$ and $y = 1–4$), which ions are also observed in the mass spectrum. Similar mass spectra are recorded at locations 1–4 (see also Fig. 1b), indicating chemical homogeneity of the host mineral.

The mass spectrum displayed in Fig. 2b was measured on the inclusion at location 9 (see also Fig. 1). The major and minor elements observed are O, Na, K, Fe and C, Mg, Al, Si, S, Ca, Ni, Co, respectively, including also the trace elements, Li, N, P, Ti, V, Cr, Cu, Zn and Sr. The dynamic range in the spectra is larger than 10^4. In addition, multiple charged ions, hydrocarbons and oxides of the main constituents C, O, Ca, and Fe can be identified in the mass spectra. The resemblance of the mass spectra at location 9 and at locations 6, 7 and 8 is large, indicating negligible chemical heterogeneity in the dark inclusion. However, the depth profiling analysis shows that the inclusion material is heterogeneous.

Chemical depth profiling on calcium carbonate mineral

Fig. 3a shows the chemical depth profiles measured for the elements C, O, Na, and Si along the 30 micrometre-thick
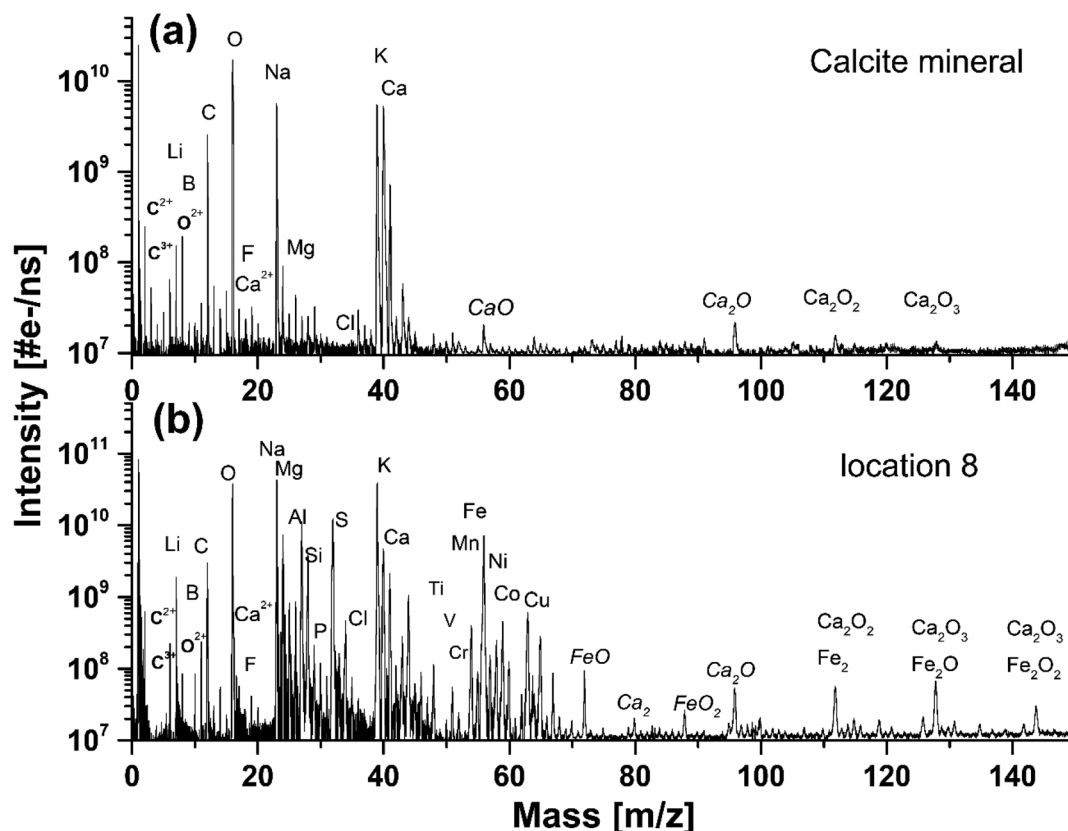

Fig. 2   (a) The mass spectrum recorded on a pure calcium carbonate mineral on location 4, (b) and on the inclusion on location 8.
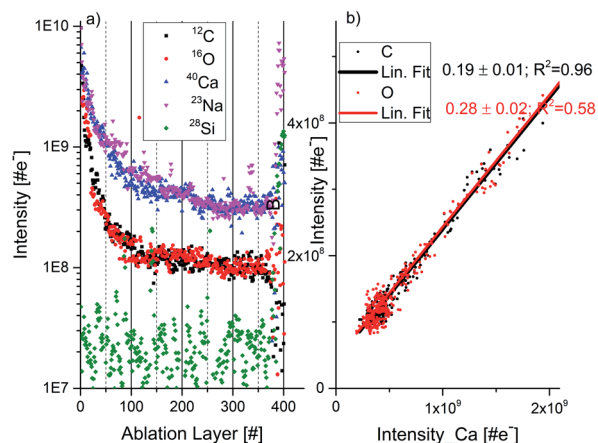
Fig. 3 (a) Depth profile of C, O, Ca, Na, and Si as a function of the ablation number. After 366 ablation layers we observe and increase of the Si and Na atomic intensities indicating transition to soda lime silica glass substrate material. (b) Correlation between C and Ca, and O and Ca atomic intensities, respectively, derived from the depth profiling data.

calcium carbonate sample as a function of the ablation layer number. The measured mass peak intensities are observed to decrease fast for about first 30 ablation layers and for the remaining ablation layers – approximately monotonically with the ablation layer number. Because of the mass peak shape distortions and peak intensity fluctuations in the spectra recorded during an initial ablation phase (here for the first 30 ablation layers) these spectra are removed from the analysis. Above the ablation layer number 366 we observe a sudden signal increase for all elements, including Na and Si. The Na and Si signals show the transition to the soda lime glass substrate, which indicates that the laser has drilled a crater through approximately 30 μm thick thin-section sample. Approximately circular craters with a diameter of about ø10 μm were produced as observed with the internal optical microscope.[17] More advanced analyses of the crater characteristic were not conducted in the current study.

Mass peak intensity correlation of C, O and Ca and RSC values

Calcium carbonate mineral. The mass spectrometric analysis of the composite spectrum (see Fig. 2a) recorded on pure calcium carbonate yield C, O and Ca atomic intensities from which we can calculate atomic fractions of each element and compare these results to the expected atomic fraction in $CaCO_3$ using formula (2). These calculation yields RSCs for C, O and Ca of 0.55 ± 0.03, 0.36 ± 0.05, and 3.3, respectively (see also ESI, section A, Table 1 for more details†). The intensity correlation plots of C and Ca and O and Ca, derived from individual mass spectra measured along the ablation depth are displayed in Fig. 3b. The mass peak intensities correlated in plots are calculated from the peak areas. We observe linear correlations for these data and the slope coefficient as shown in Fig. 3b together with the R^2 correlation coefficient. The slope coefficients derived from the linear fits are 0.19 ± 0.01 and $0.28 \pm$

0.02 for the correlation of C with Ca and of O with Ca, respectively. Considering the expected atomic fractions of the elements in the pure calcium carbonate mineral, we derive relative sensitivity coefficients, RSCs from the formula (3) by taking RSC_{Ca} determined using the first method as reference. The RSCs values obtained with this method are 0.63 ± 0.02, and 0.30 ± 0.03 for C, O, respectively (see also ESI, Section A, Table 2†) indicating consistency of the RSCs values determined by the two methods.

Inclusion material, heterogeneities

Fig. 4 compares the O depth profiles in a function of ablation layer number recorded at various locations on the inclusion including also location at the interface between pure calcium carbonate and inclusion material (Fig. 4, panel (a), see also location 5 in Fig. 1b). Similar depth profiles are observed for C and Ca although some differences at specific ablation layers in the profile shapes can be noticed. An increase of mass peak intensities at specific ablation layer form characteristic peaks on the depth-profiling curve. A systematic shift of these peaks towards lower ablation layer numbers is observed while the measurement locations become closer to the centre of the inclusion, as indicated by the arrows in Fig. 4(b–d). This shift is likely due to measurement of the same assembly of dark needles embedded in the matrix with a certain angle to the surface normal. The closer one gets to the centre of the dark inclusion, the earlier one measures the needle assembly because the darker area is closer to the surface and the needles were growing in opposite direction, along the direction of depth of the thin section. Fig. 4 (panels (e) to (h)) displays the mass peak intensity correlations of C with Ca, and O with Ca, measured at locations from 5 to 8. At location 5, the mass peak intensities correlate linearly well and only a small group of data points lay outside the correlation line. Contrary, at location 6, the linear correlation is observed only for a small part of data points (Fig. 4, panel (f)). At locations 7 and 8, no obvious linear correlation can be observed; the data points form a diffuse distribution with a large intensity spread. The data points in the depth profiles measured in locations 5 and 6 (Fig. 4, panels (a) and (b), respectively), for which the calcium carbonate mineral was identified via the linear correlations (red and black data points on the linear regression in panels (e) and (f)) are indicated as green circles in panels (a) and (b), respectively. These are found at the ablation layer numbers between regions of increased peak intensities on the depth-profiling curve. These layers with calcium carbonate, if present, are not easy to detect at locations 7–9. By fitting the fraction of the data that correlates to the calcium carbonate mineral at location 6, one obtains slope coefficient values for O, which are larger compared to that determined at the location of a pure calcium carbonate. The slope coefficient values in correlation of C and Ca are observed to be similar to each other at these locations with the slope coefficient values of 0.41 ± 0.06 and 0.31 ± 0.02 at location 6 and 0.45 and 0.33 at location 7 for C/Ca and O/Ca correlations, respectively (Fig. 4e and f). Considering data at locations 6 we can repeat the procedure applied for pure calcium carbonate

84 | J. Anal. At. Spectrom., 2021, 36, 80–91

98
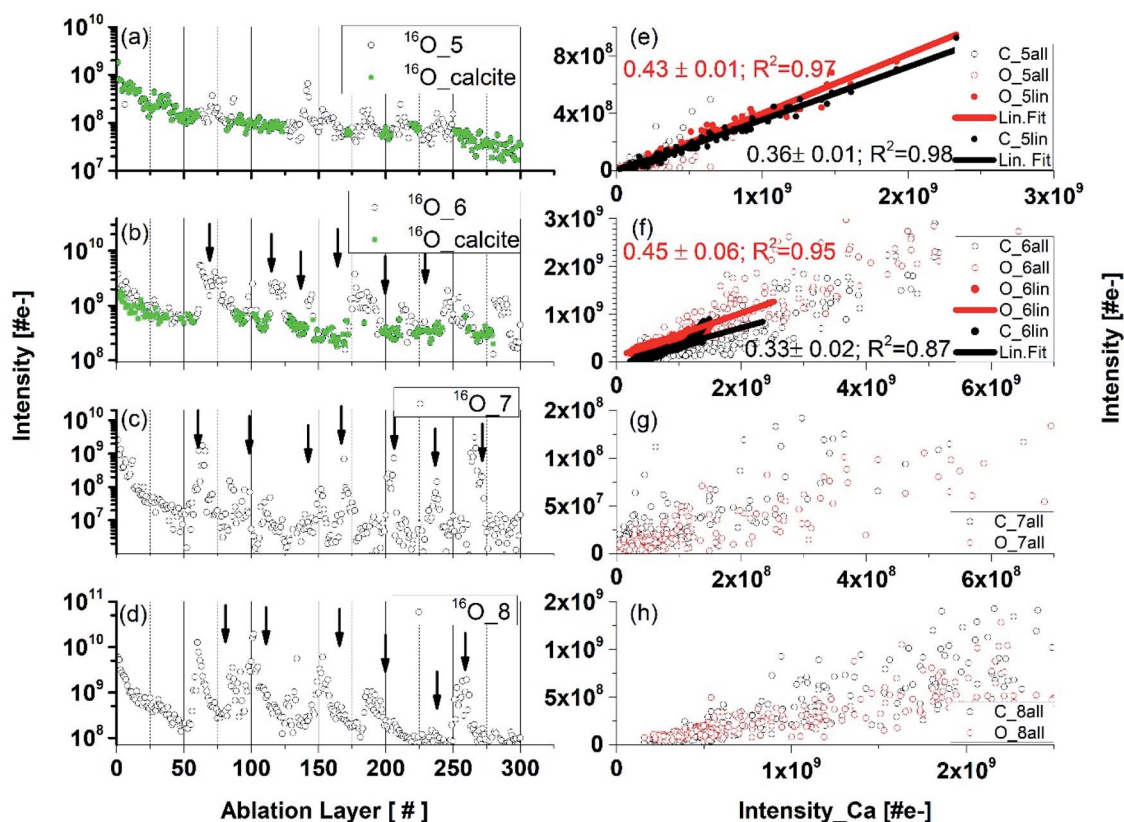
This journal is © The Royal Society of Chemistry 2021

Fig. 4 Depth profiles and mass peak correlations. Panels (a–d): Depth profiles obtained by measuring the O mass peak intensities at locations 6 panel (b), 7 panel (c), and 8 panel (d) on the dark inclusion. The green-coloured points in the left panels (a) an (b) are the data points that correspond to the calcium carbonate mineral. Right panels: the correlations of the mass peak intensities of elements C and O (y-axis) with Ca (x-axis) measured at corresponding locations. Only part of the data points correlates linearly (black and red full circles) in panels (e) and (f). There was no clear correlation between O and Ca in the data plotted in panel (g), although a C–Ca correlation can be noticed. Contrary, for a part of the data points the correlation between O and Ca is readily observed but the correlation between C and Ca is less pronounced in panels (g) and (h).

phase and derive RSC coefficients of C, O and Ca at the inclusion location (see ESI Section A, Table 3†) using the composite spectrum obtained by summing up only the contributing the data points to linear correlations. These calculations yield RSCs of C, O, and Ca of 0.91 ± 0.05, 0.44 ± 0.05 and 2.71 ± 0.04, respectively. Applying again eqn (3) with RSC of Ca equal 2.71 ± 0.04 as reference and considering slope coefficients C/Ca and O/Ca we can obtain the RSC of C and O equal to 0.97 ± 0.09 and 0.38 ± 0.05, respectively. The RSC of Ca is chosen as the reference (see Section A, Tables 3 and 4 in ESI for more details†).

Depth profiles at location 9

Fig. 5 shows depth profiles of several major and minor elements recorded at location 9. Significant mass peak intensity variations for all elements can be observed along the depth profiles at specific ablation layer numbers. Signal intensity variations indicate either a deficit or an increase of the element concentration at these specific ablation layers. Some increase of the ablation efficiency can occur because of the increased absorption of laser radiation at darker spots such as needles. In agreement with the signal intensity correlation studies at locations 7 and 8, we also observe that the Ca, C, and O signals do

not correlate linearly. Moreover, at certain depths, a depletion of Ca is observed shown as shaded areas in Fig. 5a. At the ablation layer ranges 53 : 58, 65 : 71, 128 : 132, 156 : 161, and 165 : 171, we can observe that the drop of the K intensity is accompanied by an increase of the intensities of Na, C, and Ca (Fig. 5a). Considering correlations of the minor elements, an increase of the S, Ni, Co, Mg, Mn, Cu, and B at several locations is typically accompanied by a decrease of the Fe and Ti mass peak intensities (see shaded areas in Fig. 5b and c). Only at the ablation, layer 110 all peaks intensities are observed to correlate. From the microscope images of the needle-like structures we can estimate their size to be in the range of 0.8–2 μm (Fig. 1b). The narrowest peaks in the depth profile involve 5 ablation layers (see Fig. 5). Thus, the thickness of the measured individual ablation layers is estimated to be in a range 0.2–0.4 μm.

Determination of RSCs from the data collected at locations 6 to 9

If the atomic signals were measured stoichiometric, the slope in the correlation would represent the abundance ratios of the elements of the relevant chemical formula. So far, we have

This journal is © The Royal Society of Chemistry 2021

99

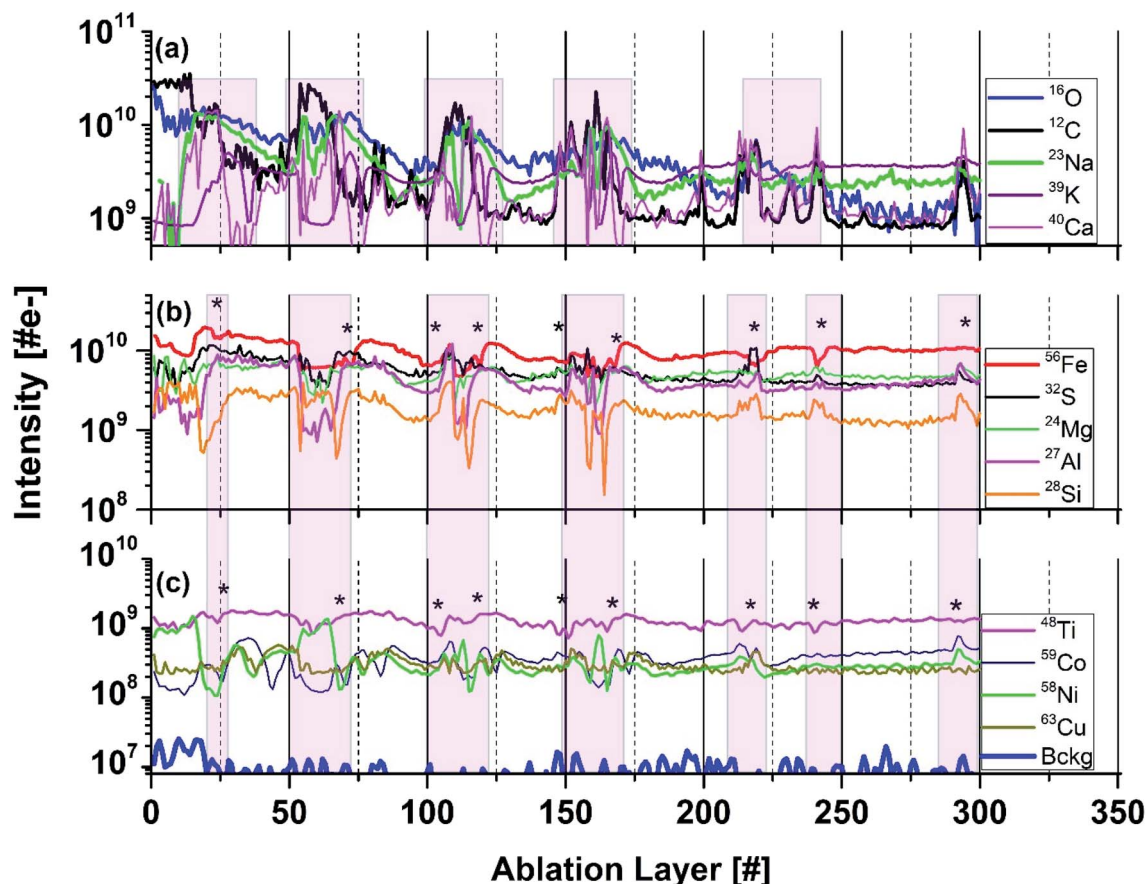J. Anal. At. Spectrom., 2021, **36**, 80–91 | **85**

Fig. 5 Variation of the intensities as function of ablation layer number for several elements (isotopes) measured on the dark inclusion at location 8 (Fig. 1b). Panel (a) shows the intensity variations of the C, O, Ca, Na, and K intensities. Panel (b) shows variations of the Fe, S, Mg, Al, and Si intensities. Panel (c) shows variations of Ti, Ni, Co, and Cu intensities. At ablation layers marked by (*) the intensities of Ni and Co correlate well with S, and anticorrelate with Fe. Shaded area highlights layer with specific element correlation, respectively anticorrelations. The traces are shifted to each other vertically for better visualisation of their shapes.

determined RSCs for the elements C and O through their abundance correlation with Ca. Based on our previous studies we know that in the inclusion material other small chemical compounds are present including oxides and sulphides.

By correlating the intensities of other elements with oxygen and sulphur intensities, we can gain the information about the RSCs for these elements providing that they form oxides and/or sulphides. The details of RSCs calculations are given in ESI, Section A Tables 5–7.† Here we briefly discussed the results obtained with the linear intensity correlation method.

In Fig. 6, the intensities of several elements are set in correlation with S (left panel) and O (right panel) intensities, respectively. For some parts of the data set, a linear correlation, either with oxygen or sulphur is observed indicating that these elements may have their origin in oxides and/or sulphides. This agrees with the results of our earlier studies, where identification of oxides and sulphides with other analytical methods were performed.[7] Typically, the fs-LIMS measurements on Standard Research Materials show that RSCs for almost all metallic elements are, generally, close to one within about 15%. An exception are Al, Ti, Cu and Mg. RSCs of non-metallic elements such as C, S, O, Si and Ca various within broader range and

depends sensitively on the ablation conditions and material properties.[19,22]

Iron is an abundant element in the inclusion. Typically, iron occurs ubiquitously in hydrothermal ore deposits in sulphides, oxides, silicates, and in carbonate minerals. In the correlation of the Fe with O intensities, we observe three different linear slopes (Fig. 6d). The coefficient values for the red and blue slope are 1.4 ± 0.1 and 3.3 ± 0.3, respectively. With RSC of O derived in the analysis of calcium carbonate on the inclusion equal to 0.38 ± 0.05, we derive RSC of Fe equal 1.25 ± 0.04 considering chemical formulas FeO for blue slope. From the slope coefficient of the red curve we can derive RSC of Fe equal to 1.05 ± 0.05 assuming that FeO_2 is present in the sample (Fig. 6 panel (d)). From these data we obtain average RSC of Fe equal to 1.15 ± 0.05. RSC value for Fe is typically close to 1 (within 15% uncertainty) as it can be determined in the analyses of various materials including metallic or rock samples.[19,22,51] The latter compound, FeO_2, however, is unlikely to be present in our sample. The most common oxidation states of Fe are +2 and +3. With these valences Fe can be found in minerals such FeO (ferric oxide), Fe_2O_3 (hematite) or Fe_3O_4 (magnetite). Most viable candidate instead can be iron(II) hydroxide or ferrous

86 | *J. Anal. At. Spectrom.*, 2021, **36**, 80–91

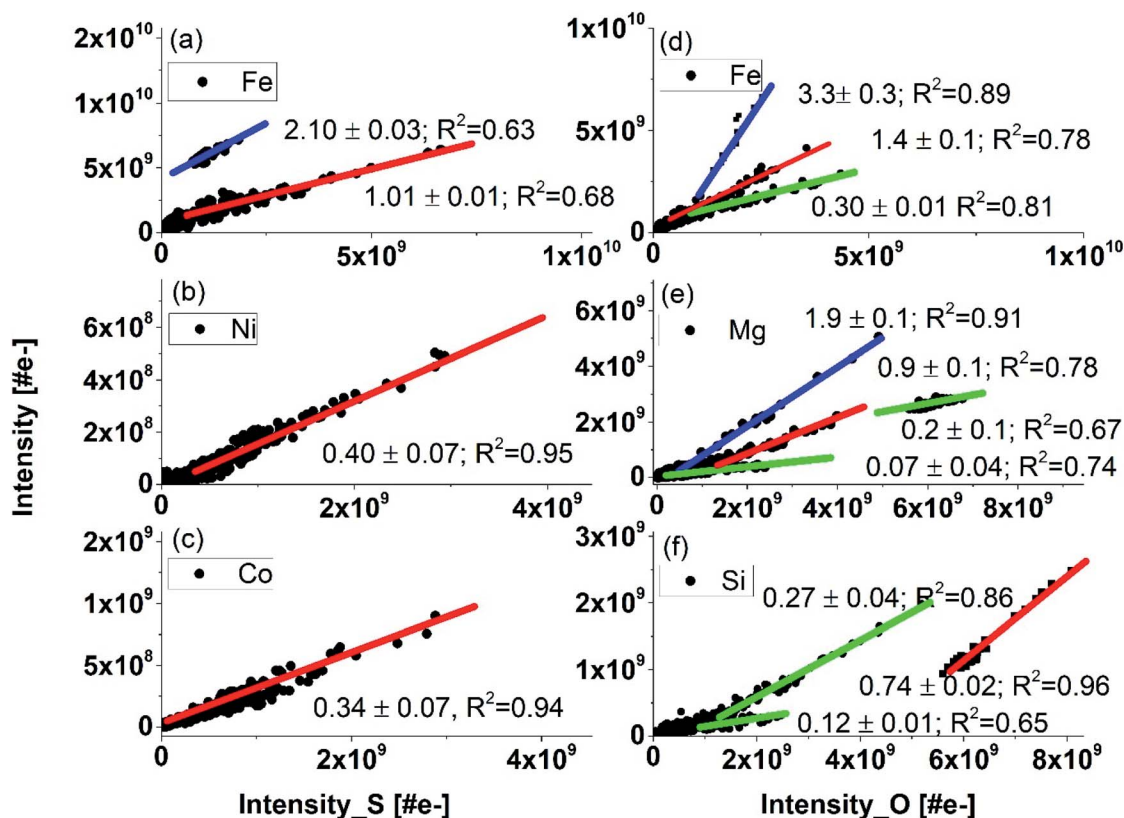This journal is © The Royal Society of Chemistry 2021

Fig. 6 The correlation of Fe, Ni, and Co with S (panels (a–c)), and the correlation of Fe, Mg, and Si with O (panels (d–f)) based on all data measure on locations 6 to 9 (see Fig. 1b). The y-axis represents the intensity of the indicated element in the legend. The slopes (blue) are given for each correlation curve for the oxides and sulphides. RSCs are determined from the blue slopes. Lower slope values than expected for sulphides and oxides indicates that the analysed element was within another compound or a more complex.

hydroxide with the formula $Fe(OH)_2$. Ferrous hydroxide is produced when iron(II) salts, *e.g.*, iron(II) sulphate is treated with hydroxide ions. This process is likely to occur in aqueous environments including hydrothermal systems.[52] The slope coefficient of the third correlation of Fe with O observed in Fig. 6c is smaller than the one assigned to the oxides. This result implies that those measured atomic signals may originate from more complex mixture of relevant chemical compounds or mineral phases. For a sample consisting of heterogeneities within a volume smaller than the sampled volume, the determination of RSC from the mass spectrum at this location would yield inaccurate results, because an unique element composition is assumed for the entire volume, thus entire data set. Using depth profiling, the mass peak intensity correlation method can be applied for each individual layer and can yield more accurate results. The specific chemical composition of each layer is preserved and the spatial resolution with the profiling depth is expected to be in sub-micrometre range. In this way an isolation between various compound can be achieved and the different mineralogical phases are recognised by means of their unique elemental composition (abundance ratio) within the various ablation layers. Hence, they can be identified from the correlation studies. Nevertheless, the lateral resolution of this study is at least order of magnitude lower and

may affect an isolation of the signals if more than one unique mineralogical grains will be ablated simultaneously.

In similar analysis conducted for the correlation of Fe with S, we observe two linearly correlated populations (Fig. 6a). Taking the RSC value of Fe as 1.15 ± 0.05, we can determine the RSC value of S. The two correlation curves give the slope coefficients 2.10 ± 0.03 (blue) and 1.01 ± 0.01 (red), respectively. Assigning the blue and red slopes as being due to the compounds FeS and FeS_2, we can derive RSC of S equal to 0.54 ± 0.06 and 0.57 ± 0.04, respectively. The FeS and FeS_2 are the most common forms of iron sulphides known in natural environments, such as mackinawite (metastable form of FeS), pyrrhotite (thermodynamically more stable form of FeS), cubic FeS, (formed in presence of foreign ions *e.g.* chloride ions), and pyrite (FeS_2). These compounds are formed frequently also in hydrothermal environments. In our previous study, no pyrite could be identified in the inclusion though this compound is the most important sulphide gangue mineral and occurs in virtually all major hydrothermal base and precious metal mineral deposits. Nevertheless, if the deposits are subjected to high-grade metamorphism, pyrrhotite (FeS) replaces pyrite. This occurs typically by dissolution of pyrite from aqueous solutions and presence of $Fe(SH)_2$ instead FeS_2 in inclusion material.[52–54]

Similar analyses of RSCs of other elements including Mg, Al, Si, and Ca are conducting assuming that these elements can be

This journal is © The Royal Society of Chemistry 2021

101

J. Anal. At. Spectrom., 2021, **36**, 80–91 | **87**

present as sulphides and oxides. For the Mg and O intensity correlation, we observe four slopes and with S one slope. The slope with the coefficient 1.9 ± 0.1 corresponds likely to MgO. The correlation slope coefficient 0.91 ± 0.06 indicates that MgO_2 could be present in the inclusion. Again, in this case similarly to other above cases, it is more likely that $Mg(OH)_2$ is present instead. The slope value in correlation Mg and S is consistent with the measurement on MgS. Three linear correlations of Si with O are observed. Quartz (SiO_2) is typically, the gangue mineral, most commonly encountered in ore-forming hydrothermal systems. Slope coefficient 0.74 in correlation with O is taken to derive the RSC of this element. It is likely that again $Si(OH)_2$ instead of SiO_2 is present in the sample. The correlation curves with smaller slope coefficients can reflect the presence of other mineralogical phases with a lower Si fraction in the compound. Similarly, for Al correlation with O, the measured slope coefficient is consistent with presence $Al(OH)_3$ rather than common compound Al_2O_3. The analysis of Mg and Ca correlation curves with S yield RSCs similar to that obtained for correlation with O or from the analysis of calcium carbonate. The correlation of Ca with S with measured slope coefficient indicates presence of $Ca(SH)_2$ and the analysis of correlation Al with S is consistent with Al_2S_3. The results of calculations of RSCs are given in Table 7.†

The slope coefficients of the Ni S, and Co and S correlation curves are 0.34 ± 0.07 and 0.4 ± 0.07, respectively, yield however much lower values of RSCs as expected for these metals (Table 7†). These coefficients are somewhat too small to yield the expected chemical formula NiS and CoS, assuming that RSCs of these elements are typically close to one.[22,51] One of the reason for these RSCs being lower than 1 can be that the compounds NiS and CoS are well mixed and effectively a (Ni, Co)S compound is sampled in our experiment. The determination of RSC for S (from Fe-sulphides) also allows to conclude that the size of the other Ni and Co sulphides inclusions are smaller than the ablated layer thickness ($\lesssim 150$ nm). The ablation layers where NiS and CoS are measured cannot be resolved to see separate compounds. The optical imaging of the inclusion area shows large density of needles characteristic of millerite mineral.[7] In addition to Co, also Cu, Mn, and Fe are known impurities, which are identified in millerite. Fig. 5 shows that at the ablation layers at which an increase of the Ni, Co and S signals is observed also signals of Mn, Cu or Mg increases. An exception are Fe and Ti signals, which decrease for these specific ablation layers.

A small part of data points of Ti are observed to linearly correlate with the slope coefficient 0.22 ± 0.03 indicating that Ti may be a part of more complex mineralogical phase(Fig. 5, panel (c)). In the analysed locations on the inclusion there is also no correlation of Ti with O observed. TiO_2 (anatase mineral) was identified, however, in the previous studies in other inclusion locations. For the other trace elements including B, V, Cr, Mn, Zn and Cu measured here the correlation curves are too diffuse and the linear fit is not applied. However, their RSCs are expected to be close to one.[18,20,51] The RSCs for non-metals such as Si, C, O, S or P are typically smaller than one. RSCs of Na and K are typically larger than 1, because it is relatively easy to ionise

alkali elements comparing to other elements. RSCs of Na and K could not be derived from the correlation of these elements with O or S. For the minor and trace elements, the measurement intensities have larger statistical fluctuations resulting in less accurate determination of the slope coefficients. There are ablation layers in which the relevant atomic signals are weak or not detectable. For too large signals, the mass peak intensities can be under-measured due to the detector saturation effects. For large differences in material properties from one ablation layer to the other, the ablation process and efficiency of ion production may change affecting linearity of the atomic intensity correlation. The effects affecting the mass peak intensity measurements are discussed in more detail in our recent publication.[55] The RSCs of investigated elements are summarised in Table 1.†

For a sample consisting of heterogeneities within a volume smaller than the sampled volume, the determination of RSC from the mass spectrum at this location would yield inaccurate results, because an unique element composition is assumed for the entire volume, thus entire data set. Using depth profiling, the mass peak intensity correlation method can be applied for each individual layer, which yields more accurate results.[55] The specific chemical composition of each layer is preserved and the spatial resolution with the profiling depth is expected to be in sub-micrometre range. In this way an isolation between various compounds can be achieved and the different mineralogical phases are recognised by means of their unique elemental composition (abundance ratio) within the various ablation layers. Hence, they can be identified from the correlation studies. Nevertheless, the lateral resolution of this study is at least order of magnitude lower and may affect an isolation of the signals if more than one unique mineralogical grains will be ablated simultaneously.

Only, RSCs of major and some minor elements can be obtained using present method (see also ESI†). The corrections were applied to the atomic intensities of the major species Ca, O, S, Mg and Si. Such corrections can be important when the analysis of mineralogical phases are conducted as the rock chemistry depends on the concentrations of these elements. In ESI,† Section B, we show also the atomic intensities (background level value 10^7) and atomic fractions of all elements identified in this study.

Mineralogical phases

Calcium carbonate. Ternary plots are frequently used in the mineralogical analysis in geology. They can be helpful in identification of the mineralogical context of the analysed sample just by applying a minimum number of chemical species, which are necessary to describe the composition. In the conventional analysis of carbonate phases, ternary diagram correlating abundance fractions (ciderite–magnesite–calcite) is applied. For pure calcium carbonate, the ternary plot of the C, O, and Ca abundance fractions is shown in Fig. 7, panel (a). The experimental data points are observed to be located close to each other, with element abundance ratios (atomic fractions, 20 : 20 : 60) that agrees with the chemical formula $CaCO_3$.
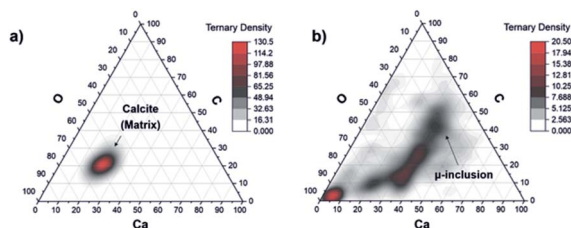
Fig. 7 Ternary diagram of the C, O, and Ca concentration fractions with applied RSCs that were derived on a pure calcium carbonate phase at location 1–4 (a) and on the dark inclusion (b) at locations from 6 to 9. Only a part of the data points in panel (b) can be assigned to calcium carbonate mineral. On the O axes the 3 times O concentration is applied thus the fractions of C, Ca and 3O are the same for carbonate mineral and experimental points are in the middle area in the triangle. The other remaining points indicate presence of other mineralogical compounds.

Contrary, the plot of the data representing the measurements on locations 6 to 9 shows very diffuse distribution with large spread from the area expected for the calcite mineral (Fig. 7, panel (b)). This may indicate that the other minerals instead a calcium carbonate are present at the inclusion location. This goes along with our previous investigations with spectroscopic methods, which identified the presence of mineralogical phases such as clinochlore $[(Mg_5Al)(AlSi_3)O_{10}(OH)_8]$, clinochlore-transformed augite $(Ca,Na)(Mg,Fe,Al,Ti)(Si,Al)_2O_8$ and plagioclase (feldspar group) $[NaAlSi_3O_8–CaAl_2Si_2O_8]$.

Clinochlore, pyroxene-augite, plagioclase. Clinochlore is a member of chlorite group of phyllosilicate minerals and often contains considerable amounts of Fe^{2+}, grading into chamosite. Clinochlore is described by the chemical formula $Mg_5Al(AlSi_3O_{10})(OH)_8$, and Fe, Mn, Zn, Ca, and Cr are common impurities. Fig. 8a shows the ternary plot of the Mg, Al, and Fe element abundances to illustrate their relative proportions measured in the dark inclusion (Fig. 1b). These proportions indicate that the clinochlore mineral is a major mineral in the dark inclusion with the average Mg/(Mg + Fe) abundance ratio of $\sim$0.34 $\pm$ 0.14. Within ultramafic rocks, metamorphism produces predominantly clinochlore chlorite, an Mg-rich end member, which could be identified as the inclusion.

Clinochlore-transformed augite mineral is a member of pyroxene group namely calcic clinopyroxenes (diopside, hedenbergite, augite). Pure diopside of $CaMgSi_2O_6$ chemical composition can undergo as a complete solid solution to hedenbergite $CaFeSi_2O_6$ or augite $(Ca,Mg,Fe^{2+},Fe^{3+},Al)_2(Si,Al)_2O_6$. In another chemical formula proposed for pyroxene group $(Ca_xMg_yFe_z)(Mg_{y1}Fe_{z11})Si_2O$, for augite, the relative composition of Ca can vary within the range $0.4 \leq x \leq 0.9$ where $x + y + z = 1$ and $y_1 + z_1 = 1$.[56] The ternary plot of Mg, Fe, and Ca proportions indicates presence of magnesian augite in the inclusion material (Fig. 8b).

Fig. 8c shows the ternary diagram of the abundances of the elements K, Na, and Ca, which is applied for the classification of feldspar minerals. One can identify the distribution of data along the base of the K coordinate, and the miscibility gap in the centre. The sequence of minerals along the base of the triangle indicate the presence of the plagioclase mineral or the solid solution series of plagioclase being a mixture between albite (sodium aluminosilicate: $[Na(AlSi_3O_8)]$) and anorthite (calcium aluminosilicate: $[Ca(Al_2Si_2O_8)]$) and the notable absence of alkali feldspars.[57] The red line denoting location of the feldspars is schematic. The RSC values of K and Na are chosen 1, which can be the reason that we observe the plagioclase on ternary diagram shifted more in the alkali region. The RSCs of K and Na are expected to be larger than 1 but they cannot here be derived accurately using the abundance correlation method applied to simple compounds.

In our previous studies the analyses conducted by various analytical techniques showed that the dark inclusion investigated here is highly weathered basalt with olivine, and pyroxene (augite) phenocrysts with a dominating amygdales phase.[7] Typically, amygdales in basaltic lavas are formed after lava cools down, and hydrothermal fluids pass through the cavities. Their precipitation in the bubbles form secondary minerals such as calcium carbonate, zeolites, and quartz, depending on the fluid composition. The depth profiles indicate also presence of layers of various mineralogical phases around the millerite needles. Oxides and sulphides of several metals are present as separate entities and some of the metals fill the sites of complex mineralogical compounds.[52]
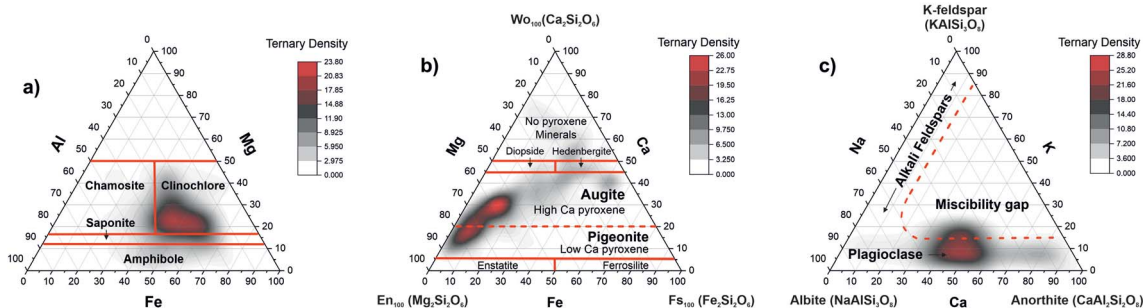

Fig. 8 Ternary plots of the elements relevant to chlorite and feldspar groups. (a) The correlation of the Al, Mg, and Fe abundances showing characteristic regions for various minerals of the chlorite group; (b) the correlation between the Mg, Fe, and Ca abundance ratios indicating the presence of magnesium augite and pigeonite in the sample material; (c) the correlation diagram of the Ca, K, and Na elements indicate the presence of the plagioclase feldspars in the inclusion material.

This journal is © The Royal Society of Chemistry 2021

103

J. Anal. At. Spectrom., 2021, **36**, 80–91 | **89**

Summary

We describe the mass spectrometric measurements conducted by fs-laser ablation/ionisation time-of-flight mass spectrometry and introduce principles of data analysis, which yields to true elemental abundances. The corrections to measured atomic abundances are obtained by analysing atomic intensity correlations using depth-profiling data. In heterogeneous sample such as analysed inclusion embedded in a calcium carbonate mineral, chemical compounds are present at specific depths and can be identified in specific ablation layers. By correlating the measured atomic intensities, one can select the data points for which atomic intensities, correlate linearly. With the slope coefficients of these correlations, the relative abundances of the correlated elements are derived. The method was tested successfully on a pure calcium carbonate phase followed by locations at the interface between inclusion and calcite mineral and on the inclusion. RSCs for major and minor elements were derived by correlating atomic intensity of these elements with either oxygen or sulphur. Subsequently, with the corrected elemental abundances more complex mineralogical phases were analysed by applying the ternary atomic abundance fraction correlations on specific elements typically chosen in geological analysis of the mineralogical context.

The mass spectrometric results are consistent with the results presented in our former studies of this sample conducted with several other instruments. However, LIMS studies cannot differentiate between minerals with the same chemical formula such as calcium carbonate mineral, which can occur as aragonite, calcite or vaterite. To conduct such analysis the Raman measurements will be necessary. Similarly, with the IR analysis one would approve our ambiguous assignment of presence of hydroxides or hydrosulphides in the sample. The presented method can be important in studies of highly heterogeneous materials, because matrix-matched reference materials are often not available for complex natural samples. The *in situ* determined RSCs can overcome the problems associated with the local heterogeneity, which can occur in the micrometre range similar to the sample probing laser spot. Thus, the *in situ* atomic abundance calibration introduced here can allow quantitative chemical composition investigation of heterogeneous structures. With the detection of heterogeneity in the micrometre scale, one can analyse grain-size objects such as microsized minerals or microfossils in detail. The presented method can be also of interest in searches for ancient fossilised life forms on planetary surfaces because in addition to context mineralogy, high depth profiling resolution provides means for isolation and analyses of microfossils embedded in mineralogical phases.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 K. E. Peters, J. M. Walters and J. M. e. Moldowan, *The Biomarker Guide*, Cambridge University Press, 2005.

2 A. H. Stevens, A. McDonald, C. de Koning, A. Riedo, L. J. Preston, P. Ehrenfreund, P. Wurz and C. S. Cockell, *Sci. Rep.*, 2019, **9**, 9706.

3 A. Riedo, C. de Koning, A. H. Stevens, C. S. Cockell, A. McDonald, A. C. Lopez, V. Grimaudo, M. Tulej, P. Wurz and P. Ehrenfreund, *Astrobiology*, 2020, **20**, 1224–1235.

4 A. Riedo, C. de Koning, A. Stevens, A. C. o. L. p. McDonald, A. Tulej, P. Wurz, C. S. Cockell and P. Ehrenfreund, *Astrobiology*, 2020, **20**, 1224–1235.

5 J. M. Reess, M. Bonafous, L. Lapauw, O. Humeau, T. Fouchet, P. Bernardi, P. Cais, M. Deleuze, O. Forni, S. Maurice, S. Robinson, R. C. Wiens and S. Team, *Proc. SPIE*, 2018, 11180.

6 P. Sarrazin, D. Blake, M. Gailhanou, F. Marchis, C. Chalumeau, S. Webb, P. Walter, E. Schyns, K. Thompson and T. Bristow, *J. Instrum.*, 2018, **13**, C04023.

7 A. Neubeck, M. Tulej, M. Ivarsson, C. Broman, A. Riedo, S. McMahon, P. Wurz and S. Bengtson, *Int. J. Astrobiol.*, 2016, **15**, 133–146.

8 M. Tulej, A. Riedo, M. B. Neuland, S. Meyer, P. Wurz, N. Thomas, V. Grimaudo, P. Moreno-Garcia, P. Broekmann, A. Neubeck and M. Ivarsson, *Geostand. Geoanal. Res.*, 2014, **38**, 441–466.

9 M. Tulej, A. Neubeck, M. Ivarsson, A. Riedo, M. B. Neuland, S. Meyer and P. Wurz, *Astrobiology*, 2015, **15**, 669–682.

10 Y. M. Lin, Q. A. Yu, W. Hang and B. L. Huang, *Spectrochim. Acta, Part B*, 2010, **65**, 871–883.

11 A. A. Sysoev and A. A. Sysoev, *Eur. J. Mass Spectrom.*, 2002, **8**, 213–232.

12 A. A. Sysoev, A. V. Karpov, V. V. Milyaeva and A. A. Sysoev, *Eur. J. Mass Spectrom.*, 2018, **24**, 96–107.

13 U. Rohner, J. A. Whitby and P. Wurz, *Meas. Sci. Technol.*, 2003, **14**, 2159–2164.

14 M. Tulej, M. Iakovleva, I. Leya and P. Wurz, *Anal. Bioanal. Chem.*, 2011, **399**, 2185–2200.

15 A. Riedo, A. Bieler, M. Neuland, M. Tulej and P. Wurz, *J. Mass Spectrom.*, 2013, **48**, 1–15.

16 R. Wiesendanger, M. Tulej, A. Riedo, S. Frey, H. Shea and P. Wurz, *J. Anal. At. Spectrom.*, 2017, **32**, 2182–2188.

17 R. Wiesendanger, D. Wacey, M. Tulej, A. Neubeck, M. Ivarsson, V. Grimaudo, P. Moreno-Garcia, A. Cedeno-Lopez, A. Riedo and P. Wurz, *Astrobiology*, 2018, **18**, 1071–1080.

18 J. He, R. F. Huang, Q. Yu, W. Hang and B. L. Huang, *Chem. J. Chin. Univ.*, 2009, **30**, 57–59.

19 B. C. Zhang, M. H. He, W. Hang and B. L. Huang, *Anal. Chem.*, 2013, **85**, 4507–4511.

20 S. D. Zhang, B. C. Zhang, W. Hang and B. L. Huang, *Spectrochim. Acta, Part B*, 2015, **107**, 17–24.

21 Z. B. Yin, L. Hang, R. Liu, W. Hang and B. L. Huang, *J. Mass Spectrom.*, 2018, **53**, 435–443.

22 A. Riedo, M. Neuland, S. Meyer, M. Tulej and P. Wurz, *J. Anal. At. Spectrom.*, 2013, **28**, 1256–1269.

23 V. V. Milyaeva, E. E. Sil'nikov, A. M. Mikhailov and A. A. Sysoev, *Eur. J. Mass Spectrom.*, 2017, **23**, 167–173.

24 M. Tulej, R. Wiesendanger, A. Riedo, G. Knopp and P. Wurz, *J. Anal. At. Spectrom.*, 2018, **33**, 1292–1303.

25 R. F. Huang, Y. M. Lin, L. F. Li, W. Hang, J. He and B. L. Huang, *Anal. Chem.*, 2010, **82**, 3077–3080.

26 R. Wiesendanger, V. Grimaudo, M. Tulej, A. Riedo, R. Lukmanov, N. Ligterink, R. Fausch, H. Shea and P. Wurz, *J. Anal. At. Spectrom.*, 2019, **34**, 2061–2073.

27 M. B. Neuland, S. Meyer, K. Mezger, A. Riedo, M. Tulej and P. Wurz, *Planet. Space Sci.*, 2014, **101**, 196–209.

28 S. Frey, R. Wiesendanger, M. Tulej, M. Neuland, A. Riedo, V. Grimaudo, P. Moreno-Garcia, A. C. Lopez, M. Mohos, B. Hofmann, K. Mezger, P. Broekmann and P. Wurz, *Planet. Space Sci.*, 2020, **182**, 104816.

29 A. Riedo, M. Tulej, U. Rohner and P. Wurz, *Rev. Sci. Instrum.*, 2017, **88**.

30 V. Grimaudo, P. Moreno-Garcia, A. C. Lopez, A. Riedo, R. Wiesendanger, M. Tulej, C. Gruber, E. Lortscher, P. Wurz and P. Broekmann, *Anal. Chem.*, 2018, **90**, 5179–5186.

31 V. Grimaudo, P. Moreno-Garcia, A. Riedo, M. B. Neuland, M. Tulej, P. Broekmann and P. Wurz, *Anal. Chem.*, 2015, **87**, 2037–2041.

32 V. Grimaudo, M. Tulej, A. Riedo, R. Lukmanov, N. F. W. Ligterink, C. de Koning and P. Wurz, *Rapid Commun. Mass Spectrom.*, 2020, **34**, e8803.

33 V. Grimaudo, P. Moreno-Garcia, A. Riedo, S. Meyer, M. Tulej, M. B. Neuland, M. Mohos, C. Gutz, S. R. Waldvogek, P. Wurz and P. Broekmann, *Anal. Chem.*, 2017, **89**, 1632–1641.

34 M. H. He, Y. F. Meng, S. S. Yan, W. Hang, W. G. Zhou and B. L. Huang, *Anal. Chem.*, 2017, **89**, 565–570.

35 B. Fernandez and R. Pereiro, *Anal. Bioanal. Chem.*, 2010, **396**, 2723–2724.

36 B. Fernandez, R. Pereiro and A. Sanz-Medel, *Anal. Chim. Acta*, 2010, **679**, 7–16.

37 B. Hattendorf, J. Pisonero, D. Gunther and N. Bordel, *Anal. Chem.*, 2012, **84**, 8771–8776.

38 A. Pelster, M. Korsgen, T. Kurosawa, H. Morita and H. F. Arlinghaus, *Anal. Chem.*, 2016, **88**, 9638–9646.

39 J. Pisonero, *Anal. Bioanal. Chem.*, 2006, **384**, 47–49.

40 J. Pisonero, J. Koch, M. Walle, W. Hartung, N. D. Spencer and D. Gunther, *Anal. Chem.*, 2007, **79**, 2325–2333.

41 J. Fietzke, V. Liebetrau, D. Guenther, K. Gurs, K. Hametner, K. Zumholz, T. H. Hansteen and A. Eisenhauer, *J. Anal. At. Spectrom.*, 2008, **23**, 955–961.

42 J. Fietzke, V. Liebetrau, D. Gunther, A. Frische, K. Zumholz, T. H. Hansteen and A. Eisenhauer, *Geochim. Cosmochim. Acta*, 2008, **72**, A267.

43 V. N. Epov, S. Berail, M. Jimenez-Moreno, V. Perrot, C. Pecheyran, D. Amouroux and O. F. X. Donard, *Anal. Chem.*, 2010, **82**, 5652–5662.

44 M. Ivarsson, *Biogeosciences*, 2012, **9**, 3625–3635.

45 M. Ivarsson, W. Bach, C. Broman, A. Neubeck and S. Bengtson, *Geomicrobiol. J.*, 2018, **35**, 460–467.

46 M. Ivarsson, S. Bengtson, V. Belivanova, M. Stampanoni, F. Marone and A. Tehler, *Geology*, 2012, **40**, 163–166.

47 S. McMahon, T. Bosak, J. P. Grotzinger, R. E. Milliken, R. E. Summons, M. Daye, S. A. Newman, A. Fraeman, K. H. Williford and D. E. G. Briggs, *J. Geophys. Res.: Planets*, 2018, **123**, 1012–1040.

48 S. McMahon, J. Parnell and N. J. F. Blamey, *Int. J. Astrobiol.*, 2012, **11**, 163–167.

49 R. Wiesendanger, M. Tulej, V. Grimaudo, A. C. Lopez, R. Lukmanov, A. Riedo and P. Wurz, *J. Chemom.*, 2018, 1–10, DOI: 10.1002/cem.3081.

50 S. Meyer, A. Riedo, M. B. Neuland, M. Tulej and P. Wurz, *J. Mass Spectrom.*, 2017, **52**, 580–590.

51 M. B. Neuland, V. Grimaudo, K. Mezger, P. Moreno-Garcia, A. Riedo, M. Tulej and P. Wurz, *Meas. Sci. Technol.*, 2016, **27**, 035904.

52 T. M. Seward, A. Williams-Jones and A. Migdisov, in *Tratise on Geochemistry*, Elsevier Ltd, 2013, vol. 13.

53 U. M. Graham and H. Ohmoto, *Geochim. Cosmochim. Acta*, 1994, **58**, 2187–2202.

54 H. Ohmoto, K. Hayashi and Y. Kajisa, *Geochim. Cosmochim. Acta*, 1994, **58**, 2169–2185.

55 M. Tulej, A. Neubeck, A. Riedo, R. Lukmanov, V. Grimaudo, N. Ligterink, M. Ivarsson, W. Bach and C. de Koning, *J. Mass Spectrom.*, 2020, **55**, e4660.

56 N. Morimoto, *Mineral. Mag.*, 1988, **52**, 535–550.

57 http://geology.com/minerals/plagioclase.shtml.

3.2 Spatially resolved 2D and 3D imaging

3.2.1 Chemical identification of microfossils from the 1.88 Ga Gunflint chert. Towards empirical biosignatures using LIMS - (Chemometrics)

Rustam A. Lukmanov[1]*, Marek Tulej[1], Niels F. W. Ligterink[1], Coenraad De Koning[1], Andreas Riedo[1], Valentine Grimaudo[1], Anna Neubeck[2], David Wacey[3], Peter Wurz[1]

[1]University of Bern, Bern, Switzerland.

[2]Department of Earth Sciences, Uppsala University, Uppsala, Sweden.

[3]Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, Perth, Australia.

*Rustam A. Lukmanov

Email: rustam.lukmanov@space.unibe.ch

RESEARCH ARTICLE

Chemical identification of microfossils from the 1.88-Ga Gunflint chert: Towards empirical biosignatures using laser ablation ionization mass spectrometer

Rustam A. Lukmanov[1] 🟢 | Marek Tulej[1] | Niels F. W. Ligterink[1] |
Coenraad De Koning[1] | Andreas Riedo[1] 🟢 | Valentine Grimaudo[1] |
Anna Neubeck[2] | David Wacey[3] | Peter Wurz[1]

[1]Space Research & Planetary Sciences (WP), University of Bern, Bern, Switzerland

[2]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

[3]Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, Perth, Western Australia, Australia

Correspondence
Rustam A. Lukmanov, Space Research & Planetary Sciences (WP), University of Bern, Bern, Switzerland.
Email: rustam.lukmanov@space.unibe.ch

Abstract

In this contribution, we investigated the chemical composition of Precambrian microfossils from the Gunflint chert (1.88 Ga) using a miniature laser ablation ionization mass spectrometer (LIMS) developed for in situ space applications. Spatially resolved mass spectrometric imaging (MSI) and depth profiling resulted in the acquisition of 68,500 mass spectra. Using single mass unit spectral decomposition and multivariate data analysis techniques, we identified the location of aggregations of microfossils and surrounding inorganic host mineral. Our results show that microfossils have unique chemical compositions that can be distinguished from the inorganic chert with high fidelity. Chemical depth profiling results also show that with LIMS microprobe data, it is possible to identify chemical differences between individual microfossils, thereby providing new insights about nature of early life. Analysis of LIMS spectra acquired from the individual microfossils reveals complex mineralization, which can reflect the metabolic diversity of the Gunflint microbiome. An intensity-based machine learning model trained on LIMS Gunflint data might be applied for the future investigations of putative microfossils from silicified matrices, where morphological integrity of investigated structures is lost, and potentially in the investigation of rocks acquired from the Martian surface.

KEYWORDS
Gunflint, Mars, mass spectrometry, microfossils, space instrumentation

1 | INTRODUCTION

In situ research and remote sensing have provided multiple lines of evidence that clement conditions were present on the surface of early Mars.[1-3] Moreover, recent radar studies reveal evidence of subglacial liquid water on Mars,[4] which supports the hypothesis that microbial life forms (extinct or extant) may be preserved within the Martian subsurface.[4,5]

All these observations provide a strong rationale for the search of biosignatures on the Red Planet. The current state of space exploration provides foundation for new measurement techniques and novel analytical approaches[6] to identify and characterize minerals and potential signatures of life, if any, on Mars. However, in situ investigation of rocks on Mars faces multiple technological difficulties, ranging from constraints on instrumentation robustness, size, and power consumption to the quality of the acquired data.[7] Some authors proposed an implementation of multicriteria approach[3,8,9] to confirm or reject a biogenic origin of the given sample. Multicriteria approach, thus, requires several instruments onboard of the rover providing a multiplex analysis of the same sample and identification of morphological, molecular, elemental, and isotopic signatures of life. However, traditional methods used in space research, for example, bulk analysis and remote sensing instrumentation, might not be sufficiently sensitive to detect faint features from micrometer-sized (and below) organic material or microbial remains[6,10] and, in some cases, can alter the chemistry of the sample. There is a growing demand for sensitive, in situ instruments with high spatial resolution and minimal sample processing, providing elemental and organic composition detection, which will enhance the scientific return from the missions to Mars and icy moons of Jupiter.[6,10–12]

In addition to the development of analytical methods, development of chemometric tools has also proven to be a field of high importance to the current and future space exploration programs.[13] For example, multivariate curve resolution alternating least squares have been shown to successfully identify various minerals and compounds from Raman hyperspectral images, overcoming the spectral overlap issues.[14,15] The linear mixture model (LMM) was successfully used to quantify the abundance of major elements using the laser-induced breakdown spectroscopy (LIBS) spectra.[16] Furthermore, a data fusion approach was reported for complementary analytical techniques (Raman and LIBS) that improved the classification limits of investigated binary compounds.[17]

Laser-based mass spectrometry is an emerging and sensitive technique that has shown to be capable of measuring extremely low concentrations (fmol) of amino acids in desorption mode,[18] elemental detection of single microbes in Martian mudstone analog material,[11] and provides chemical (element, isotope) analysis in ablation mode of any solid material. The latter can be conducted with a depth resolution on the scale of tens of nm[19,20] and with high element detection sensitivity down to the ppb level.[21,22] This makes laser-based mass spectrometry an attractive method in the field of in situ chemical analysis on planetary surfaces.[23] The upcoming ExoMars mission/Rosalind Franklin Rover contains a Laser Desorption/Ionization (LDI)-Quadrupole Mass Spectrometer in its instrument suite,[24,25] stimulating further development of LDI instruments.[26–30] However, laser ablation ionization mass spectrometer (LIMS) capability to detect and identify billion-year-old microfossils has not been shown so far.

Since the discovery of Precambrian microbial communities in the 1.88-Ga Gunflint Formation (Ontario, Canada) in the early 1960s,[31–33] many more examples of Precambrian life have been found, but the Gunflint Formation retains its place as a premier Precambrian Fossil-Lagerstätten, demonstrating that Paleoproterozoic life was widespread, already complex, and diversified.[12,34,35] The Gunflint chert sample in this study has been used as a Martian analog, reflecting the iron-rich nature of the Martian sediments, as well as taking into account that siliceous sediments have high preservation capacity and are of interest to upcoming astrobiological missions.[34,36–38] However, despite being among the best example of Precambrian life, phylogenetic affinities and metabolic speciation of the Gunflint microfossils remain largely unknown. Similarly, microfossils of unknown affinities dominate the majority of the Precambrian record.[39] Traditionally, the classification of types of microfossils has relied on morphological features[31,40] and later advanced to include isotope fractionation[41] and multielement nano-characterization of individual microfossils.[35] However, within the space instrumentation domain, detection of individual microfossils remains a technological challenge.[10]

In this contribution, we present results on mass spectrometric imaging (MSI) and chemical depth profiling from the Gunflint chert using a miniature time-of-flight reflectron-type mass spectrometer developed for in situ space applications, equipped with a femtosecond (fs) UV-258-nm laser ablation ion source. Mass spectrometric studies conducted on microfossils allowed their identification within the inorganic host material. Utilizing the depth profiling approach, we removed the contaminated layer present on the rock surface and probed the original chemical composition of the microfossils. The network-based approach used for the interpretation of hundreds of recorded mass spectra revealed a new topological dimension, where separate mineralogical inclusions present within the same analytical spot (inclusions smaller than the size of the ablation spot) can be readily separated.

The analysis of the mass spectra from the depth profiles revealed the presence of major biorelevant elements (CHNOPS), microscopic inclusions of Cu, Cr sulfides, rare earth element (REE) minerals in addition to the majority of Fe-dominated mineralization associated with the microfossils. These observations can indicate the presence of sulfur-processing and iron-processing species[42] but also can indicate the presence of intracellular biomineralization machinery

(passive mineralization) within the Gunflint microbiome to withstand possible Cr and Cu toxicity within an already highly ferruginous environment.[39] H/C, O/C, Si/C ratios, and principal component analysis (PCA) scores calculated from the depth profiling dataset show ratios and intensity regions in which microfossils can be identified. Large-scale mass spectra sampling allowed the construction of binary classification machine learning (ML) models, which can be used for the identification of microfossils from other Precambrian cherts and other rocks (once calibrated), where morphological integrity of the putative microfossils is lost, thus providing a way to assess biogenicity, by comparing spectra from other cherts to our truly biogenic model data.

Overall, the LIMS microprobe shows an ability to identify micro- and nano-mineralization associated with the microfossils. LIMS imaging combined with accurate depth profiling has the potential to reveal new insights into the distribution, preservation, and elemental speciation of microfossils from Precambrian cherts and a potential to deliver insights into the chemical composition of samples acquired from the Martian surface.

2 | MATERIALS AND METHODS

A standard double-polished thin section of the Gunflint chert (collected from the Schreiber beach locality, Ontario, Canada) with a thickness of ∼30 μm was used in the current study. The sample was mounted on a metal sample holder with vacuum-compatible copper tapes to fix the sample on the surface of the steel holder. No additional treatment of the sample was performed. A miniature time-of-flight mass spectrometer (TOF-MS) developed at the University of Bern has been used to study the Gunflint sample. The mass analyzer has small dimensions: Ø 60 mm × 160 mm, which makes it suitable for space exploration programs as part of the lander or rover.[7,23,30] The time-of-flight mass analyzer works in the positive ion detection mode and provides a single unit mass resolution and a ppm-level sensitivity. For the detailed characterization of the instrument and method in general, we refer the interested reader to the reviews[43,44] and technical reports describing figures of merit.[21,23,45,46] In the current laboratory setup, a mass spectrometer is accompanied by an integrated microscopy system[12] and a fs UV-258-nm ionization source. The instrument is designed to have spatial molecular,[18,47] elemental, and isotopic[22,29] mapping capabilities of solid samples. The microscopy system utilized in the current laboratory environment is not space qualified; however, a separate space prototype was developed in our group that combines the microscopy system, mass analyzer, and an ion source.[23] High precision XYZ translation stage is used for the accurate sample positioning under the instrument: 1-μm positioning accuracy is typically achieved between the internal microscope and laser focal point positions.[12] Laser power output stability, beam profile, and crater shapes are checked prior to the measurement campaign.

The Gunflint sample studied in this work could be approximated as a dielectric glass with inclusions of dark absorptive features (microfossils). To enhance the ionization efficiency of the quartz mineral, which constitutes most of the sample, and shorten the gap between the ionization efficiency of microfossils and host mineral, we implemented a frequency tripling system (STORC) into the beam delivery line, which allowed us to reach a stable fs UV-258-nm laser radiation. Photon energy, at this point, reached 4.8 eV, which is well suited for the ablation and ionization of any solid materials. Ti:sapphire fs Laser from Clark Inc. generates infrared (IR) 775-nm fundamental wavelength (180 fs), which is guided into the frequency doubling and tripling system. ß-BBO crystals are used in the doubling and tripling parts of the STORC system to achieve UV light generation.

In this contribution, we performed two separate data acquisition campaigns: (1) the MSI of the 1.5 × 2-mm area with 10-μm gaps between ablation craters and (2) depth profiling campaign within specified areas—microfossil-rich zone and a host mineral (quartz) area. The MSI campaign yielded 30,000 (150 × 200 pixels) mass spectra, where every pixel represents a histogram of 200 single laser shot spectra. The output mass spectra were processed using a single mass unit spectral decomposition (extraction of consecutive mass peak areas, following the footsteps of Meyer et al., 2017). Overall, 182 single mass unit intensities, retrieved from every pixel, were used to form the isotope intensity maps. Further, the Gaussian process (Kriging) interpolation was used to increase the resolution of the output maps by a factor of 2. In total, ^{28}Si, ^{16}O, ^{12}C, and ^{1}H maps were calculated to visualize the distribution of microfossils and filling quartz mineral on the surface of the sample. The second data acquisition campaign was performed to remove the organic contamination layer present on the surface and probe the original chemical composition of the microfossils and subsequently build a binary classification model. The depth profiling campaign resulted in the acquisition of 38,500 spectra from 15 depth profiles acquired from the microfossils-rich zone and a host mineral (see the supporting information for detailed information about data processing and filtering). The output spectral intensity profiles from both locations were log-transformed and z-score standardized.

The PCA of the depth profiling dataset was conducted using the correlation matrix computed on centered data. The first three principal components were extracted from the depth profiling dataset. The weighted correlation networks of inclusions present in the depth profiles were calculated using the direct Pearson pairwise correlation scores of 182 single unit mass intensities. The correlation scores retrieval resulted in the acquisition of 16,380 Pearson correlation pairs for given inclusion. Further, extracted correlation pairs were used as weights defined on pairwise edges in the construction of the undirected network. The force-directed layout (ForceAtlas2) provided with Gephi[48,49] was used to visualize the network structure. The edge weight threshold was implemented to remove the insignificant correlation values from the network (see further down in the text). The pairwise kernel density estimates of biorelevant ratios and element intensities were used to visualize the density distributions of investigated locations (silicified host mineral and microfossils). Furthermore, 24 binary classification ML models (including classification trees, support vector machines, and ensemble models) were scored using the Matlab ML presets (see supporting information for more details). The isotope ratios and synthetic metrics (e.g., geometric mean values of light masses) were added to the ML dataset, making 196 variables in total. The fivefold cross-validation was used to avoid overfitting of the dataset. Within best-performing models, an additional 30-step Bayesian optimization (search through different learning hyperparameters) procedure was implemented to test for potential improvements in the output performance. For a more detailed description of ML models, information on PCA, and weighted correlation networks, we refer to the supporting information.

3 | RESULTS AND DISCUSSION

3.1 | LIMS imaging and depth profiling

All experimental measurement procedures—surface imaging and depth profiling—performed on the sample with a miniature LIMS system are schematically illustrated in Figure 1. The drawing is out of scale and intended to give a better understanding of the subject of this study. The distribution of microfossils embedded in a quartz matrix is shown with gray lenticular structures. Most of the preserved species studied by nano-microscopy are hollow and represent partially collapsed cell walls. An approximate estimate of the thickness of the cell walls mentioned in the figure is 500 nm; however, actual thicknesses are varying. For detailed morphological studies of these microfossils, we refer to literature.[35,39] The small layer on the surface represents surface contamination with recent organic material. A focused UV-258-nm fs laser beam shown in Figure 1 was used to ablate and ionize material from the Gunflint chert. The produced ions were transmitted into a miniature TOF-MS (LIMS) developed for the operation on planetary surfaces.

To acquire information on the chemical composition of the stromatolitic layering from the Gunflint chert sample, we performed MSI of an area of a thin section containing two morphological features: (1) dense population of microfossils and (2) clean host mineral area. We identified a rectangular area (1.5 mm × 2 mm) (Figure 2A,B) where these features were present. To accurately sample the area under investigation, a 10-μm gap between the ablation craters was chosen for the imaging, resulting in a grid with 30,000 ablation spots (with a single mass spectrum corresponding to each spot—see Figures 5 and S3–S5). To avoid material displacement and crater-to-crater cross-contamination from the ablation processes, a pulse train of 200 laser pulses was applied to each location, yielding a single mass spectrum. Because imaging implies probing material from different parts of the sample with different light absorption properties, suitable pulse energies were determined on preliminary craters from various locations prior to the imaging. Laser pulses with an energy of 0.36 [μJ] per pulse (measured at the sample surface) were found to be appropriate for both the dense microfossil assemblage and the clean host area. The diameter of the analytical spot was determined to be around 4–5 μm for the imaging campaign within dominantly quartz locations and 7–8 μm within microfossils.

Figure 2D,E shows the distribution of ^{1}H and ^{12}C signal intensities on the surface of the Gunflint sample, extracted from the mass spectra using Simpson integration (details about integration procedure can be found in literature[50]) and defined as an output current registered from the detector per unit of time ($\log_{10}$ electrons * ns^{-1}). As is clear from Figure 2D,E, hydrogen and carbon are spatially correlated with the location of the microfossils identified by the optical microscopy (see Figure 2B). Resolution of the imaging reached a single cell level (see Figures S4 and S5) and could be improved by a factor of 2 in future campaigns without any analytical interference. Figure 2F,G displays the distribution of ^{28}Si and ^{16}O intensities recorded on the surface of the sample. Both isotopes show a relatively homogeneous distribution, with a good correlation to each other. The host material in which the microfossils are embedded is diagenetic quartz with varying crystal sizes. It is also possible to observe an enhanced signal of ^{28}Si and ^{16}O within the microfossil's lamination area due to the enhanced absorption of light by microfossils. Imaging of minor and trace elements was not
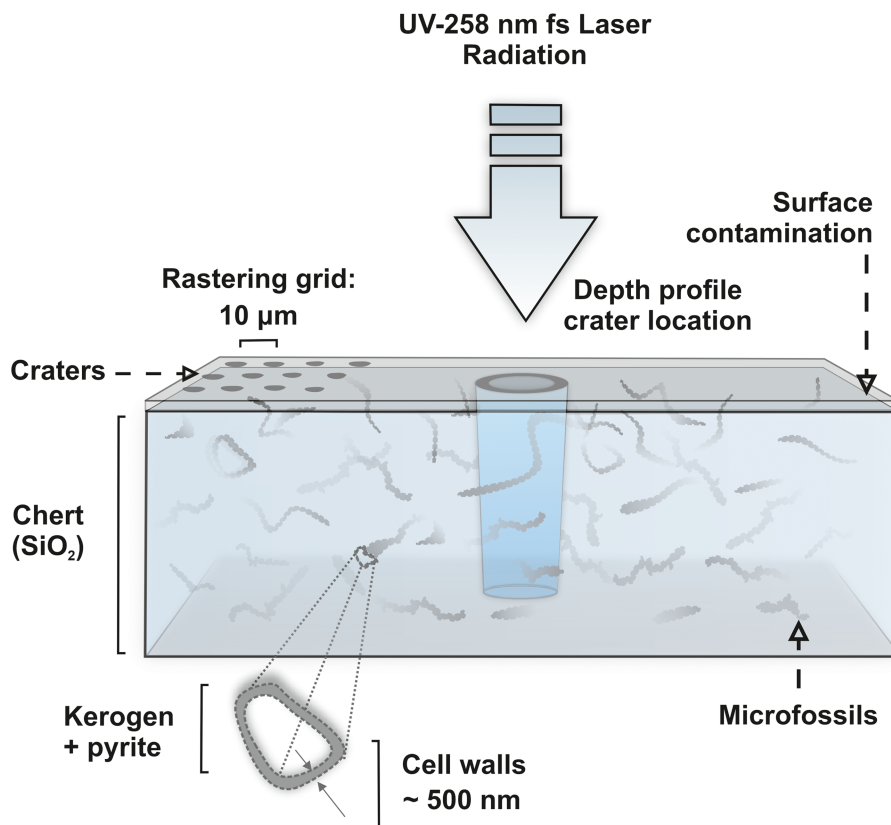
FIGURE 1   Schematic representation of the sample with three-dimensional distribution of various microfossils within the chert. The arrow represents focused fs laser light forming an ablation crater. The blue cylinder represents an ablation crater, formed during the depth profiling (not to scale). On the left side of the arrow, smaller ablation craters from the surface imaging are present (ablation craters are at different scales). On the bottom, the schematic composition of the hollow microfossils and approximate thickness of the cell walls are shown

possible due to isobaric contamination by hydrocarbon clusters in the mass spectra (see Figure S12). The separate depth profiling campaign was performed to remove the contaminated layer on the surface and probe the original chemical composition of the microfossils and host area.

3.2 | Depth profiling

Figure 3 shows the locations at which depth profiles were measured within the lamination area containing the population of microfossils and the host quartz area. Depth profiling analysis was conducted on nine spots containing microfossils and six spots of clean host quartz. Spots studied by depth profiling are independent of the grid that was used for MSI. For seven spots from the microfossils and six spots from the host area, 2500 spectra were measured simultaneously on two acquisition cards and summed together (forming a 5000-spectra dataset for a single depth profile before summation), where each spectrum consists of 200 single laser pulse spectra and corresponds to the single ablation layer. Two additional spots on a microfossil-rich area were measured with reduced histogramming down to 64, and 32 single laser shot spectra, in an attempt to obtain an even finer sampling of the microfossils. In total, 3000 spectra were collected from each of these spots (6000 spectra before preprocessing). A data extraction procedure is performed, retrieving intensities of the single mass units, utilizing direct Simpson integration[50] of the time-of-flight windows determined for each mass. In total, 182 single unit masses from each spectrum have been retrieved, including a background signal (noise measurements), which were determined in the time-of-flight window, free of any ion signal.

Figure 3A,B displays location and morphology of exemplary craters acquired during the depth profiling campaign. The dark patches in Figure 3B represent an aggregation of microfossils, where an arrow indicates location of the single lenticular microfossil. Figure 3C shows variation of the ^{12}C mass peak intensities measured at the selected location. As is clear from Figure 3C, the ^{12}C mass peak intensity measured within 32,500 spectra reveals clear intensity separation
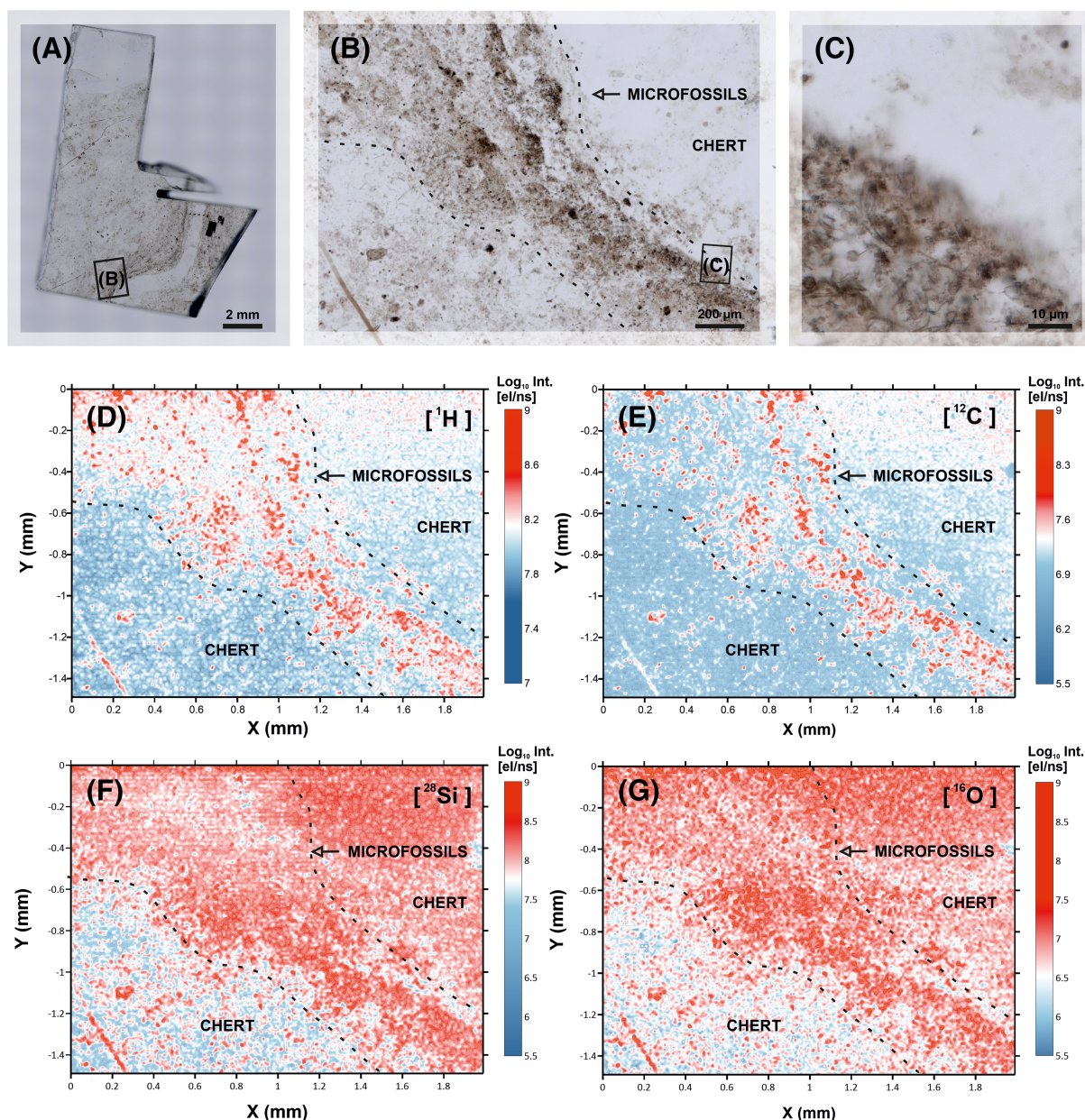
FIGURE 2 Optical images of the Gunflint thin section and results of the MSI. (A) Panoramic picture of the Gunflint thin section sample with a designated location of the area chosen for the MSI. (B) Image of the area chosen for the MSI with a designated location of the close-up picture (images are constructed by stacking of multiple focal points). The actual distribution of microfossils exposed to the surface is less dense but follows the same structure. (C) Close-up image of the distribution of microfossils within a densely populated area at the edge of the lamination. (D, E) Intensity maps of the ^{1}H and ^{12}C. Red areas correspond to the intense signal from the microfossils, and blue areas to the lower intensities from the surrounding chert. Black dashed line outlines the edge of the lamination area. To compare with the optical image of the same area, see Figure 2B. Lower panels (F) and (G): Intensity maps of ^{28}Si and ^{16}O, which constitutes inorganic silicified host of the Gunflint sample

boundary between two locations (note that intensities are presented in the log scale). A limited amount of ^{12}C signal, registered from the host region, may originate from the surface of the sample due to the widening of the crater, while ablation is progressing. Figure 3D shows the single depth profile from the microfossil-rich location and reveals increased ^{12}C intensities within specific depth regions, indicating that the measured carbon originates from the inclusions (Figure 3D, bars with location X and Y). In contrast, the depth profiles registered from the host region (Figure 3C) show the presence of a significantly reduced amount of carbon, in comparison with the depth profile acquired from the microfossil-rich lamination area. Localized aggregation of intense peaks of carbon within the bulk of the quartz matrix
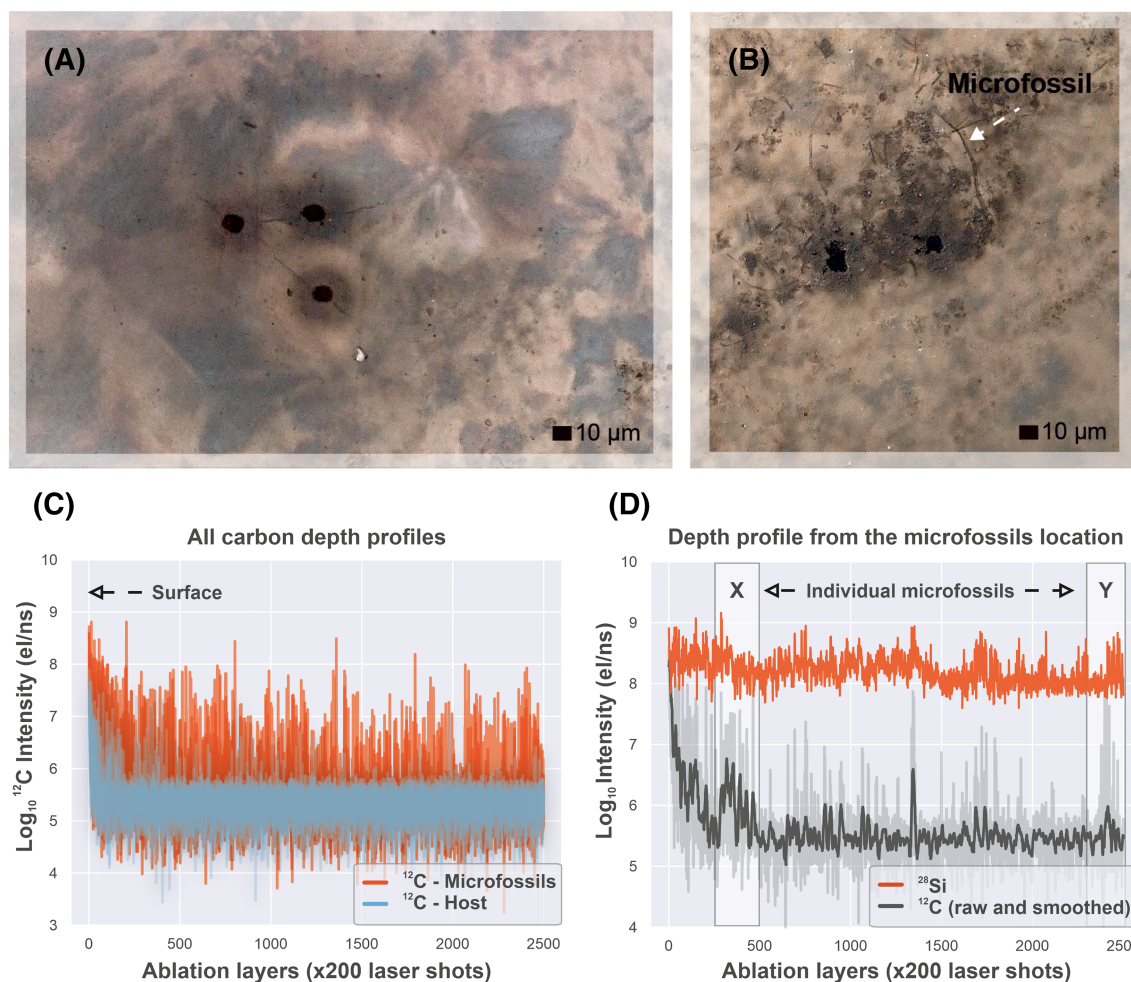
FIGURE 3 Location of the exemplary craters from the depth profiling campaign and registered ion yields from different locations. (A) Exemplary depth profile craters from the host area. (B) Exemplary depth profile craters from the microfossil aggregation zone. Arrow indicates location and relative size of the microfossil in comparison with the size of the analyzed crater. (C) Ion yield curves registered from the depth profiles. Superimposed ^{12}C depth profiles (integrated ^{12}C peak intensity) from the seven microfossil locations (orange line) and six host locations (blue line) (32,500 spectra in total) are shown. An intense carbon signal on the surface indicates the presence of surface contamination. (D) Single depth profile registered from the microfossil-rich spot (see Figure 3B). The ^{28}Si (orange line), ^{12}C (gray line—raw data, black—smoothed by moving average ^{12}C ion yield curve) depth profile constructed by extracting peak areas from 2500 consecutive spectra. Location of individual microfossils within the depth profile indicated by X and Y

is interpreted to be individual bodies of microfossils (see the sketch—Figure 1). Moreover, carbon-enriched inclusions are associated with other biorelevant elements: CHNOPS (see further in the text and Figure 4B,C), which indicates that these inclusions are indeed individual microfossils, located in the distinct depth regions. However, to prove that these localized spectra are acquired from a single source (microfossil), we calculated the correlation networks, which will be presented later in the text. Here, we need to mention that LIMS, being a destructive method, provides sensitive and spatially resolved measurements that are hard to achieve using bulk characterization methods. For example, in the low-biomass simulation of Martian sediments,[10,11] the results have shown that LIMS can identify spatially constrained biosignatures in Mars analog environments.

Figure 4 shows spectra measured from different mineralogical inclusions present within the Gunflint subsurface. The spectrum measured at the host area (Figure 4A) reveals the chemical composition corresponding to the diagenetic quartz. Intense peaks of Si and O are readily recognized. Additionally, peaks of H, C, Na, K, and chain of SiO clusters can be identified in the spectrum (Figure 4A). The mass spectrum in Figure 4B was measured within location X (Figure 3D) at depth position 250–500. The spectrum reveals the complex chemical composition of the microfossil body intermixed with the host chemistry. The chemical composition measured within this spot represents a mixture of the
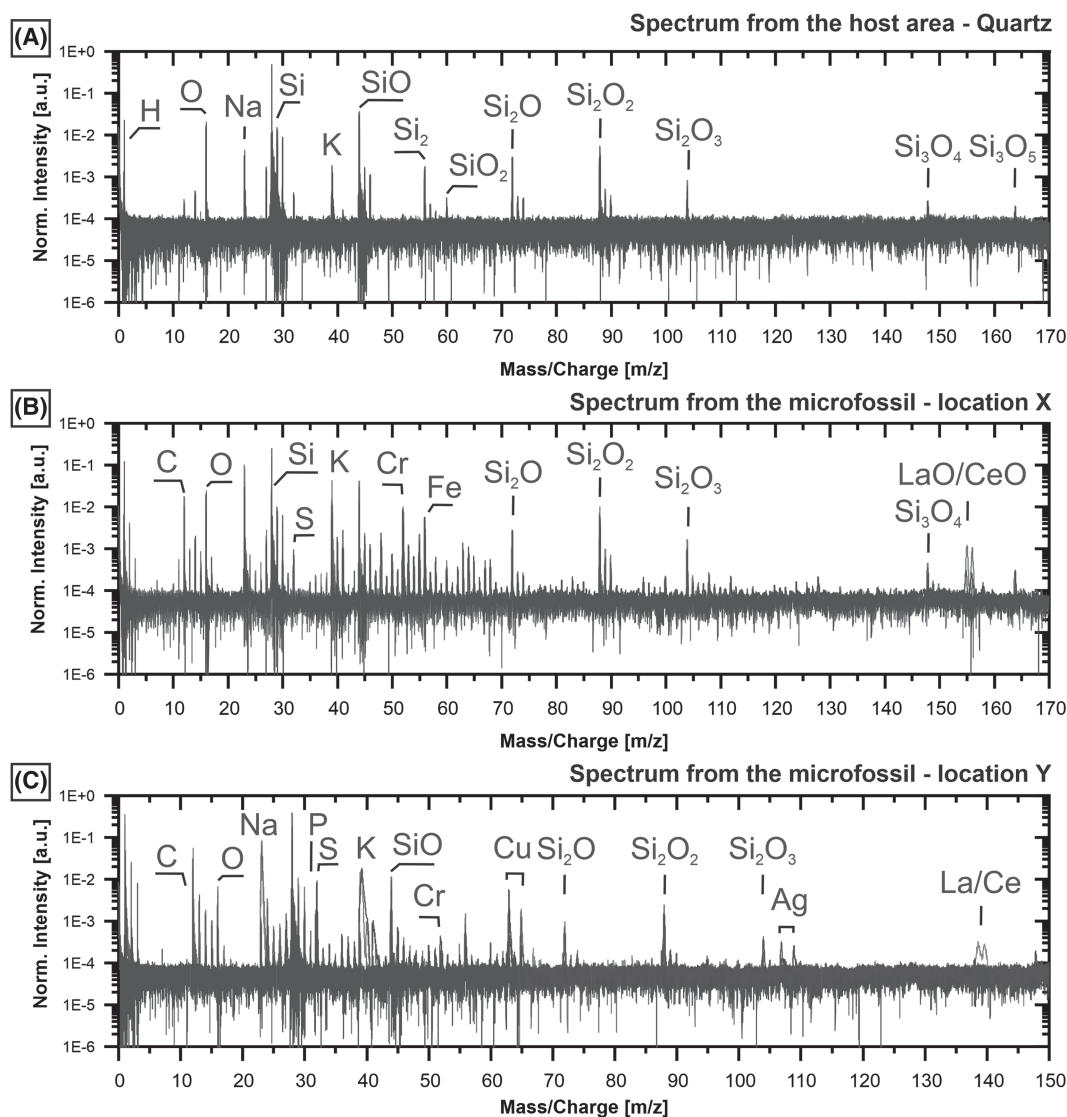
FIGURE 4 Mass spectra from various parts of the sample. (A) Mass spectrum from the host area. (B) Mass spectra from the microfossil (location X in the depth profile; see Figure 3D). (C) Mass spectrum from the microfossil (location Y in the depth profile; see Figure 3D). Note the increased peak intensities of biorelevant elements (CHNOPS) and richer chemistry within microfossils in comparison with the host mineral (quartz; see Figure 4A)

quartz mineral, kerogen from microfossil cell walls, and a polymetallic inclusion associated with the microfossil. Intense mass peaks of transition metals, Ti, Mn, Fe, and Cu, can be identified in the spectrum with a major contribution from Cr and additional minor contribution from La and Ce monoxides. The presence of kerogen is identified from the detection of C, H, N, O, P, and S. Additionally, multiple low-intensity C_xH_y compounds can be identified in the spectrum.

Figure 4C shows the mass spectrum measured at depth locations 2350–2500 (location Y; see Figure 3D) and reveals a different composition compared with the inclusion described above. In addition to the mass peaks registered from quartz and CHNOPS, there are also peaks of Mg, Al, K, Ca, Cr, Mn, La, and Ce, and mass peaks of Cu and Ag could be noted. The latter are rather unexpected to be found within microfossils because they are known to be elements with high cytotoxicity (i.e., they are toxic to cells).

Although LIMS can yield sensitive measurements of elements and isotopes, it can be challenging to determine mineralogical composition of multiple microinclusions (smaller than the size of the probing laser spot). The pairwise correlation factors between single mass unit intensities are calculated in an attempt to identify mineralogical composition of investigated inclusions. Figure 5 shows two networks calculated from the individual microfossils (Figure 3D) and
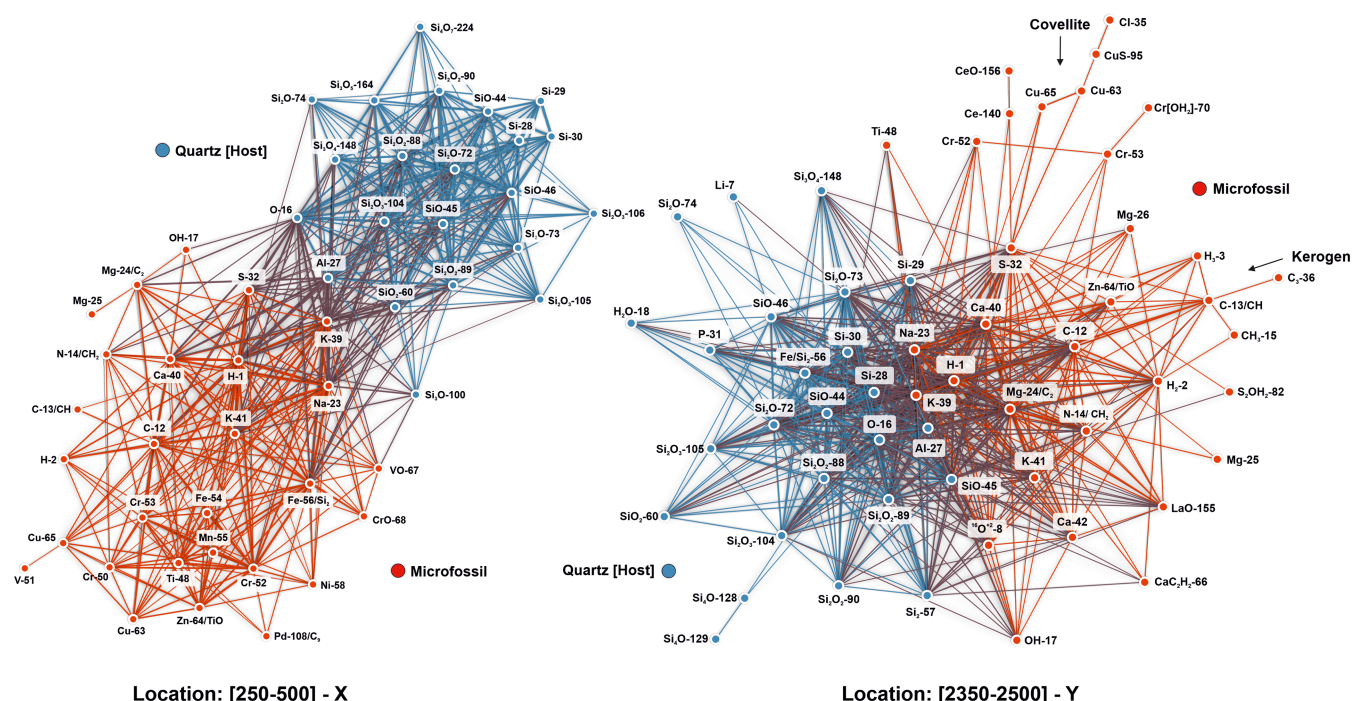
FIGURE 5　Weighted mass correlation networks calculated from the inclusions registered at locations X and Y (see Figures 3D and 4B, C). Left: Network from the location X (see Figures 3D and 4B) (ρ larger than 0.4). The network shows presence of two separate entities within studied location—microfossil (red nodes) and quartz (blue nodes). Right: Network from the depth location Y (see Figures 3D and 4C) (ρ larger than 0.2). The network shows two separate inclusions—microfossil (red nodes) and quartz (blue nodes). Chemical composition and topology of network X are different from network Y, which could be attribute to the different size and chemical composition of microfossils. The straight line between two nodes (single unit masses—m/z) represents a linear correlation factor measured within the microscopic inclusions. The node colors are assigned according to the Louvain modularity score. Note the central elements within each subgroup—C, H are among central elements within microfossils and silicon oxides within quartz

visualized using an open-source graph drawing platform Gephi.[49] The mass correlation network shown on the left side was calculated from location X, which is located at the depth region: 250–500. Correlation factors (ρ) larger than 0.4 were used to visualize only strongly correlated masses (see full correlation matrices in the supporting information). Node colors are chosen according to the modularity rank,[51] calculated from the network topology (modularity Q is a parameter that measures the density of links in the graph inside communities as compared with links between communities). Blue nodes correspond to the elements registered from the quartz mineral, whereas red nodes represent elements and isotopes registered from the microfossil body. Clear separation of the inclusion from the host chemistry could be observed within this network—host mineral (quartz) colored with blue nodes and a microfossil (red nodes).

The mass correlation network shown on the right side of Figure 5 was calculated from the depth region 2350–2500, at location Y (see Figure 3D), and represents the chemical composition of another microfossil. Due to the lower number of spectra registered from this inclusion (size of the inclusion was considerably smaller than the inclusion described above), the ρ cutoff value was set to 0.2. However, even with low correlation factors, it is possible to obtain the appropriate modularity ranks and to increase the interpretability of the data. Within the analyzed microscopic inclusion, S is interconnected with Cu, implying the presence of covellite within the body of the microfossil (see Figure 5, location Y, top red nodes), or closely attached to it. P measured in the microfossil is more interconnected with the host (quartz) elements (left blue node in network Y). This observation can be attributed to the "shouldering" effect of the intense ^{30}Si peak, which affects the integration window of P, thus modifying its ρ value and positioning in the graph. One notable feature of this correlation network is that by identifying the least and most interconnected nodes, it is possible to find predictive masses that are unique for each of the given mineralogical classes. The topology of the correlation networks also reveals a centrality measure, which indicates the importance of the element in a network. We can see that in Figure 5 left, there are some elements that are present both within the quartz and a microfossil—H, Na, K, O—these nodes could be characterized with high betweenness centrality. ^{56}Fe is also among the central elements in a network, due to the presence of the isobaric contribution of Si₂. Analyzing the microfossil-related network separately (only red

nodes), one can see that ^{12}C, ^{1}H, ^{41}K, ^{54}Fe are among the central isotopes, which also reflects the importance of this isotope in the chemical composition of the microfossil bodies. Any organic matter preserved within the bulk of the host mineral requires the existence of such networks with measurable centrality of C and H and a separate modularity rank. Depending on the chemical integrity (state of decay) of the inclusions, better preserved microfossils also have better connectivity and higher covariance, whereas finely dispersed carbon incorporated into the body of the matrix will not have such metrics and likely to be located within the outer nodes of the network.

The co-occurrence of P, Ce, and La (Figure 4B,C) within studied locations is indicative of the presence of the monazite microscopic inclusion. There are two possible interpretations for the presence of REEs in association with the microfossils: first is the intracellular incorporation and passive mineralization of La and Ce by living organisms and, second, postmortem mineralization (secondary incorporation of these elements). However, because La and Ce have been measured within our dataset only in association with microfossils, intracellular incorporation seems more plausible. The connection of Fe with S (Figure 5 left) indicates the presence of pyrite, which was described as a byproduct of the metabolic activity of sulfate-reducing microbes.[35] However, localization of metallic nodes in the lower part of the network indicated presence of the third inclusion, which likely represents Cr-rich nodule with impurities of Fe, Mn, Cu, and Ti (see bottom red nodes in Figure 5, location X). The complex chemical composition associated with some of the microfossils might also be indicative of the development of tolerance to the polymetallic toxicity. For example, Cu, registered at location Y, is an essential trace element for aerobic organisms; however, Cu might be lethal to microbes if homeostasis is not maintained.[52] Cr has also been reported as a highly toxic and mutagenic element for bacterial colonies.[53] The C and H bearing microfossils reveal close association with Na, K, Mg, Ca also with Fe and S. Overall, analysis of the inclusions from the depth profiles reveals a complex chemical composition indicating presence of chemically distinct microfossils with identifiable hydrothermal mineralization patterns (presence of typical hydrothermal elements like Cu, Fe, and Ag).

3.3 | Identification of empirical biosignatures

As it was shown in Figures 2, 4, and 5, carbon and hydrogen peak intensities are correlated with the location of the microfossils, and within our sample, carbon might be used as a tracer of the microfossils. The full depth profiling dataset sampled from the microfossil-rich locations was divided into host and microfossil data by thresholding the carbon signal. Depth profiles from the microfossil locations were sorted using a threshold of 5.8 $\log_{10}$ el/ns (higher than the noise level) to create a subset of data that represent only microfossil-related spectra, assuming that the C signal originates from the microfossils. The depth profiling dataset was additionally filtered to the depth region 500–2500 within both locations (host mineral area or aggregation of microfossils), to avoid contribution from the surface data. In total, after filtering, we formed a dataset with 12,000 spectra from the host location and 1454 spectra from the microfossil-rich location. Figure 6 presents kernel density estimates (two-dimensional density maps) calculated for two specific regions and represents a probability distribution function of element intensities for two groups: quartz and microfossils (i.e., inorganic or bioorganic intensity regions). Figure 6A–D represents the variation of the signal from the microfossils plotted against the same mass intensities measured from the host region—red kernels are calculated from the microfossil-rich location, and blue kernels from the host area (see Figure 3A,B). Multiple non-overlapping intensity regions associated with the microfossils could be observed. These intensity regions can serve as predictive borderlines for the identification of the organic remnants from other Gunflint-like cherts. In contrast, Figure 6E–G represents the variation of the elements associated with the inorganic host, shows mostly overlapping intensities, and indicates that most of the spectra from the microfossils have a significant contribution from the quartz mineral. Figure 6F shows that the Fe signal registered from the microfossils interferes with the Si_2 molecule, and significant parts of it protrude into the higher Fe content area, indicating increased Fe content within the microfossil bodies. Figure 6H displays a perfectly overlapping variation of the Gaussian background signal derived from the two locations.

Figure 7A shows partially overlapping clusters of $^{1}H/^{12}C$ and $^{16}O/^{12}C$ ratios measured from the host (blue kernels) and the microfossils (red and orange kernels). Because most of the microfossils are hollow[39] (see Figure 1) and smaller than the LIMS analytical spot size, they will be sampled with the encapsulating host mineral. In addition to the chemical composition of the microfossils, compositional details of the host mineral are likely to be registered. Hence, intensity values of ^{1}H and ^{16}O, formerly occurring within microfossils, are interfering with the same isotopes from the quartz mineral; hence, they can be subtracted. The results of these corrections are shown in Figure 7A,B with orange kernels. As can be seen from this figure, the locations of the kernels from the microfossils coincide with empirically determined
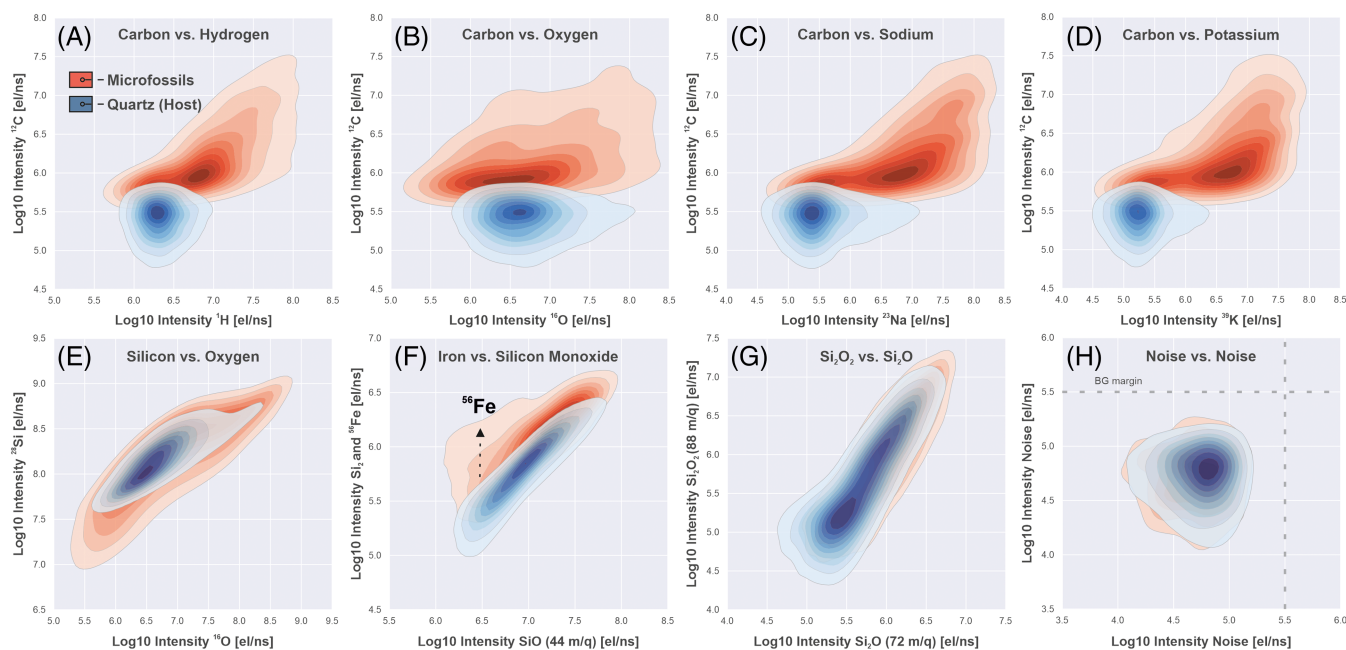
FIGURE 6 Kernel density estimates (KDE) of m/z intensities for two groups: Red kernels are calculated from the measurements performed within the microfossil-rich locations. A total of 13,454 spectra shown within each plot (1454 spectra measured from microfossils-rich location). Blue kernels are calculated from the measurements acquired from the host locations (12,000 spectra). (A)–(D) represent the total variation of C, H, O, Na, and K signal measured within microfossils and host locations. (E)–(G) represent the total variation of Si, O, Si_2 (^{56}Fe), Si_2O, and Si_2O_2 signal intensities measured from the inorganic host (quartz) and microfossil-rich locations. (H) A total variation of the noise signal; dashed line outlines conservative background margin

regions of organic compounds (lipids, peptides, sugars, and condensed hydrocarbons [kerogen]). The identification of ratio boundaries for different organic compounds was demonstrated in the literature using ultra-high-resolution mass spectrometry.[54,55] Black boxes shown in Figure 7A schematically represent the location of those boundaries. Empirically determined shapes of the complex organic compounds are derived at overlapping areas and typically have more complex shapes to those presented on the plot.

Data collected from the densely populated microfossil area (Figure 7A red and orange kernels) hint at the presence of lipid and peptide signatures, which aligns with the previously reported identification of amides from the Gunflint microfossils and exceptional preservation capacity of cherts.[34,37,56] However, they likely represent a mixing ratio between the original kerogen and quartz ratios. A fraction of the data also intrudes into the area of kerogens, which can be classified as a Kerogen Type I (Algal).[57] However, there is a part of the red kernels that overlap with the host mineral data as well. Data collected from the host area contain mostly quartz mineral, and most of the measured ratios represent signal-to-noise ratios of ^{16}O and ^{1}H, due to the low concentration of ^{12}C within analyzed depth profiles from the host locations. The results presented in Figure 7A indicate that organic hydrocarbons measured at the Gunflint sample are significantly reduced in ^{16}O and relatively saturated in ^{1}H (increased abundance of ^{1}H). A shift of the $^{1}H/^{12}C$ values towards relative ^{1}H saturation (mentioned in the plot as hydrogenation) can also be promoted by the different optical absorption of the laser energy. Different absorption rates might shift the total energy balance for ablation and ionization of the microfossils, consecutively resulting in a more efficient release of ^{1}H ions. Whereas a significant part of the ablated and ionized material can also be expected to represent a mixture of the signal from the microfossils and quartz mineral (due to the effective diameter of the focused laser), and it might be challenging to subtract the contribution of the isobaric input from the quartz entirely. Figure 7B, similarly to the Van Krevelen plot (Figure 7A), represents bivariate distribution of $^{28}Si/^{12}C$ and $^{16}O/^{12}C$ values and indicates presence of the separate ratio space characteristic of microfossils. The same correction procedure as in Figure 7A has been applied for $^{16}O/^{12}C$ distribution, which is shown with orange kernels.

Figure 7C,D shows the distribution of PCA values obtained from the two distinct groups—measurements from the host area in contrast to the measurements from the microfossil-rich zone (12,863 spectra in total). Spectra from the microfossil location could be separated from the spectra collected within the host area, except for a small number,
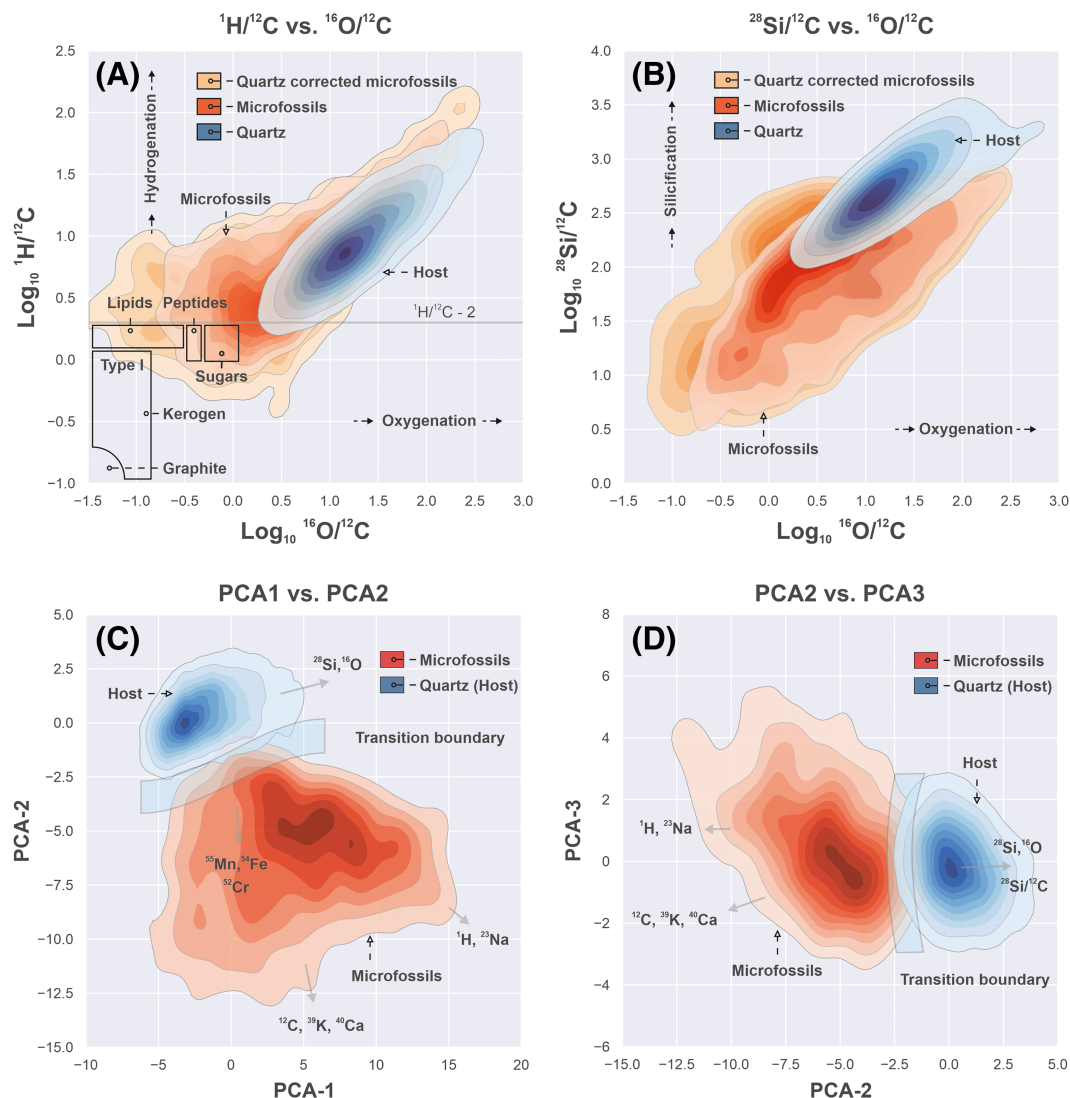
FIGURE 7 (A) Kernel density estimates (KDE) of $^1H/^{12}C$ and $^{16}O/^{12}C$ ratios (Van Krevelen plot). Blue kernels represent values measured from the host area. Red kernels show the distribution of $^1H/^{12}C$ and $^{16}O/^{12}C$ values calculated from the area densely populated with microfossils. Orange kernels represent the same data, as in red kernels, but corrected for ^1H and ^{16}O interfering signal, registered from the co-occurring with microfossils quartz mineral (details about correction procedure described further in the text). Black bars: Schematic representation of the boundaries of lipids, peptides, sugars, and kerogen within the Van Krevelen space (values are taken from the literature[45–47]). (B) Kernel density estimates (KDE) of $^{28}Si/^{12}C$ and $^{16}O/^{12}C$ ratios. Blue kernels represent values measured from the host area. Red kernels show the distribution of $^{28}Si/^{12}C$ and $^{16}O/^{12}C$ values measured from the area densely populated with microfossils. Orange kernels represent the same data, as in red kernels, but corrected for ^{16}O interfering signal, registered from the co-occurring with microfossils quartz mineral. (C and D) 2D projections of the 196 features into the principal components space. Blue kernels represent data from the host locations. Red kernels represent data collected from the microfossil-rich zone. Arrows schematically represent loading directions for different single unit masses. The blue transparent line indicates the location of the transition boundary between two classes. Among 196 features, 179 are normalized intensities (z-scores) of single unit masses, 15 features are normalized (z-scores) element ratios, and two last features are z-scores of averaged intensities of light masses (1–74 m/z) and geometrical mean values for light masses (1–74 m/z)

which interferes with the host measurements. To increase the separability of the dataset, we further thresholded C intensities to 6 $\log_{10}$ el/ns (see supporting information), which reduced the microfossils dataset from 1454 spectra to 863 spectra. We schematically identified the transition boundary with a light blue transparent line, which shows the location of the estimated transition boundary between the inorganic host (quartz) spectra and spectra from the microfossils. The transition of one class into another could be explained by the ablation of the small portions of the microfossils (nm thick cell walls in the bulk of the host mineral). The thickness of the rims of the collapsed cell walls reported from the Gunflint microfossils varies from the tens of nm[39] to first micrometers,[35] which might explain statistically

more prevalent low-intensity regions in KDE plots (Figures 6 and 7). Figure 7C,D indicates that intermediate levels between classes are sampled when spectra from one class gradually turn into the spectra from the other class. It is worth noting that even with only the first two principal components, a clear separation between two main classes can be seen. Additionally, as was demonstrated before (Figures 4B,C and 5), and noting the dispersion of the PCA loadings, it is possible to identify that there are potentially more than one class of microfossils within the Gunflint dataset. A significant part of the dataset is clustered within the dark red areas, representing the majority of simple kerogen containing microfossils. However, we could see from Figure 7C that kernels protrude from the hydrocarbon saturated area towards areas with notable metallic content, pointing towards the presence of uptake of Fe, Mn, and Cr. This observation agrees with the results reported previously on the diversity of the microbiome within Gunflint waters. Cyanobacteria were proposed to be a dominant part of the Gunflint stromatolites[31,40,58]; however, other interpretations are possible. Presence of Mg in the spectra (Figures 4B,C and 5) can indicate the presence of degradation products of chlorophyll, because all chlorophyll molecules share chlorin magnesium ligand in their structure, supporting the photosynthetic hypothesis. A community of saprophytic heterotrophs was proposed as part of the microbiome,[35] which are assumed to have a different set of metallic catalysts and enzymes, hence, identifiable chemical fingerprints. Nevertheless, clear separation of the chemically distinct subclasses of microfossils requires even higher statistics and linkage to the morphological features.

Overall, the broad set of mass spectrometric characteristics, measured from the Gunflint microfossils, can be identified using the full feature space, with the ML classification models (see Table S1). By applying an ensemble classification algorithm (adaptive boosting), we achieved a 99.7% separation rate between classes: inorganic host (quartz) or organic inclusions (microfossils). A small subset of misclassifications is attributed to the transition boundary line, where spectra are protruding from the host region towards microfossils. By extracting the 196 features from the single mass spectra, we created 19,110 unique sets of mass and ratio pairs, in which two classes might be separated (i.e., C vs. H, and C vs. O). It is possible to achieve a full separation score of 100% by further limiting the C intensities; however, this will affect the quality of the learned borderlines between classes. Such models, containing the empirical biosignatures from the known terrestrial samples with a proven biogenic origin, might be used as a deployable solution onboard of the Martian rovers, providing an additional line of evidence towards establishing the biogenicity of a given putative sample by assessing the proximity of the data to the spectra acquired from the Gunflint and/or other model samples. However, future work will be required to identify the capacity of LIMS system for distinguishing between true biological organic material (life) as opposed to organic material that was produced non-biologically, for example, via Fischer-Tropsch-Type (FTT) synthesis.

4 | CONCLUSION

In summary, the chemical composition of Precambrian microfossils from the 1.88-Ga Gunflint Formation was investigated using a laser-based miniature TOF-MS. Locations of microfossils were identified on the surface of the sample utilizing MSI. The composition of individual microfossils embedded within the chert was identified using depth profiling and single mass unit spectral decomposition. Utilizing MSI, weighted mass correlation networks, isotopic ratios, and projection of intensities into the low dimension using PCA, it was demonstrated that the microfossils, and associated with them mineralization, have a unique chemical composition that could be identified from the LIMS microprobe data. High-throughput LIMS imaging combined with depth profiling has been shown to be capable of yielding new insights into the distribution, preservation, and elemental speciation of the microfossils in Precambrian cherts.

CONFLICT OF INTEREST
No competing interests exist.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3370.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Rustam A. Lukmanov https://orcid.org/0000-0001-9257-7410
Andreas Riedo https://orcid.org/0000-0001-9007-5791

REFERENCES

1. McMahon S, Bosak T, Grotzinger JP, et al. *A Field Guide to Finding Fossils on Mars*. Hoboken, New Jersey: John Wiley & Sons, Ltd; 2018:1012-1040.
2. Westall F, Foucher F, Bost N, et al. Biosignatures on Mars: what, where, and how? Implications for the search for Martian life. *Astrobiology*. 2015;15(11):998-1029.
3. Vago JL, Westall F, Pasteur Instrument Teams, Landing S, et al. Habitability on early Mars and the search for biosignatures with the ExoMars rover. *Astrobiology*. 2017;17(6-7):471-510.
4. Orosei R, Lauro SE, Pettinelli E, et al. Radar evidence of subglacial liquid water on Mars. *Science*. 2018;361(6401):490-493.
5. Grotzinger JP, Sumner DY, Kah LC, et al. A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars. *Science*. 2014;343(6169):1242777.
6. Navarro-González R, Navarro KF, de la Rosa J, et al. The limitations on organic detection in Mars-like soils by thermal volatilization-gas chromatography-MS and their implications for the Viking results. *Proc Natl Acad Sci U S A*. 2006;103(44):16089-16094.
7. Wurz P, Abplanalp D, Tulej M, et al. Mass spectrometric analysis in planetary science: investigation of the surface and the atmosphere. *Sol Syst Res*. 2012;46(6):408-422.
8. Hofmann BA. Morphological biosignatures from subsurface environments: recognition on planetary missions. *Space Sci Rev*. 2008; 135(1-4):245-254.
9. Brasier MD, Wacey D. Fossils and astrobiology: new protocols for cell evolution in deep time. *Int J Astrobiol*. 2012;11(4):217-228.
10. Stevens AH, McDonald A, de Koning C, et al. Detectability of biosignatures in a low-biomass simulation of martian sediments. *Sci Rep*. 2019;9(1):9706.
11. Riedo A, de Koning C, Stevens AH, et al. The detection of elemental signatures of microbes in martian mudstone analogs using high spatial resolution laser ablation ionization mass spectrometry. *Astrobiology*. 2020;20(10):1224-1235.
12. Wiesendanger R, Wacey D, Tulej M, et al. Chemical and optical identification of micrometer-sized 1.9 billion-year-old fossils by combining a miniature laser ablation ionization mass spectrometry system with an optical microscope. *Astrobiology*. 2018;18(8):1071-1080.
13. Veneranda M, Lopez-Reyes G, Manrique-Martinez JA, et al. ExoMars Raman laser spectrometer (RLS): development of chemometric tools to classify ultramafic igneous rocks on Mars. *Sci Rep*. 2020;10:1-14.
14. Smith JP, Smith FC, Booksh KS. Spatial and spectral resolution of carbonaceous material from hematite (α-Fe_2O_3) using multivariate curve resolution-alternating least squares (MCR-ALS) with Raman microspectroscopic mapping: implications for the search for life on Mars. *Analyst*. 2017;142(17):3140-3156.
15. Smith JP, Smith FC, Booksh KS. Multivariate curve resolution–alternating least squares (MCR-ALS) with Raman imaging applied to lunar meteorites. *Appl Spectrosc*. 2018;72(3):404-419.
16. Konstantinidis M, Lalla EA, Daly MG, et al. Elemental estimation of terrestrial analogues from the CanMars rover field campaign using LiRS: implications for detecting silica-rich deposits on Mars. *Icarus*. 2021;358:114113-114124.
17. Manrique-Martinez JA, Lopez-Reyes G, Alvarez-Perez A, et al. Evaluation of multivariate analyses and data fusion between Raman and laser-induced breakdown spectroscopy in binary mixtures and its potential for solar system exploration. *J Raman Spectrosc*. 2020;51(9): 1702-1717.
18. Ligterink NFW, Grimaudo V, Moreno-García P, et al. ORIGIN: a novel and compact laser desorption–mass spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces. *Sci Rep*. 2020;10(1):9641.
19. Grimaudo V, Tulej M, Riedo A, et al. UV post-ionization laser ablation ionization mass spectrometry for improved nm-depth profiling resolution on Cr/Ni reference standard. *Rapid Commun Mass Spectrom*. 2020;34:e8803.
20. Riedo A, Grimaudo V, Moreno-García P, et al. Laser ablation/ionisation mass spectrometry: sensitive and quantitative chemical depth profiling of solid materials. *Chimia*. 2016;70(4):268-273.
21. Riedo A, Bieler A, Neuland M, Tulej M, Wurz P. Performance evaluation of a miniature laser ablation time-of-flight mass spectrometer designed for in situ investigations in planetary space research. *J Mass Spectrom*. 2013;48(1):1-15.
22. Wiesendanger R, Tulej M, Riedo A, Frey S, Shea H, Wurz P. Improved detection sensitivity for heavy trace elements using a miniature laser ablation ionisation mass spectrometer. *J Anal At Spectrom*. 2017;32(11):2182-2188.
23. Tulej M, Riedo A, Neuland MB, et al. CAMAM: a miniature laser ablation ionisation mass spectrometer and microscope-camera system for in situ investigation of the composition and morphology of extraterrestrial materials. *Geostand Geoanal Res*. 2014;38(4):441-466.
24. Goesmann F, Brinckerhoff WB, Raulin F, et al. The Mars Organic Molecule Analyzer (MOMA) instrument: characterization of organic material in martian sediments. *Astrobiology*. 2017;17(6-7):655-685.
25. Li X, Danell RM, Pinnick VT, et al. Mars Organic Molecule Analyzer (MOMA) laser desorption/ionization source design and performance characterization. *Int J Mass Spectrom*. 2017;422:177-187.

26. Briois C, Thissen R, Thirkell L, 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.

27. Selliez L, Briois C, Carrasco N, et al. Identification of organic molecules with a laboratory prototype based on the Laser Ablation-CosmOrbitrap. *Planet Space Sci*. 2019;170:42-51.

28. Arevalo R Jr, Selliez L, Briois C, et al. An Orbitrap-based laser desorption/ablation mass spectrometer designed for spaceflight. *Rapid Commun Mass Spectrom*. 2018;32(21):1875-1886.

29. Tulej M, Neubeck A, Ivarsson M, et al. Chemical composition of micrometer-sized filaments in an aragonite host by a miniature laser ablation/ionization mass spectrometer. *Astrobiology*. 2015;15(8):669-682.

30. Neuland MB, Grimaudo V, Mezger K, et al. Quantitative measurement of the chemical composition of geological standards with a miniature laser ablation/ionization mass spectrometer designed for in situ application in space research. *Meas Sci Technol*. 2016;27(3): 035904.

31. Barghoorn ES, Tyler SA. Microorganisms from the Gunflint chert: these structurally preserved Precambrian fossils from Ontario are the most ancient organisms known. *Science*. 1965;147(3658):563-575.

32. Awramik SM, Semikhatov MA. The relationship between morphology, microstructure, and microbiota in three vertically intergrading stromatolites from the Gunflint Iron Formation. *Can J Earth Sci*. 1979;16(3):484-495.

33. Cloud PE. Significance of the Gunflint (Precambrian) microflora. *Science*. 1965;148:563-575.

34. Alleon J, Bernard S, Le Guillou C, et al. Molecular preservation of 1.88 Ga Gunflint organic microfossils as a function of temperature and mineralogy. *Nat Commun*. 2017;7:11977.

35. Wacey D, McLoughlin N, Kilburn MR, et al. Nanoscale analysis of pyritized microfossils reveals differential heterotrophic consumption in the ∼1.9-Ga Gunflint chert. *PNAS*. 2013;110:1053-1060.

36. Cady SL, Skok JR, Gulick VG, Berger JA, Hinman NW. Siliceous hot spring deposits: why they remain key astrobiological targets. In: *From Habitability to Life on Mars*. Amsterdam, Netherlands: Elsevier; 2018:179-210.

37. Alleon J, Bernard S, le Guillou C, et al. Early entombment within silica minimizes the molecular degradation of microorganisms during advanced diagenesis. *Chem Geol*. 2016;437:98-108.

38. Schelble RT, Westall F, Allen CC. ∼1.8 Ga iron-mineralized microbiota from the Gunflint Iron Formation, Ontario, Canada: implications for Mars. *Adv Space Res*. 2004;33:1268-1273.

39. Lepot K, Addad A, Knoll AH, et al. Iron minerals within specific microfossil morphospecies of the 1.88 Ga Gunflint Formation. *Nat Commun*. 2017;8(1):1-11. https://doi.org/10.1038/ncomms14890

40. Awramik SM, Barghoorn ES. The Gunflint microbiota. *Precambrian Res*. 1977;5(2):121-142.

41. House CH, Schopf JW, McKeegan KD, Coath CD, Harrison TM, Stetter KO. Carbon isotopic composition of individual Precambrian microfossils. *Geology*. 2000;28(8):707-710.

42. Javaux EJ, Lepot K. The Paleoproterozoic fossil record: implications for the evolution of the biosphere during Earth's middle-age. *Earth Sci Rev*. 2017;176:68-86.

43. Azov VA, Mueller L, Makarov AA. Laser ionization mass spectrometry at 55: Quo Vadis? *Mass Spectrom Rev*. 2020. https://doi.org/10.1002/mas.21669

44. Tulej M, Ligterink NFW, de Koning C, et al. Current progress in femtosecond laser ablation/ionisation time-of-flight mass spectrometry. *Applied Sciences*. 2021;11(6):2562.

45. Riedo A, Meyer S, Heredia B, et al. Highly accurate isotope composition measurements by a miniature laser ablation mass spectrometer designed for in situ investigations on planetary surfaces. *Planet Space Sci*. 2013;87:1-13.

46. Riedo A, Tulej M, Rohner U, Wurz P. High-speed microstrip multi-anode multichannel plate detector system. *Rev Sci Instrum*. 2017; 88(4):045114.

47. Moreno-García P, Grimaudo V, Riedo A, Tulej M, Wurz P, Broekmann P. Towards matrix-free femtosecond-laser desorption mass spectrometry for in situ space research. *Rapid Commun Mass Spectrom*. 2016;30(8):1031-1036.

48. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Palo Alto, California: Association for the Advancement of Artificial Intelligence (AAAI); 2009.

49. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*. 2014;9(6):e98679.

50. Meyer S, Riedo A, Neuland MB, Tulej M, Wurz P. Fully automatic and precise data analysis developed for time-of-flight mass spectrometry. *J Mass Spectrom*. 2017;52(9):580-590.

51. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp*. 2008; 2008(10):P10008.

52. Ladomersky E, Petris MJ. Copper tolerance and virulence in bacteria. *Metallomics*. 2015;7(6):957-964.

53. Viti C, Marchi E, Decorosi F, Giovannetti L. Molecular mechanisms of Cr(VI) resistance in bacteria and fungi. *FEMS Microbiol Rev*. 2014;38(4):633-659.

54. Brockman SA, Roden EV, Hegeman AD. Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. *Metabolomics*. 2018;14(4):48–53.

55. Kim S, Kramer RW, Hatcher PG. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. *Anal Chem*. 2003;75(20):5336-5344.

56. Alleon J, Bernard S, Le Guillou C, et al. Organic molecular heterogeneities can withstand diagenesis. *Sci Rep*. 2017;7:1508.

57. Hutton A, Bharati S, Robl T. Chemical and petrographic classification of kerogen/macerals. *Energy Fuel*. 1994;8(6):1478-1488.

58. Boal D, Ng R. Shape analysis of filamentous Precambrian microfossils and modern cyanobacteria. *Paleobiology*. 2010;36(518):555-572.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

3.2.2 On topological analysis of fs-LIMS data. Implications for in situ planetary mass spectrometry (Frontiers in Artificial Intelligence)

Rustam A. Lukmanov[1*], Andreas Riedo[1], David Wacey[2], Niels F.W. Ligterink[1], Valentine Grimaudo[1], Marek Tulej[1], Coenraad de Koning[1], Anna Neubeck[3], Peter Wurz[1]

[1]Space Research and Planetary Sciences (WP), University of Bern, Bern, Switzerland
[2]Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, Perth, Australia
[3]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

*Correspondence:
Rustam A. Lukmanov: rustam.lukmanov@space.unibe.ch

On Topological Analysis of fs-LIMS Data. Implications for in Situ Planetary Mass Spectrometry

Rustam A. Lukmanov[1]*, Andreas Riedo[1], David Wacey[2], Niels F. W. Ligterink[1], Valentine Grimaudo[1], Marek Tulej[1], Coenraad de Koning[1], Anna Neubeck[3] and Peter Wurz[1]

[1]Space Research and Planetary Sciences (WP), University of Bern, Bern, Switzerland, [2]Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, Perth, WA, Australia, [3]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

In this contribution, we present results of non-linear dimensionality reduction and classification of the fs laser ablation ionization mass spectrometry (LIMS) imaging dataset acquired from the Precambrian Gunflint chert (1.88 Ga) using a miniature time-of-flight mass spectrometer developed for *in situ* space applications. We discuss the data generation, processing, and analysis pipeline for the classification of the recorded fs-LIMS mass spectra. Further, we define topological biosignatures identified for Precambrian Gunflint microfossils by projecting the recorded fs-LIMS intensity space into low dimensions. Two distinct subtypes of microfossil-related spectra, a layer of organic contamination and inorganic quartz matrix were identified using the fs-LIMS data. The topological analysis applied to the fs-LIMS data allows to gain additional knowledge from large datasets, formulate hypotheses and quickly generate insights from spectral data. Our contribution illustrates the utility of applying spatially resolved mass spectrometry in combination with topology-based analytics in detecting signatures of early (primitive) life. Our results indicate that fs-LIMS, in combination with topological methods, provides a powerful analytical framework and could be applied to the study of other complex mineralogical samples.

Keywords: fs-LIMS, mass-spectrometry, UMAP (uniform manifold approximation and projection), mapper, microfossils, mars, Gunflint

INTRODUCTION

The current state of space exploration is on the verge of new frontiers, holding promise for discoveries on other planetary bodies through *in-situ* robotic exploration (Vago et al., 2015). For example, Mars and the icy moons of Jupiter and Saturn, once thought to be lifeless, have gained more attention from the scientific community in recent decades due to new data informing upon the potential habitability of these bodies (Priscu and Hand 2012; Garcia-Lopez and Cid 2017; McMahon et al., 2018). Thus, there is an ongoing need for sensitive and high-throughput space instrumentation providing precise analytical data on a microscale (Navarro-González et al., 2006; Goesmann et al., 2017). However, space-type instruments are usually small and provide only a fraction of the sensitivity and overall capability of their full-scale laboratory counterparts. Reduction in performance occurs due to the strict constraints on size, power consumption, and weight of the scientific payload. Therefore, the development of new miniature instruments and analytical methods with improved capabilities is a continuously pressing issue (Li et al., 2017; Arevalo et al., 2018; Stevens et al., 2019; Wurz et al., 2020).

Laser-based mass spectrometry (Laser Ablation Ionization and Desorption Mass Spectrometry–LIMS and LDMS) is a modern and compact analytical method that promises to greatly enhance the quality of chemical analysis on planetary bodies (Riedo et al., 2013b; Arevalo et al., 2018). The first LIMS instrument selected and built for a planetary lander was LASMA, developed for the Phobos-Grunt mission (Managadze et al., 2010). Recently, the second LIMS instrument was chosen for the upcoming ExoMars mission/Rosalind Franklin Rover (Goesmann et al., 2017), further facilitating developments in this field. Laser-based mass spectrometry, developed for *in-situ* planetary exploration, as a versatile method, can provide a description of molecular composition and element, isotope characterization of solids (Moreno-García et al., 2016; Arevalo et al., 2020; Tulej et al., 2020). The time-of-flight version of LDMS has been shown to be capable of measuring extremely low concentrations (fmole) of amino acids in the desorption mode (Ligterink et al., 2020). LIMS modification of this instrument has been reported to measure ppbw level trace elements and routinely measure fine chemistry from a variety of samples (Riedo et al., 2013a; Neuland et al., 2016; Wiesendanger et al., 2017). Moreover, a number of reports have indicated that LIMS, particularly fs-LIMS, might be applicable to the detection of faint signatures of life from microscopic inclusions (Tulej et al., 2015; Wiesendanger et al., 2018) and low-biomass Martian analogs (Stevens et al., 2019; Riedo et al., 2020). However, the field of study of early and primitive life remains profoundly complex (Brasier and Wacey 2012; Westall et al., 2015; Wacey et al., 2016) with no single chemical criterion that can be assigned as definitive proof of biogenicity. A number of authors have proposed a multi-criteria approach, where a multitude of methods needs to be applied before any conclusions can be drawn (Hofmann 2008; Brasier and Wacey 2012; Hand et al., 2017; Vago et al., 2017; Neveu et al., 2018; Chan et al., 2019). The multi-method approach enhances the size of parametric space and reduces the probability of false-positive detection. Therefore, any advancement within each of the applied methods can increase the overall confidence of the correct identification of signatures of life.

In this contribution, we hypothesize that on the basis of the full feature scale (mass range) present in the fs-LIMS spectral datasets, it is possible to identify minerals and compounds of specific chemistry using an unsupervised data-driven approach. We describe a topology-based analysis pipeline to define the complexity of the fs-LIMS imaging data in low dimensions and identify groups of spectra that share a significant degree of similarity. We apply the aforementioned method to 18,000 composite spectra acquired from the Gunflint chert (1.88 Ga), which contains populations of well-preserved Precambrian microfossils of proven biological origin (Wacey et al., 2013). The analysis of the data reveals four distinct populations of fs-LIMS spectra, which correspond to two groups of microfossils, the quartz matrix in which microfossils are entombed and organic surface contamination spectra. Moreover, we describe a fine transitional structure between classes and argue that low dimensional representations are of high utility in *in-situ* mass-spectrometry and space research. Further, we speculate that our approach is scalable to non-space instruments and may,

therefore, prove useful in the field of Precambrian micropaleontology and high-dimensional analytical chemistry in general.

METHODS

In this study, we use laser ablation ionization mass spectrometry for the characterization of the chemical composition of the Gunflint sample and optical microscopy to identify morphological features. A thorough review of LIMS operation principles, current state-of-the-art, and application case studies can be found in a number of previous reports (Tulej et al., 2014; Wiesendanger et al., 2017; Grimaudo et al., 2020; Ligterink et al., 2020) and reviews (Grimaudo et al., 2019; Azov et al., 2020), and therefore, only a short description will be given here. In the simplest case, LIMS instruments include two main parts–a pulsed laser system to ablate and ionize materials and a mass analyzer to separate and register ions produced during the ablation and ionization process. The fs-LIMS is a successor of ns-LIMS, with the only difference that the mass analyzer is coupled to the fs laser system. Current commercial fs lasers can provide peak power fluences up to terawatt/cm^2, compressed to very short pulses of femto-second duration. Such high powers can ionize any material, thus, providing means for an isotope and element characterization of any solid with very small detection limits and reduced matrix effects (Riedo et al., 2013c). As an ion source, we have installed a Ti:Sapphire laser with chirped pulse amplification, which provides a stable IR-775 nm, ~190 fs laser radiation. Conversion of the fundamental wavelength from IR-775 nm to UV-258 nm was made using a commercially available third-harmonic generation module.

The fs-LIMS system used in this study consists of a miniature time-of-flight (TOF) mass analyzer (⌀ 60 × 160 mm) (*see* **Figure 1**) with an axially symmetric design and single unit mass resolution. The instrument was developed for *in-situ* space applications, and due to its miniature design it could be placed on a rover, lander, or even used as a handheld instrument (Wurz et al., 2020). In normal operation mode, fs-LIMS could identify major chemical composition along with ppmw-level concentrations of trace elements. As shown in **Figure 1**, the TOF mass analyzer consists of entrance ion optics (where ions are confined and accelerated), a drift tube (where ions experience mass/charge separation), a reflectron (ion mirror which uses an electric field), and a microchannel plate (MCP) detector system (Riedo et al., 2017) to register ion flux. The schematic illustration of the fs-LIMS sample analysis is shown in **Figure 1B**. The focused blue light indicates the fs-UV-258 nm laser radiation that passes through the instrument and ablates the small area of the sample with a diameter of the ablation spot of about 5 µm. The positioning of the ablation spots is determined by the internal microscopy system. The objective of the microscope is located at a fixed offset from the instrument. After ablation and ionization, positively charged ions are guided by an electric field of the instrument into a defined parabolic trajectory so that every ion that enters the instrument will land on the surface of the detector. Incoming time-separated ion flux launches an electron avalanche
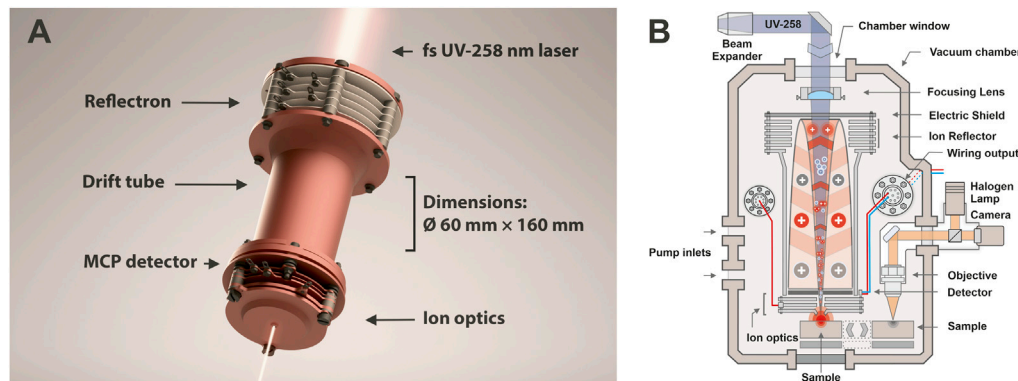
FIGURE 1 | (A) 3D render of our miniature time-of-flight mass analyzer. Location of the reflectron, drift tube, entrance ion optics, MCP detector, and dimensions of the instrument are denoted. The focusing fs-UV laser light shown on the top and the bottom and illustrates an axial design of the mass analyzer. Sample positioning is not shown. However, in the laboratory setting, the investigated sample is positioned in close vicinity to the entrance plate of the ion optical system of the mass analyzer, right in the position of the laser focus, to achieve ablation and subsequent ionization of target material. **(B)** Schematic illustration of the fs-LIMS. An fs-laser radiation (blue line) ablates and ionizes material from the sample. The positively charged ions are separated and detected using the time-of-flight mass spectrometer. The ablation position can be precisely located using an integrated microscopy system.
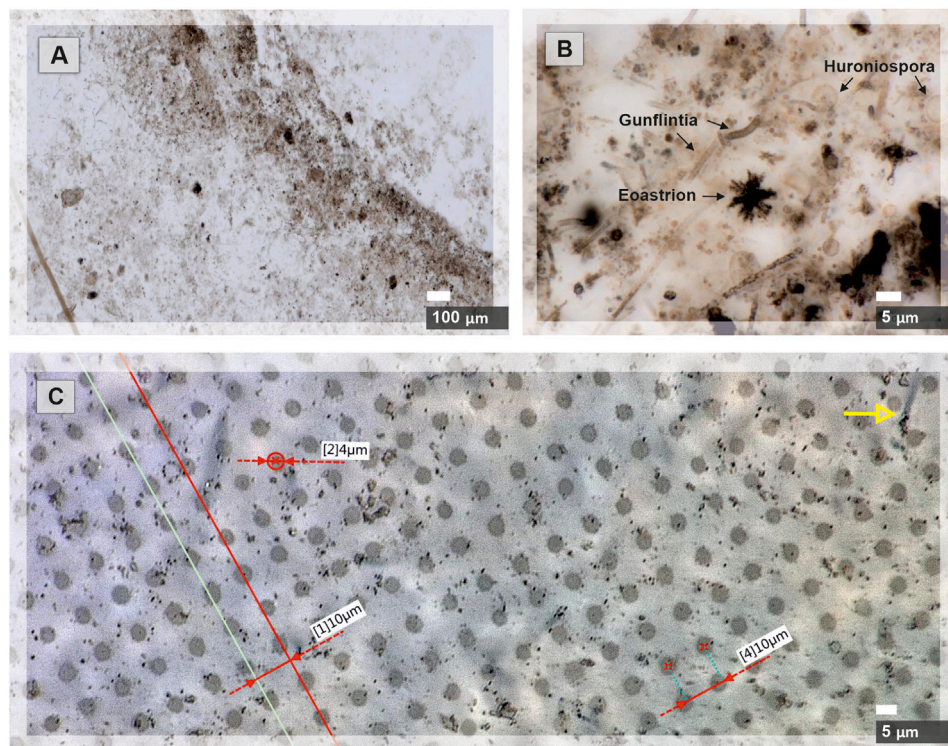

FIGURE 2 | Microscope images of the Gunflint chert before and after the fs-LIMS imaging campaign are shown. **(A)** Microscope image of the area (0.9*2 mm^2) chosen for the chemical imaging with our fs-LIMS system. The dark brown patches distributed through the sample and forming a diffuse layer in the middle of the picture represent a bio-lamination surface. **(B)** Close-up microscope picture of individual microfossils from the bio-lamination surface. Filamentous (Gunflintia), star-shaped (Eoastrion), and spherular microfossils (Huroniospora) can be seen. **(C)** Microscope picture of laser ablation craters (0.9*2 mm^2 area covered with 90*200 pixels–18,000 ablation positions) formed after the fs-LIMS imaging campaign. Red lines denote the accuracy of sample positioning (the gap between ablation craters is consistently 10 μm) and identify the ablation crater diameters. Individual craters range in diameter from 4 to 5 μm. Note, on the upper part of the image, the yellow arrow indicates an individual microfossil body. The size of the microfossil can be compared with the diameter of the analytical spot.

within the microchannels of the detector system and creates a measurable current on the output anodes. Thus, time-of-flight LIMS measures an output current per unit of time, which is correlated with the element and isotope abundances of an investigated spot. Note that the image of the fs-laser beam passing through the instrument (**Figure 1A**) is exaggerated - in

the laboratory setting, the laser focal point is located in close proximity to the entrance electrode of the mass analyzer (see **Figure 1B**).

The investigation of a 30 μm thick thin-section of Gunflint chert has been conducted with our miniature fs-LIMS system. The sample acquired from the Gunflint Formation (Schreiber beach locality, Ontario, Canada; Wacey et al., 2012, 2013) represents a finely polished thin slice of the original rock, glued to the glass substrate and mounted on a steel holder. Preliminary optical microscopy was performed on the sample to identify specific areas of microfossil aggregation (see **Figure 2A**). Matrix material in which microfossils are preserved was identified to be microcrystalline quartz. Chemical imaging of the rectangular area, containing a bio-lamination surface (aggregation of microfossils within a stromatolite) and a clear host area (quartz filled matrix) was done with the LIMS system using the fs UV-258 nm laser, which provides a flux of 4.8 eV UV photons, which is well suited for ionization of glasses and other non-conductive materials with low absorption coefficients.

Figure 1A. 3D rendering of our miniature time-of-flight mass analyzer. Location of the reflectron, drift tube, entrance ion optics, MCP detector, and dimensions of the instrument are denoted. The focusing fs-UV laser light shown on the top and the bottom and illustrates an axial design of the mass analyzer. Sample positioning is not shown. However, in the laboratory setting, the investigated sample is positioned in close vicinity to the entrance plate of the ion optical system of the mass analyzer, right in the position of the laser focus, to achieve ablation and subsequent ionization of target material. **Figure 1B**. Schematic illustration of the fs-LIMS. The fs-laser radiation (blue line) ablates and ionizes material from the sample. The positively charged ions are separated and detected using the time-of-flight mass spectrometer. The ablation position can be precisely located using the integrated microscopy system.

Data Acquisition

A rectangular area of 0.9×2 mm^2 was investigated using the fs-LIMS system (see **Figure 2A**). A relatively low number of laser pulses were applied to each surface position – 200 laser shots, to avoid material displacement and crater-to-crater cross-contamination. The spatially resolved measurements conducted on the Gunflint chert resulted in the collection of 18,000 composite spectra (collected from the grid - 90 by 200 position or ablation sites). A composite spectrum collected from the given position (or ablation site) resulted in the accumulation of 200 single-shot spectra, with 64,000 data points digitized per spectrum. Thus, the total number of registered shots resulted in 3.6×10^6 single-shot spectra. The laser energies applied to each position amounted to ~360 nJ/pulse (measured at the sample surface) using UV-258 nm laser. This energy was appropriate to produce the optimal quality for the mass-spectrometric signal, both from the microfossils and the quartz-filled host area. Analytical conditions were held constant during the data collection. The diameter of the average ablation crater was measured to be ~5 μm, and gaps between ablation craters were set to 10 μm (see **Figure 2C**). A custom-built software package

was used to control the translation stage and the laser firing intervals. A fast data acquisition system from Keysight was used for digitizing current from the anodes, providing a 3.2 GSa/s sampling rate. An example of a single composite spectrum (representing a histogram of 200 individual single-shot spectra) registered from the Gunflint sample is shown in **Figure 3**. A single mass spectrum consists of 64,000 individual datapoints sampled with a digitizer, where each digitized data point corresponds to ~0.33 ns of a flight time. Thus, every recorded spectrum contains information about ~20 μs of a flight time, which provides a mass/charge (m/z) coverage of up to 800 m/z, providing a complete record of all stable isotopes and simple molecular compounds. Overall, 18,000 composite spectra were collected from the Gunflint sample, with a 10 μm gap between ablation craters. Additionally to the mass spectra collection, noise measurements were recorded, which allowed the enhancement of the recorded signal.

Figure 2. Microscope images of the Gunflint chert before and after the fs-LIMS imaging campaign are shown. A) Microscope image of the area (0.9*2 mm^2) chosen for the chemical imaging with our fs-LIMS system. A. The dark brown patches distributed through the sample and forming a diffuse layer in the middle of the picture represent a bio-lamination surface. B) Close-up microscope picture of individual microfossils from the bio-lamination surface. Filamentous (Gunflintia), star-shaped (Eoastrion), and spherular microfossils (Huroniospora) can be seen. C) Microscope picture of laser ablation craters (0.9*2 mm^2 area covered with 90*200 positions – 18,000 ablation sites) formed after the fs-LIMS imaging campaign. Red lines denote the accuracy of sample positioning (the gap between ablation craters is consistently 10 μm) and identify the ablation crater diameters. Individual craters range in diameter from 4 to 5 μm. Note, on the upper part of the image, the yellow arrow indicates an individual microfossil body. The size of the microfossil can be compared with the diameter of the analytical spot.

Data Preprocessing

The entire imaging dataset, which consists of ~50 GB of recorded composite mass spectra, was preprocessed before any analysis was applied to the data. A mass spectrometry preprocessing routine applied to the dataset consists of several typical steps that largely follow methods described in (Gil and Marco 2007) and (Meyer et al., 2017). The fs-LIMS preprocessing routine applied to the imaging data consisted of:

1) Noise removal for an improvement of the signal-to-noise ratio (SNR) of the signal. The noise signal (empty composite mass spectra) was recorded after the imaging campaign was completed. The recorded noise waveform was subtracted from the imaging observations.

2) Baseline subtraction. A filter function was applied to the noise-removed mass spectra to estimate varying baseline within multiple windows and regressed using spline approximation.

3) Jitter correction. Since materials within the analyzed sample might be of better or worse ionization efficiency (mainly due to topography), temporal variation of ion yields is expected to
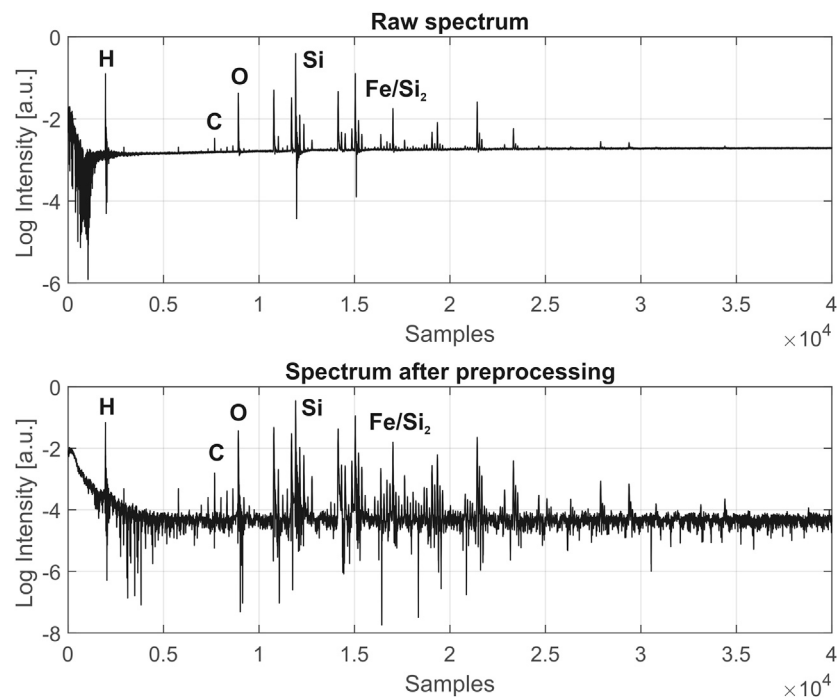
FIGURE 3 | Comparison of fs-LIMS spectra (composite spectrum of 200 laser shots, recorded from single position), before and after data preprocessing, acquired from the Gunflint chert sample. Each line in the spectrum represents a single unit mass. The increase of SNR to 10^4 and correction of the baseline can be noted. *See* the text for the full description of preprocessing procedures. Exemplary atomic lines are denoted on top of the spectrum.

occur so that times-of-flight of given ion packets might slightly vary. Typically, this effect is small and affects the peak shapes in a minor way. However, since we collected a relatively large dataset, a correction procedure has been applied. To correct for mismatch of times-of-flight, we have used an autocorrelation function described in (Gil and Marco 2007).

4) Low pass filtering. The low pass filter with normalized cutoff frequency at 0.13 πrad/sample and stopband attenuation of 60 dB was applied to each composite mass spectrum. This step removes the remaining high-frequency component from the recorded signal. Typically, it improves the SNR by two to five and does not alter the peak shapes.

5) Parametric peak preserving smoothing. The Savitzky–Golay filter function (Press and Teukolsky 1990) was applied to flatten the baseline further and increase the SNR.

6) Mass scale assignment. An average time-of-flight spectrum of all 18,000 spectra was recalculated for mass calibration purposes. A simple quadratic equation was used to calibrate the mass scale with the time-of-flight spectrum (Riedo et al., 2013a).

7) Single mass unit decomposition. An integration of consecutive 260 single unit masses, starting from ^{1}H, was achieved by recalculating the time-of-flight windows from the mass calibration equation and utilizing direct Simpson's integration (Meyer et al., 2017).

Figure 3 shows a typical raw spectrum (top panel) acquired from the Gunflint sample before any data preprocessing has been

applied. The bottom panel shows a spectrum after preprocessing and reveals significantly improved SNR (10^4) and a flat baseline. After step number seven, multiple isotope maps were calculated using Kriging interpolation (further information in the text and *see* **Figure 4**) for an investigation of the distribution of major abundant elements. The imaging dataset was z-score normalized to remove the imbalanced scales. An assessment of the pairwise correlation factors was made, showing that approximately half of the dimensions (single unit masses) are empty or very weakly expressed.

The principal component analysis (PCA) reduction down to the first 60 principal components was applied to remove empty dimensions dominated by noise from the original dataset. The Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) was used to further characterize non-linear dependencies present in the PCA reduced data matrix. The overall classification of the UMAP scores was made using a hierarchical density-based clustering algorithm (HDBSCAN) (Campello et al., 2013; McInnes et al., 2017). The specific spectra identified from the microfossils were further visualized using the Mapper algorithm (Singh et al., 2007). The identification of the modules present in the Mapper network was conducted using a greedy modularity optimization algorithm (Louvain) (Blondel et al., 2008).

Figure 3. Comparison of fs-LIMS spectra (composite spectrum - 200 laser shots, recorded from single pixel), before and after data preprocessing, acquired from the Gunflint chert sample. Each line in the spectrum represents a single unit mass.
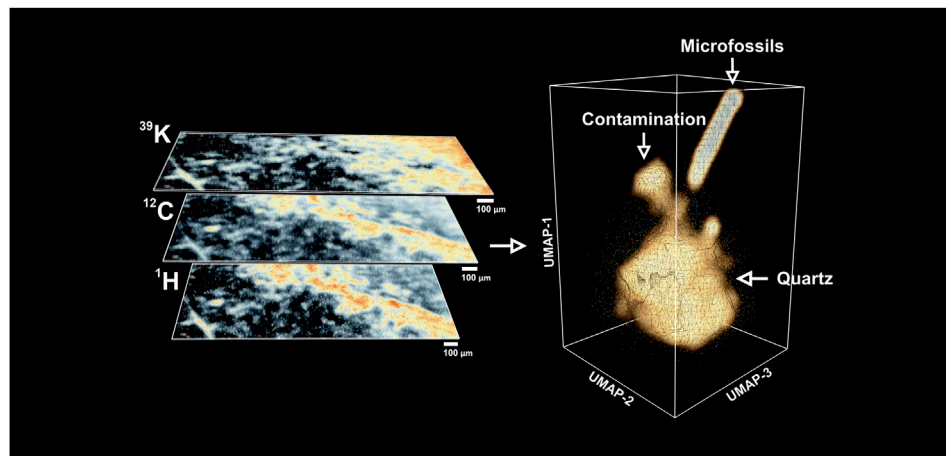
FIGURE 4 | Left panel–Exemplary isotope intensity maps (warmer colors indicate high concentrations) retrieved from the fs-LIMS mass spectra. The bio-lamination surface (aggregation of microfossils) could be identified from ^{12}C and ^{1}H maps (bright yellow to red areas), distribution of ^{39}K indicates the presence of the surface contamination (upper left corner, bright yellow to the red area). Dark areas on the isotope maps indicate low-intensity regions and correspond to the quartz matrix. To compare with an optical image, *see* **Figure 2A**. Right panel–Low dimensional structure of the imaging data cube revealed by UMAP. Triangulated mesh represents volumetric isodensity surface of UMAP scores calculated from the 18,000 fs-LIMS mass spectra. Three separate entities could be observed from the spectral neighborhood, namely quartz, contamination, and microfossils. The point cloud data plotted along with the density surface.

The increase of SNR to 10^4 and correction of the baseline can be noted. *See* the text for the full description of preprocessing procedures. Exemplary atomic lines are denoted on top of the spectrum.

RESULTS

We calculated the intensity maps of major (abundant) isotopes to understand a basic representation of the data. In **Figure 4**, the spatial distributions of ^{12}C, ^{1}H, and ^{39}K are illustrated and the chemical maps reveal specific areas where isotopes show elevated intensities. In comparison with the optical image of the same area (*see* **Figure 2A**), one can see that most of the dark brown patches identified from optical microscopy as microfossils preserved in the bio-lamination surface are spatially correlated with increased values of ^{12}C and ^{1}H. This observation is consistent with the fact that major elements within microfossil bodies are C and H. However, the intensity map of ^{39}K reveals different distribution. A top-right corner of the sample, which was previously identified as a clean matrix (milky quartz), reveals elevated concentrations of ^{39}K and relatively intense ion yields of ^{12}C. In fact, after a closer investigation of the mass spectra recorded from that region, we identified a full range of biorelevant elements (CHNOPS).

Additionally, a full range of Si isotopes, various silicon oxides, and small chains of hydrocarbon clusters were observed from that region. Considering that a particular location from optical microscopy does not show any distinct mineralogical association with described elements, we concluded that the identified area could belong to the organic contamination. From our previous studies of the Gunflint sample (Wiesendanger et al., 2018), particularly the chemical depth

profiling of the neighboring region, it was identified that organic contamination is present as a thin surface layer and organic spectral features quickly decay with increasing depth. The organic contamination potentially comes from the sample handling and preparation procedures and likely represents a small layer of lipids finely distributed on the surface.

In general, the manually inspected mass spectra from various regions appeared to be somewhat similar. They contain the same elements with varying concentrations–Si, CHNOPS, and polyatomic molecules of similar composition. This observation makes it difficult to manually define compounds observed from the Gunflint sample since they seem to represent continually mixing variants. The borders between chemical classes are fused into each other. Thus, the deterministic classification solely based on isotope intensity maps cannot be made. However, we can further explore the chemical variations within different parts of the sample. For example, the spectral features from the top-right corner also show very close proximity to the chemical composition of the host mineral - Si, O, and various Si oxide chains indicate that ablation craters were deep enough to pass through the layer of organic contamination and probe the chemical composition of the original underlying mineral. Lower parts of the isotope maps, shown with black regions (**Figure 4**-left) after a closer investigation of the mass spectra, were proposed to be from quartz, showing previously described simple chemistry–Si, O, and minor amounts of Na, K, Al. The latter elements (Na, K, Al) could be found as impurities within the chert since they are relatively common in the seawater and could have precipitated together with Si during the time of the rock formation, or they could be from phyllosilicates (clay minerals) that can occasionally occur in the matrix of Gunflint Formation stromatolites, e.g., (Lepot et al., 2017). Since the ^{12}C and ^{1}H maps outline the structure of the bio-lamination surface, previously

identified from the optical microscopy, we can investigate the spectra from the lamination site. The spectra from that area can be characterized by the presence of the bio-relevant elements–with increased concentrations of CHNOPS and an additional minor contribution from Fe, Mn, and Cr. Another notable observation is that spectra from the lamination surface reveal relatively strong polyatomic molecules formation patterns. Various hydrocarbon molecules accompanied by Si oxide chains populate the mass spectra up to 200 m/z.

A dimensionality reduction algorithm was applied over the full mass range of the fs-LIMS imaging data (1–260 amu) to find similar spectra in the dataset. We used the UMAP algorithm (McInnes et al., 2018) to analyze our observations. The first six UMAP components were retrieved from the dataset precompressed with PCA. The UMAP scores were calculated using the Euclidean distance as a metric; every 15 nearest neighbors were used in the construction of the k-nearest neighborhood graph, with a small minimal distance (0.1), and iterated over 400 epochs. This particular set of hyperparameters were found to be appropriate for an approximation of the global structure of the manifold. In the right panel of **Figure 4**, a distribution of the first three UMAP components is shown. The spectral neighborhood appears to be relatively busy (see point cloud data). However, three main protrusions can be observed from the equal density surface of the UMAP scores. The composition of protruding clusters matches our previous interpretation of the data. The lower part of the plot represents a relatively large cluster of mass spectra acquired from the Quartz-filled matrix. A smaller cluster observed in the vicinity of the main body corresponds to the spectra measured from the area with signatures of organic contamination. It is noteworthy that the contamination cluster is more connected to the main quartz cluster and that the structure of the density surface indicates a smooth transition from pure quartz to the spectra from the surface contamination. The transition structure forms a narrow neck where the similarity of spectra gradually changes from one class to another. From the point cloud data, we could *see* that the contamination cluster is relatively fuzzy, and the fine structure of the transition could be observed on the isodensity surface.
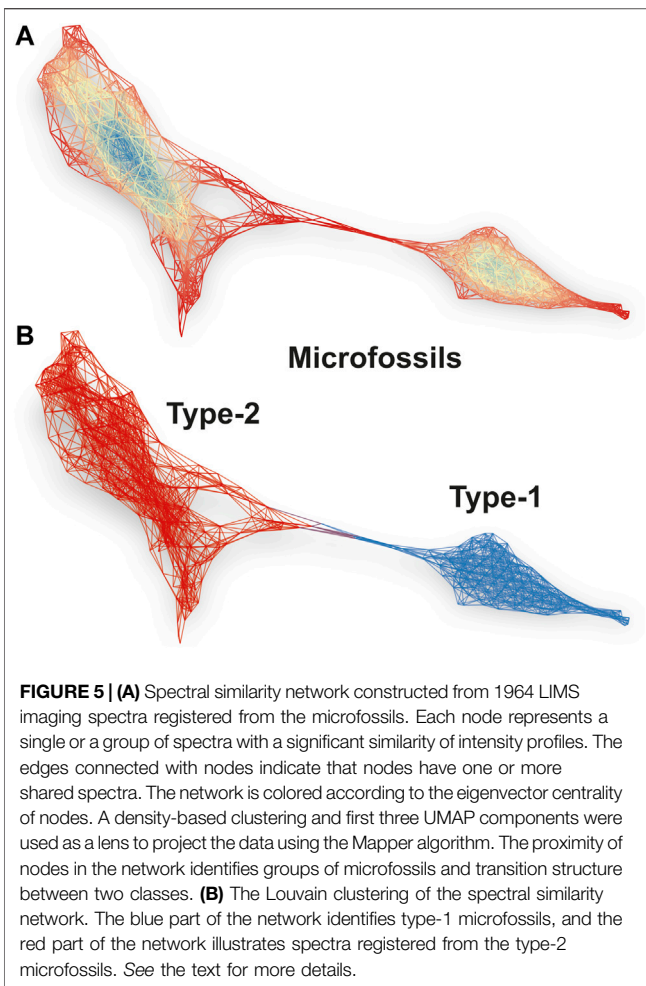
Through the same transition pathway, a cluster of spectra that corresponds to the microfossils preserved within the bio-lamination layer could be observed. In comparison to the cluster of spectra with organic contamination, the density surface of the microfossils cluster forms a separate transition line. The cluster of microfossils forms a smooth identifiable shape, which gradually rises further apart from the quartz and contamination clusters. As one can see, the relative proximity of the spectra located closer to the transition "neck" indicates the ablation of small parts of microfossils. From the investigation of the individual spectra (*see* **Figure 3**), we have noted that almost all spectra from microfossils contain spectral features from the filling quartz mineral. This observation could be explained by the fact that bodies of microfossils represent partially collapsed and degraded cell walls. The thicknesses of the partially decayed cell walls vary from the first tens of nm to the first hundreds of nm, and these walls are all entombed in the silica matrix. By ablation of small portions of the microfossils and larger portions of the silica matrix, we can explain the smooth transition

structure, where similarity of spectra transitions from the clean silica matrix. Thus, the end members of the microfossil cluster represent the best volumetric sampling of microfossils, as well as the best chemical composition of the fossils.

Overall, the volumetric density estimate of the UMAP scores provides a good overview of the spectral types and their transition structures. Also, it is possible to identify outliers (e.g., microscopic inclusions of other minerals) from this graph, for example, by recalculating the isolation forest scores (or any other outlier detection algorithm) – the data points that are weakly connected to the main clusters will have high values, thus, easily identifiable. In the fs-LIMS analysis, where fine chemistry is often of great interest, such information might be valuable because it allows the identification of detached spectra from the bulk of very similar ones.

Figure 4. Left panel–Exemplary isotope intensity maps (warmer colors indicate high concentrations) retrieved from the fs-LIMS mass spectra. The bio-lamination surface (aggregation of microfossils) could be identified from ^{12}C and ^{1}H maps (bright yellow to red areas), distribution of ^{39}K indicates the presence of the surface contamination (upper left corner, bright yellow to the red area). Dark areas on the isotope maps indicate low-intensity regions and correspond to the quartz matrix. To compare with an optical image, *see* **Figure 2A**. Right panel–Low dimensional structure of the imaging data cube revealed by UMAP. Triangulated mesh represents volumetric isodensity surface of UMAP scores calculated from the 18,000 fs-LIMS mass spectra. Three separate entities could be observed from the spectral neighborhood, namely quartz, contamination, and microfossils. The point cloud data plotted along with the density surface.

The UMAP isodensity estimate reveals the continuous structure of spectral similarities, and therefore it is not clear where to define a boundary between different classes. A density-based clustering approach was used to define discreet classes from the low dimensional UMAP scores. The six UMAP components were used to discretize distributions using a Hierarchical Density-Based Spatial Clustering (HDBSCAN) algorithm (Campello et al., 2013; McInnes et al., 2017). An HDBSCAN provides relatively conservative class assignments compared to other clustering algorithms and potentially more accurate in its predictions. An advantageous side of HDBSCAN over DBSCAN, for example, is that it can find clusters with varying densities, which is precisely the case with our data, where we have an oversampled data from the silicified matrix and a relatively small number of spectra from the microfossils. Moreover, it is possible to calculate the confidence probabilities of the assignment of each spectrum to the cluster, which makes troubleshooting of clustering results more intuitive and less bothersome. However, the downside of the conservative clustering is that some portions of the data might be classified as noise if they do not tightly belong to the densely packed cluster. In contrast to the previous interpretation of UMAP scores, the clustering algorithm found two microfossil populations, a cluster of surface contamination, and quartz from the matrix. The additional cluster of microfossils was hidden on the backside of the quartz-related spectra (*see* **Figure 4**). The Mapper networks were applied to the spectra registered from the

FIGURE 5 | (A) Spectral similarity network constructed from 1964 LIMS imaging spectra registered from the microfossils. Each node represents a single or a group of spectra with a significant similarity of intensity profiles. The edges connected with nodes indicate that nodes have one or more shared spectra. The network is colored according to the eigenvector centrality of nodes. A density-based clustering and first three UMAP components were used as a lens to project the data using the Mapper algorithm. The proximity of nodes in the network identifies groups of microfossils and transition structure between two classes. **(B)** The Louvain clustering of the spectral similarity network. The blue part of the network identifies type-1 microfossils, and the red part of the network illustrates spectra registered from the type-2 microfossils. *See* the text for more details.

microfossils to visualize the proximity structure between these two classes.

Figure 5A shows a spectral similarity network constructed from the fs-LIMS spectra registered from the microfossils, using the first three UMAP components as a lens. A python implementation - Kepler Mapper (Van Veen et al., 2019) of the Mapper algorithm (Singh et al., 2007) was used to calculate the similarity network of LIMS spectra. However, other open-source implementations exist - e.g., recently published Giotto-TDA (Tauzin et al., 2020). The density-based clustering was applied to identify clusters within overlapping filter function windows. In total, twenty windows were applied to construct the network with 40% overlap over three UMAP components, forming 8,000 sampling windows and resulting in a complex network with 417 nodes and 2,967 edges (from 1,964 composite spectra registered from the microfossils). Note that the number of filter dimensions is user-defined, and in principle, they might be defined as an *n*-dimensional hypercube, though two-dimensional filters provide the best interpretability. The nodes present in the network indicate groups of fs-LIMS spectra with a high degree of similarity. The nodes might contain one or hundreds of spectra, depending on the size of the filter function window. The edge between nodes is drawn if nodes share the same observations (it

might be one or many more spectra). The coloring of the network is conducted according to the eigenvector centralities of the nodes. Blue parts of the network indicate the central nodes, and red parts indicate less connected network components.

The structure of the network identifies the presence of two connected communities. **Figure 5B** shows the same spectral similarity network as in **Figure 5A** but colored according to the Louvain modularity, calculated from the network topology. The red part of the network (nodes are not shown) categorizes the spectra identified from the type-2 microfossils, and the blue network indicates the type-1 microfossils. The type-2 microfossils can be characterized by increased proximity to the cluster of spectra registered from the quartz. In contrast to the spectra from type-2, type-1 microfossils are almost completely detached from other groups and form a community of highly connected nodes and correspond to the spectra in a linear protrusion in **Figure 4** (right panel). Note that the HDBSCAN and Louvain clustering provides mutually supportive clustering results, although the Mapper networks provide better tolerance to noise, thus allowing for improved clustering performance. In order to check that cluster assignments are not artifactual, we performed a clustering robustness analysis. The Rand Index (RI) metric was used to assess the clustering similarity between 10 random subsamples of the data registered from microfossils. In total, 75% of the data was used to generate random subsamples. The output UMAP subsamples were clustered using the Louvain community detection algorithm. The RI similarity matrix for Louvain clustering of random samples could be found in the supplementary information (*see* **Supplementary Table S1** and **Supplementary Figure S3**). Overall, 45 different clustering pairs revealed an average RI score of 92.5% with a standard deviation of 2%, which indicates that communities shown in **Figure 5** are not artefactual and that the cluster assignments are robust. Most of the clustering uncertainty can be attributed to the transition zone between two types of microfossils. The type-2 microfossils reveal more inhomogeneity (*see* **Supplementary Figures S1, S2**) in comparison to the type-1 microfossils and represent more intermixed with the host mineral material.

The spectral similarity network calculated from the first three UMAP components reveals a better visualization of internal structure and detects outliers and irregularities. Moreover, the force-directed layout (ForceAtlas2 (Jacomy et al., 2014)), applied to the network, exaggerates the positioning of weakly connected nodes, which makes them easier to detect. Moreover, interpretation of the low-dimensional embedding of fs-LIMS data can be easily achieved by coloring the network with original isotope intensities and synthetic features such as isotope ratios. Any other functions might be applied to the data (e.g., Kernel Density Estimate (KDE), Singular Value Decomposition (SVD), and Principal Component Analysis (PCA)), which makes Mapper networks a versatile and powerful tool for insight extraction and hypothesis generation. Furthermore, by reducing the large fs-LIMS intensity space down to a network, we can additionally define a multitude of secondary statistics that could be calculated from the graph topology. Metrics such as centrality, modularity (e.g., *see* **Figures 5A,B**), average degree, path length (e.g., between the host mineral and microfossils), and many more, can be applied to the specific
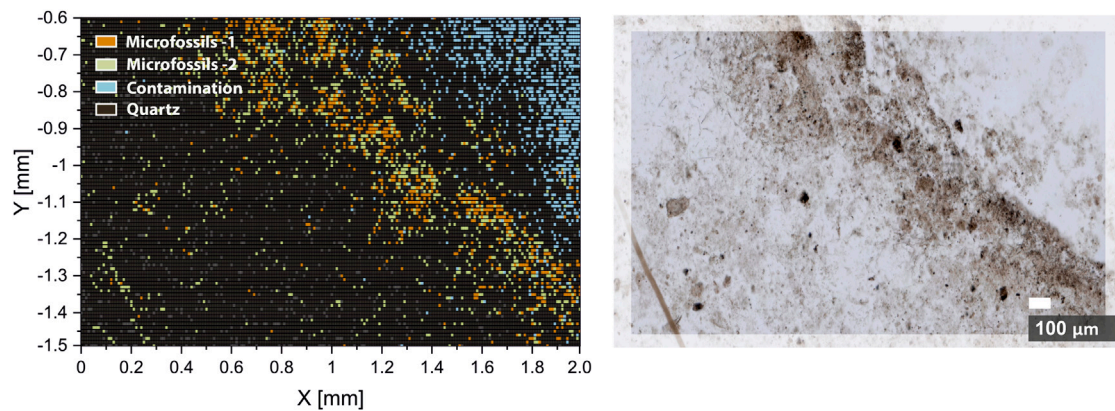
FIGURE 6 | Hierarchical density-based spatial clustering (HDBSCAN) of six UMAP components of the imaging dataset (left panel). The orange pixels represent spectra registered from the type-1 microfossils. The green pixels represent spectra registered from the type-2 microfossils. The blue pixels represent spectra registered from the surface contamination. Black and grey pixels–spectra registered from the quartz matrix of the Gunflint chert. Right panel–the optical microscopy image of the analyzed area. Note the aligned distribution of classified spectra with the bio-lamination surface crossing the image.

minerals and microfossils to define the multiparametric space further and enhance the potential for definitive identification.

Figure 5A. Spectral similarity network constructed from 1,964 LIMS imaging spectra registered from the microfossils. Each node represents a single or a group of spectra with a significant similarity of intensity profiles. The edges connected with nodes indicate that nodes have one or more shared spectra. The network is colored according to the eigenvector centrality of nodes. A density-based clustering and first three UMAP components were used as a lens to project the data using the Mapper algorithm. The proximity of nodes in the network identifies groups of microfossils and transition structure between two classes. **Figure 5B**. The Louvain clustering of the spectral similarity network. The blue part of the network identifies type-1 microfossils, and the red part of the network illustrates spectra registered from the type-2 microfossils. See the text for more details.

The overall results of the density-based clustering can be seen in **Figure 6**. Clustering results reveal a very close match with results of optical microscopy (*see* **Figure 6**, right panel) and conclusions from previous single isotope maps investigations. Moreover, we have identified two types of microfossils and a contamination zone, which were not acknowledged from the microscope image. The type–1 microfossils represent spectra obtained from the microfossils with the best microfossil over host (matrix mineral) sampling ratio. Thus, spectra from type-1 can be counted as the most representative of microfossils. On the other hand, type-2 represents the microfossils with an increased contribution from the host mineral, which is also shown in **Figure 6**. The chemical composition of type-1 microfossils can be characterized with increased content of carbon and oxygen (^{12}C, $^{12}C^{2+}$, and $^{16}O^{2+}$ peaks in the mass spectra), whereas type-2 microfossils contain less ^{12}C and more hydrocarbons, which indicates lower volumetric ablation and colder plasma temperatures, thus, more prevalent recombination processes. Higher plasma temperatures observed in the type-1 microfossils can be attributed to the higher volumetric contribution from absorptive kerogen. This observation also finds confirmation from the spatial distribution of microfossils. In

Figure 6, the first type is mainly distributed in the densely populated area (*see* **Figure 6**, right panel), in contrast to type-2, which is largely distributed outside of the dense zone, and more likely to be sampled with larger portions of the host mineral. The identification of microfossils from the host mineral using fs-LIMS and low dimensional analysis provides topological biosignatures. As it was shown in **Figures 4**, **5**, the structure of spectral similarities identifies the positionings of spectra from different classes and provides means for identification, classification of large datasets, and has a potential for the prediction of spectral classes from previously unseen spectra, given that a sufficiently rich spectral library is provided.

Figure 6. Hierarchical density-based spatial clustering (HDBSCAN) of six UMAP components of the imaging dataset (left panel). The orange pixels represent spectra registered from the type-1 microfossils. The green pixels represent spectra registered from the type-2 microfossils. The blue pixels represent spectra registered from the surface contamination. Black and grey pixels–spectra registered from the quartz matrix of the Gunflint chert. Right panel–the optical microscopy image of the analyzed area. Note the aligned distribution of classified spectra with the bio-lamination surface crossing the image.

DISCUSSION

The identification and chemical characterization of minerals and prospective biosignatures from large spectral databases generated using fs-LIMS as well as other *in-situ* spectroscopic techniques is a longstanding problem that can be generalized to other analytical methods as well. For example, other important methods proposed for *in-situ* space exploration, such as Laser-Induced Breakdown Spectroscopy (LIBS) (e.g., ChemCam, currently operates on Mars as part of the Mars Science Laboratory), Raman spectroscopy (i.e., Raman Laser Spectrometer (RLS), one of the Pasteur Payload instruments from ExoMars), and a large variety of other techniques rely on harvesting large spectral information

from the analyte material. This spectral information is often hard to interpret due to the large dimensionality, complexity, and size of generated datasets. Outside of the context of space exploration, in the field of analytical chemistry, similar data analytical challenges are often encountered in the laboratory. For example, Secondary Ions Mass Spectrometry (SIMS) or Liquid Chromatography Mass Spectrometry (LC-MS), as high-throughput techniques, provide hundreds of mass lines per spectrum, and the output spectral dataset is not always easy to interpret. As was shown in this contribution, analysis of fs-LIMS data using topological methods reveals a fast and accurate description of spectral classes and provides a good understanding of transitional structures. In the low dimensional domain, it might be easier to generate insights and formulate a hypothesis, thus accelerating the extraction of knowledge from the given sample.

The analysis of data generated by using our fs-LIMS system might also be of use for future investigations of Precambrian rocks containing signatures of putative microfossils. The Gunflint sample is rare amongst Precambrian rocks as it exhibits an exceptional level of morphological and chemical preservation, so there is little argument over the biogenicity of the encased organic material (Barghoorn and Tyler 1965; Lepot et al., 2017; Wacey et al., 2012). However, traces of early life can be destroyed or heavily altered by heat, pressure, and time (diagenetic alteration and later metamorphism). As was briefly discussed before, the full mass range spectral proximity analysis provides a means for the classification of chemically similar entities. For example, organic contamination and microfossils - similar compounds (both contain CHNOPS and Si mass peaks), can be distinguished using topological methods (see **Figures 4–6**). A big challenge in the field of Precambrian micropaleontology surrounds the fact that altered and reduced carbon found in ancient rocks could potentially be of biological origin but could also have been created by abiotic processes. For example, Fischer-Tropsch type synthesis might be responsible for the presence of some abiotic hydrocarbons in Precambrian formations (Brasier et al., 2002). However, we speculate that synthetic products of Fischer-Tropsch-like reactions will have a distinct spectral profile (e.g., polyatomic plasma chemistry products might be different), and therefore corresponding topological positioning is expected to be distinguishable from *bona fide* microfossils. Thus, there is a hope that signs of life in controversial samples might be successfully identified using sensitive methods and full-feature-based topological representations.

The current state of space exploration also faces similar challenges in the field of *in-situ* chemical analysis of solids on planetary bodies. For example, the ns-LIMS instrument proposed for Europa (Ligterink et al., 2020) reported the identification of extremely low quantities of biological and abiotic amino acids from well-defined extracts at the fmole level. However, more complex molecules (e.g., proteins, polysaccharides, etc.) combined with various undefined matrices will likely form complex fragmentation patterns with hundreds of significant mass lines, thus, making the identification challenging. The topological representation, in this case, might provide a number of compounds present in the measured mixture and their similarity to the predefined classes.

The unsupervised identification of minerals from fs-LIMS chemical imaging datasets might also be of use in the determination of relative sensitivity coefficients (RSC's). The fs-LIMS is a quantitative method; however, it requires the establishment of RSC's, which are matrix dependent. With an introduction of fs-LIMS, some matrices have been reduced to unity (RSC = 1, no correction needed). However, non-absorptive samples such as glasses typically still require the determination of RSC's for quantitative measurements. In the case of exploratory analysis, where we do not know the sample (i.e., field exploration of Martian samples), if one would know the stoichiometry of the investigated mineral, it is possible to recalculate correction factors for major elements, and then through RSC's dependence on atomic orbital ionization energy recalculate concentrations of minor and trace elements (Tulej et al., 2021). The key component here is the identification of the mineral, and as we described above, topological methods provide a means to do that.

Here we also need to point out several caveats regarding the analysis of fs-LIMS data. First, at multiple stages, the data analysis procedures require a set of hyperparameters to be chosen. For example, in UMAP embedding and Mapper network construction, we used Euclidean and cosine distances, respectively, and defined the number of neighbors, number of clusters, and filter functions. However, a more rigorous study of the effect of hyperparameters needs to be assessed in future studies regarding the analysis of fs-LIMS data or data generated by other spectroscopic techniques. Nevertheless, a recent contribution by (Belchı et al., 2020) provides an insight into the numerical stability of Mapper-type algorithms. It was shown that reliable Mapper output could be identified as a local minimum of instability, regarded as a function of Mapper input parameters. Other statistical solutions were proposed to circumvent testing large parametric spaces and keep the most representative Mapper settings (Carriere et al., 2018). Furthermore, we have used UMAP scores as a lens in the construction of the similarity network; however, a large variety of other functions might be used, and their impact on visualizations needs to be assessed. It would also be valuable to implement into the analysis pipeline some domain-specific lenses for technical usage (i.e., mass resolution, mass accuracy, etc.), which will improve the extraction of quality metrics.

Overall, in addition to the account of topological descriptors of early life, we hope that our analysis will facilitate, in time, a predictive approach in the field of study of early life. The approach described here might be expanded to more powerful, state-of-the-art standalone laboratory instrumentation (e.g., high-resolution LIMS, SIMS, LA-ICP-MS), where data quality might provide a whole new quantification perspective.

CONCLUSION

Our contribution offers several important conclusions for *in-situ* space research. First, the miniature fs-LIMS system combined with topology-based data analysis demonstrates the utility and sensitivity to distinguish organically preserved microfossils from organic contamination and inorganic host mineralogy. Second,

the proposed approach might be extended to other complex samples with multimineral compositions and used with other high-resolution spectrometric or spectroscopic methods. Third, our approach - full spectral mass range convolution down to a similarity network for life detection stands out from multielement methods. It offers great flexibility and could be further expanded to study the chemical discrepancies between individual populations of microfossils. Furthermore, our analysis reveals fine transition structures between classes and the detection of outliers. Last, the fs-LIMS system, in combination with topological methods, enables faster data analysis, accelerates the formulation of hypotheses, and the generation of insights for mineralogical compositions of investigated samples.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the **Supplementary Materials**, further inquiries can be directed to the corresponding author.

REFERENCES

Arevalo, R., Jr, Ni, Z., and Danell, R. M. (2020). Mass Spectrometry and Planetary Exploration: A Brief Review and Future Projection. *J. Mass. Spectrom.* 55, e4454. doi:10.1002/jms.4454

Arevalo, R., Jr, Selliez, L., Briois, C., Carrasco, N., Thirkell, L., Cherville, B., et al. (2018). An Orbitrap-Based Laser Desorption/Ablation Mass Spectrometer Designed for Spaceflight. *Rapid Commun. Mass. Spectrom.* 32, 1875–1886. doi:10.1002/rcm.8244

Azov, V. A., Mueller, L., and Makarov, A. A. (2020). Laser Ionization Mass Spectrometry at 55: Quo Vadis? *Mass Spectrom. Rev.* 1–52. doi:10.1002/mas.21669

Barghoorn, E. S., and Tyler, S. A. (1965). Microorganisms from the Gunflint Chert: These Structurally Preserved Precambrian Fossils from Ontario Are the Most Ancient Organisms Known. *Science* 147, 563–575. doi:10.1126/science.147.3658.563

Belchı, F., Brodzki, J., Burfitt, M., and Niranjan, M. (2020). A Numerical Measure of the Instability of Mapper-Type Algorithms. *J. Machine Learn. Res.* 21, 1–45.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. *J. Stat. Mech.* 2008, P10008. doi:10.1088/1742-5468/2008/10/p10008

Brasier, M. D., Green, O. R., Jephcoat, A. P., Kleppe, A. K., Van Kranendonk, M. J., Lindsay, J. F., et al. (2002). Questioning the Evidence for Earth's Oldest Fossils. *Nature* 416, 76–81. doi:10.1038/416076a

Brasier, M. D., and Wacey, D. (2012). Fossils and Astrobiology: New Protocols for Cell Evolution in Deep Time. *Int. J. Astrobiology* 11, 217–228. doi:10.1017/s1473550412000298

Campello, R. J., Moulavi, D., and Sander, J. (2013). "Density-Based Clustering Based on Hierarchical Density Estimates," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Heidelberg: Springer). doi:10.1007/978-3-642-37456-2_14

Carriere, M., Michel, B., and Oudot, S. (2018). Statistical Analysis and Parameter Selection for Mapper. *J. Machine Learn. Res.* 19, 478–516.

Chan, M. A., Hinman, N. W., Potter-McIntyre, S. L., Schubert, K. E., Gillams, R. J., Awramik, S. M., et al. (2019). Deciphering Biosignatures in Planetary Contexts. *Astrobiology* 19, 1075–1102. doi:10.1089/ast.2018.1903

Garcia-Lopez, E., and Cid, C. (2017). Glaciers and Ice Sheets as Analog Environments of Potentially Habitable Icy Worlds. *Front. Microbiol.* 8, 1407. doi:10.3389/fmicb.2017.01407

Gil, A., and Marco, R. (2007). *Systems Bioinformatics: An Engineering Case-Based Approach*. Artech.

Goesmann, F., Brinckerhoff, W. B., Raulin, F., Goetz, W., Danell, R. M., Getty, S. A., et al. (2017). The Mars Organic Molecule Analyzer (MOMA) Instrument:

AUTHOR CONTRIBUTIONS

RL performed the experiments and data analysis. RL wrote the main manuscript. All authors reviewed and revised the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2021.668163/full#supplementary-material

Characterization of Organic Material in Martian Sediments. *Astrobiology* 17, 655–685. doi:10.1089/ast.2016.1551

Grimaudo, V., Tulej, M., Riedo, A., Lukmanov, R., Ligterink, N. F. W., de Koning, C., et al. (2020). UV Post-Ionization Laser Ablation Ionization Mass Spectrometry for Improved Nm-Depth Profiling Resolution on Cr/Ni Reference Standard. *Rapid Commun. Mass. Spectrom.* 34, e8803. doi:10.1002/rcm.8803

Grimaudo, V., Moreno-García, P., Riedo, A., López, A. C., Tulej, M., Wiesendanger, R., et al. (2019). Review-Laser Ablation Ionization Mass Spectrometry (LIMS) for Analysis of Electrodeposited Cu Interconnects. *J. Electrochem. Soc.* 166, D3190–D3199. doi:10.1149/2.0221901jes

Hand, K., Murray, A., Garvin, J., Brinckerhoff, W., Christner, B., Edgett, K., et al. (2017). *Europa Lander Study 2016 Report: Europa Lander Mission*. La Cañada Flintridge, CA: NASA Jet Propuls. Lab.

Hofmann, B. A. (2008). Morphological Biosignatures from Subsurface Environments: Recognition on Planetary Missions. *Space Sci. Rev.* 135, 245–254. doi:10.1007/s11214-007-9147-9

Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). ForceAtlas2, A Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PloS one* 9, e98679. doi:10.1371/journal.pone.0098679

Lepot, K., Addad, A., Knoll, A. H., Wang, J., Troadec, D., Béché, A., et al. (2017). Iron Minerals within Specific Microfossil Morphospecies of the 1.88 Ga Gunflint Formation. *Nat. Commun.* 8, 14890. doi:10.1038/ncomms14890

Li, X., Danell, R. M., Pinnick, V. T., Grubisic, A., van Amerom, F., Arevalo, R. D., et al. (2017). Mars Organic Molecule Analyzer (MOMA) Laser Desorption/Ionization Source Design and Performance Characterization. *Int. J. Mass Spectrom.* 422, 177–187. doi:10.1016/j.ijms.2017.03.010

Ligterink, N. F. W., Grimaudo, V., Moreno-García, P., Lukmanov, R., Tulej, M., Leya, I., et al. (2020). ORIGIN: A Novel and Compact Laser Desorption - Mass Spectrometry System for Sensitive In Situ Detection of Amino Acids on Extraterrestrial Surfaces. *Sci. Rep.* 10, 9641. doi:10.1038/s41598-020-66240-1

Managadze, G. G., Wurz, P., Sagdeev, R. Z., Chumikov, A. E., Tuley, M., Yakovleva, M., et al. (2010). Study of the Main Geochemical Characteristics of Phobos' Regolith Using Laser Time-Of-Flight Mass Spectrometry. *Sol. Syst. Res.* 44, 376–384. doi:10.1134/s0038094610050047

McInnes, L., Healy, J., and Astels, S. (2017). Hdbscan: Hierarchical Density Based Clustering. *J. Open Source Softw.* 2, 205. doi:10.21105/joss.00205

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [Preprint]. Available at: https://arxiv.org/abs/1802.03426 (Accessed September 18, 2020).

McMahon, S., Bosak, T., Grotzinger, J. P., Milliken, R. E., Summons, R. E., Daye, M., et al. (2018). A Field Guide to Finding Fossils on Mars. *J. Geophys. Res. Planets* 123, 1012–1040. doi:10.1029/2017je005478

Meyer, S., Riedo, A., Neuland, M. B., Tulej, M., and Wurz, P. (2017). Fully Automatic and Precise Data Analysis Developed for Time-Of-Flight Mass Spectrometry. *J. Mass. Spectrom.* 52, 580–590. doi:10.1002/jms.3964

Moreno-García, P., Grimaudo, V., Riedo, A., Tulej, M., Wurz, P., and Broekmann, P. (2016). Towards Matrix-Free Femtosecond-Laser Desorption Mass Spectrometry for In Situ Space Research. *Rapid Commun. Mass. Spectrom.* 30, 1031–1036. doi:10.1002/rcm.7533

Navarro-González, R., Navarro, K. F., Rosa, J. d. l., Iñiguez, E., Molina, P., Miranda, L. D., et al. (2006). The Limitations on Organic Detection in Mars-Like Soils by Thermal Volatilization-Gas Chromatography-MS and Their Implications for the Viking Results. *Proc. Natl. Acad. Sci.* 103, 16089–16094. doi:10.1073/pnas.0604210103

Neuland, M. B., Grimaudo, V., Mezger, K., Moreno-García, P., Riedo, A., Tulej, M., et al. (2016). Quantitative Measurement of the Chemical Composition of Geological Standards with a Miniature Laser Ablation/Ionization Mass Spectrometer Designed Forin Situapplication in Space Research. *Meas. Sci. Technol.* 27, 035904. doi:10.1088/0957-0233/27/3/035904

Neveu, M., Hays, L. E., Voytek, M. A., New, M. H., and Schulte, M. D. (2018). The Ladder of Life Detection. *Astrobiology* 18, 1375–1402. doi:10.1089/ast.2017.1773

Press, W. H., and Teukolsky, S. A. (1990). Savitzky-Golay Smoothing Filters. *Comput. Phys.* 4, 669–672. doi:10.1063/1.4822961

Priscu, J. C., and Hand, K. P. (2012). Microbial Habitability of Icy Worlds. *Microbe* 7, 167–172. doi:10.1128/microbe.7.167.1

Riedo, A., Bieler, A., Neuland, M., Tulej, M., and Wurz, P. (2013a). Performance Evaluation of a Miniature Laser Ablation Time-Of-Flight Mass Spectrometer Designed Forin Situinvestigations in Planetary Space Research. *J. Mass. Spectrom.* 48, 1–15. doi:10.1002/jms.3104

Riedo, A., Koning, C. d., Stevens, A., Cockell, C., McDonald, A., López, A. C., et al. (2020). The Detection of Elemental Signatures of Microbes in Martian Mudstone Analogs Using High Spatial Resolution Laser Ablation Ionization Mass Spectrometry. *Astrobiology* 20 (10), 1224–1235. doi:10.1089/ast.2019.2087

Riedo, A., Meyer, S., Heredia, B., Neuland, M. B., Bieler, A., Tulej, M., et al. (2013b). Highly Accurate Isotope Composition Measurements by a Miniature Laser Ablation Mass Spectrometer Designed for In Situ Investigations on Planetary Surfaces. *Planet. Space Sci.* 87, 1–13. doi:10.1016/j.pss.2013.09.007

Riedo, A., Neuland, M., Meyer, S., Tulej, M., and Wurz, P. (2013c). Coupling of LMS with a Fs-Laser Ablation Ion Source: Elemental and Isotope Composition Measurements. *J. Anal. Spectrom.* 28, 1256–1269. doi:10.1039/c3ja50117e

Riedo, A., Tulej, M., Rohner, U., and Wurz, P. (2017). High-Speed Microstrip Multi-Anode Multichannel Plate Detector System. *Rev. Scientific Instr.* 88, 045114. doi:10.1063/1.4981813

Singh, G., Mémoli, F., and Carlsson, G. E. (2007). "Topological Methods for the Analysis of High Dimensional Data Sets and 3d Object Recognition," in *Eurographics Symposium on Point-Based Graphics*. Editors M. Botsch, R. Pajarola, B. Chen, and M. Zwicker (The Eurographics Association), 91, 100. doi:10.2312/SPBG/SPBG07/091-100

Stevens, A. H., McDonald, A., de Koning, C., Riedo, A., Preston, L. J., Ehrenfreund, P., et al. (2019). Detectability of Biosignatures in a Low-Biomass Simulation of Martian Sediments. *Sci. Rep.* 9, 9706. doi:10.1038/s41598-019-46239-z

Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A., et al. (2020). Giotto-Tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration. arXiv [Preprint]. Available at: https://arxiv.org/abs/2004.02551 (Accessed March 5, 2021).

Tulej, M., Lukmanov, R., Grimaudo, V., Riedo, A., de Koning, C., Ligterink, N. F. W., et al. (2021). Determination of the Microscopic Mineralogy of Inclusion in an Amygdaloidal Pillow basalt by Fs-LIMS. *J. Anal. Spectrom.* 36, 80–91. doi:10.1039/d0ja00390e

Tulej, M., Neubeck, A., Ivarsson, M., Riedo, A., Neuland, M. B., Meyer, S., et al. (2015). Chemical Composition of Micrometer-Sized Filaments in an Aragonite Host by a Miniature Laser Ablation/Ionization Mass Spectrometer. *Astrobiology* 15, 669–682. doi:10.1089/ast.2015.1304

Tulej, M., Neubeck, A., Riedo, A., Lukmanov, R., Grimaudo, V., Ligterink, N. F. W., et al. (2020). Isotope Abundance Ratio Measurements Using Femtosecond Laser Ablation Ionization Mass Spectrometry. *J. Mass. Spectrom.* 55, e4660. doi:10.1002/jms.4660

Tulej, M., Riedo, A., Neuland, M. B., Meyer, S., Wurz, P., Thomas, N., et al. (2014). CAMAM: A Miniature Laser Ablation Ionisation Mass Spectrometer and Microscope-Camera System forIn SituInvestigation of the Composition and Morphology of Extraterrestrial Materials. *Geostand Geoanal Res.* 38, 441–466. doi:10.1111/j.1751-908x.2014.00302.x

Vago, J. L., Westall, F., Coates, A. J., Jaumann, R., Korablev, O., Ciarletti, V., et al. (2017). Habitability on Early Mars and the Search for Biosignatures with the ExoMars Rover. *Astrobiology* 17, 471–510. doi:10.1089/ast.2016.1533

Vago, J., Witasse, O., Svedhem, H., Baglioni, P., Haldemann, A., Gianfiglio, G., et al. (2015). ESA ExoMars Program: The Next Step in Exploring Mars. *Sol. Syst. Res.* 49, 518–528. doi:10.1134/s0038094615070199

Van Veen, H., Saul, N., Eargle, D., and Mangham, S. (2019). Kepler Mapper: A Flexible Python Implementation of the Mapper Algorithm. *J. Open Source Softw.* 4, 1315. doi:10.21105/joss.01315

Wacey, D., McLoughlin, N., Kilburn, M. R., Saunders, M., Cliff, J. B., Kong, C., et al. (2013). Nanoscale Analysis of Pyritized Microfossils Reveals Differential Heterotrophic Consumption in the ~1.9-Ga Gunflint Chert. *Proc. Natl. Acad. Sci. U S A.* 110, 8020–8024. doi:10.1073/pnas.1221965110

Wacey, D., Menon, S., Green, L., Gerstmann, D., Kong, C., McLoughlin, N., et al. (2012). Taphonomy of Very Ancient Microfossils from the ~3400Ma Strelley Pool Formation and ~1900Ma Gunflint Formation: New Insights Using a Focused Ion Beam. *Precambrian Res.* 220-221, 234–250. doi:10.1016/j.precamres.2012.08.005

Wacey, D., Saunders, M., Kong, C., Brasier, A., and Brasier, M. (2016). 3.46 Ga Apex Chert 'Microfossils' Reinterpreted as Mineral Artefacts Produced During Phyllosilicate Exfoliation. *Gondwana Res.* 36, 296–313. doi:10.1016/j.gr.2015.07.010

Westall, F., Foucher, F., Bost, N., Bertrand, M., Loizeau, D., Vago, J. L., et al. (2015). Biosignatures on Mars: What, Where, and How? Implications for the Search for Martian Life. *Astrobiology* 15, 998–1029. doi:10.1089/ast.2015.1374

Wiesendanger, R., Tulej, M., Riedo, A., Frey, S., Shea, H., and Wurz, P. (2017). Improved Detection Sensitivity for Heavy Trace Elements Using a Miniature Laser Ablation Ionisation Mass Spectrometer. *J. Anal. Spectrom.* 32, 2182–2188. doi:10.1039/c7ja00193b

Wiesendanger, R., Wacey, D., Tulej, M., Neubeck, A., Ivarsson, M., Grimaudo, V., et al. (2018). Chemical and Optical Identification of Micrometer-Sized 1.9 Billion-Year-Old Fossils by Combining a Miniature Laser Ablation Ionization Mass Spectrometry System with an Optical Microscope. *Astrobiology* 18, 1071–1080. doi:10.1089/ast.2017.1780

Wurz, P., Riedo, A., Tulej, M., Grimaudo, V., and Thomas, N. (2020). Investigation of the Surface Composition by Laser Ablation/Ionisation Mass Spectrometry. *LPI Contrib.* 2241, 5061. doi:10.1109/AERO50100.2021.9438486

3.2.3 High-Resolution fs-LIMS 3D imaging and manifold learning reveal insight into chemical diversity of the Gunflint chert (1.88 Ga)-(Frontiers in Space Technologies)

Rustam A. Lukmanov[1*], Coenraad de Koning[1], Peter Keresztes Schmidt[1], David Wacey[2], Niels F.W. Ligterink[1], Valentine Grimaudo[1], Anna Neubeck[3], Andreas Riedo[1], Marek Tulej[1], Peter Wurz[1]

[1]Space Research and Planetary Sciences (WP), University of Bern, Bern, Switzerland
[2]Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, Perth, Australia
[3]Department of Earth Sciences, Uppsala University, Uppsala, Sweden
Correspondence:
Rustam A. Lukmanov: rustam.lukmanov@space.unibe.ch
Keywords: fs-LIMS, Mass-spectrometry, UMAP, Mapper, microfossils, Mars, Gunflint

Abstract

We examine the capability of our newly developed high-resolution laser ionization mass spectrometer (fs-LIMS-GT) for the identification of faint elemental signatures of life from 1.88 billion-year-old microfossils (Gunflint chert, Ontario, Canada). We report novel results on: 1) Mass-spectrometric volumetric imaging of the Gunflint chert sample using a high-resolution fs-LIMS instrument, that has achieved spatial resolution on a single cell level together with ppm level sensitivity; 2) The analysis of spectral neighborhood using the Uniform Manifold Approximation and Projection (UMAP) method, revealing the presence of microscopic heterogeneous sulfide inclusions and distinct cluster of spectra registered from the microfossils; 3) Large-scale mass-spectrometric sampling, allowing for the identification of unique chemical characteristics of microfossils and mapping of the full variance of the intensity profiles, thus, providing a means for unsupervised data-driven deterministic identifications of microfossils embedded in the solid silicate matrix; 4) Partial Mg enrichment revealed by the investigated microfossils, which may indicate that some of the microfossils were photosynthetic organisms. Our results show the utility of high-resolution fs-LIMS volumetric imaging in combination with manifold learning methods in studying chemically diverse samples. Our approach provides a powerful analytical framework that can be applied further to the study of early/primitive life.

Introduction

The investigation of early/primitive examples of life has a profound effect on our understanding of life's origin and evolution and potentially has an impact on expanding our capacity to identify previously unknown or unrecognized evidence of early life. Searches for evidence of early life have intensified since the mid-1960s, when the first reports of Precambrian Gunflint microfossils were published (Barghoorn & Tyler, 1965). However, despite the quality and capabilities of modern analytical techniques which have significantly improved since the 1960s, debates about the metabolic speciation of some *bona fide* microfossils and the biogenicity of other putative fossils remain highly active (Brasier *et al.*, 2002; Schopf & Kudryavtsev, 2012). The inconclusiveness of investigations are largely caused by the poor preservation of morphological and chemical signatures of early life (Wacey *et al.*, 2016a). Thus, new

approaches and modern analytical methods, that are sensitive and accurate, have to be explored in the field of early life sciences (Wacey *et al.*, 2013).

The populations of Gunflint microfossils (1.88 Ga) (Ontario, Canada) (Awramik & Barghoorn, 1977; Barghoorn & Tyler, 1965; Cloud, 1965; Wacey *et al.*, 2012) represent one of the premier examples of Precambrian life (Alleon *et al.*, 2017; Wacey *et al.*, 2013). The Gunflint formation, providing high quality chemical and morphological preservation, affords a view into life in the Precambrian, which was evidently already complex and diversified (Wacey *et al.*, 2013). However, many questions remain regarding the Gunflint microfossils. The metabolic association of various microfossils and phylogenetic affinities are mainly inferred by morphological comparison to modern examples and by carbon isotope fractionation factors (House *et al.*, 2000) that are consistent with known metabolic pathways. It is thought that many of the Gunflint microfossils represent photosynthetic mat-building microbes (Awramik & Barghoorn, 1977; Barghoorn & Tyler, 1965; Lepot *et al.*, 2017), however, other interpretations exist (Planavsky *et al.*, 2009; Schopf *et al.*, 2015).

More generally, many questions remain in the field of early life sciences, where many potential examples of life are problematic due to morphological convergence and loss of the original chemical composition (Brasier & Wacey, 2012). Various processes can contribute to the formation of microscopic objects that morphologically resemble fossils but may not be of biological origin. For example, reduced abiotic carbon can migrate along grain boundaries forming lenticular structures of undefined origin, and the alteration of certain minerals can mimic microfossil morphology (Wacey *et al.*, 2016a). However, major and trace elements present within *bona fide* microfossils can serve as a comparative landmark, they can hold a piece of information about which chemistry is potentially expected to be preserved, and therefore provide an additional line of evidence about the biogenicity of a putative microfossil, even if parts of the morphology are lost. Moreover, modern machine learning frameworks provide access to unsupervised data-driven models (McInnes *et al.*, 2018; Nolet *et al.*, 2020), that can be of high utility in identification of various compounds from spectral datasets.

Laser ablation/ionization mass spectrometry (LIMS) is a promising surface characterization method that has recently experienced a revival and a wave of modernization (Azov *et al.*, 2020; Tulej *et al.*, 2021a). Modern laser ablation/ionization time-of-flight mass spectrometers (LIMS) are capable of providing an element and isotope characterization of the investigated spots in the ablation/ionization regime (Huang *et al.*, 2011; Riedo *et al.*, 2013b; Tulej *et al.*, 2021b) and molecular characterizations in the desorption regime (Cui *et al.*, 2013; Ligterink *et al.*, 2020; Moreno-García *et al.*, 2016). Microscale spatial resolution (Wiesendanger *et al.*, 2018) and nanometer depth resolution (Cui *et al.*, 2012; Grimaudo *et al.*, 2020) make such instruments of high utility in many scientific tasks (Liang *et al.*, 2017). The fast mass separation and ion detection technology (Riedo *et al.*, 2017) makes LIMS instruments applicable to the imaging tasks of various samples (Wurz *et al.*, 2020) – ranging from biological tissue characterization (Cui *et al.*, 2013) to mineralogical and chemical investigations of rocks, up to identification of impurities in dielectrics and interconnects from the semiconductor industry (Grimaudo *et al.*, 2019; Tulej *et al.*, 2021b; Wiesendanger *et al.*, 2018).

The current study has three goals that were set prior to the experiments. First – to identify the chemical composition of Precambrian Gunflint chert and test the instrument's (LIMS-GT) sensitivity to identify faint signatures of life from populations

of Precambrian Gunflint microfossils. Second – to test whether the majority of the microfossils found in stromatolitic layers within the Gunflint chert were photosynthetic bacteria. This is done by looking for traces of Mg in association with the microfossils, given that Mg can be a trace decomposition product of chlorophyll. Third – estimate the effectiveness of relational data analysis concerning large fs-LIMS spectral data and determine challenges and potential pitfalls concerning the required data transformations and hyperparameter heuristics.

The results of this study indicate that Gunflint microfossils have a distinct chemical composition that can be successfully identified using the LIMS-GT instrument. The low dimensional representations of the full spectral database (40'000 mass spectra) deliver a clear separation between main classes (bio-organic/inorganic spectra). Moreover, the fine structure of spectral similarity provides an insight into the composition of trace inclusions and the chemical diversity of the investigated microfossils.

Sample and methods

In this contribution, we have used a newly developed high-resolution fs-LIMS instrument to chemically characterize the Gunflint chert sample. The detailed description of instrument figures of merit and performance estimations on NIST standard materials have been recently reported, thus, we refer the interested reader to the technical article from our group (Wiesendanger *et al.*, 2019). A 30 µm thick thin-section of the Gunflint chert (1.88 Ga – Ontario, Canada) (Wacey *et al.*, 2016b; Wiesendanger *et al.*, 2018) containing populations of *bona fide* microfossils preserved in a silica matrix was used in this study (Alleon *et al.*, 2017; Barghoorn & Tyler, 1965; Cloud, 1965; Wacey *et al.*, 2012). Prior to fs-LIMS characterization, optical microscopy was performed, to specify the area of interest for detailed chemical investigations. Figure 1 shows a transmitted light microscope picture of the area (0.5 mm by 1 mm) chosen for the 3D mass spectrometric imaging. Dark patches represent agglomerations of individual microfossils. Within the area chosen for the mass-spectrometric investigation, we defined a grid (100 by 200 spots) with a 5 µm gap between ablation craters. An fs-Ti:Sapphire laser (775 nm, 180 fs, CPA Series, Clark-MXR Inc., Dexter, MI, USA) was used to ablate and ionize material from the Gunflint chert. The fs-laser was coupled to the mass spectrometer – a time-of-flight instrument with high mass resolution (m/Δm = 10,000) (Wiesendanger *et al.*, 2019). Within each of the probed locations, a sequence of 5 single laser shot mass spectra was collected, forming a 3D grid that consists of 100,000 mass spectra. Every single laser shot mass spectrum collected from the Gunflint chert sample was digitized using a high-speed ADC card (U5303A, Acqiris SA, Geneva, Switzerland) at 3.2 GS/s and resulted in the collection of 320'000 data points per spectrum. The data collection and imaging process has been optimized by directly saving the digitized mass spectra using a binary data format using an in-house developed software package. In order to increase the signal to noise ratio (SNR) of individual mass spectra from given pixels, two datasets were created from the original volumetric binary dataset – 1) Two consecutive mass spectra were averaged for a given pixel (first and second spectra where averaged as a first layer, third and fourth spectra averaged as a second layer - forming two imaging layers (20,000 spots * 2 layers)), discarding the last single laser shot spectra. 2) An

average of 5 single laser shot mass spectra have been calculated, forming a single layer image (fully averaged) of the investigated area (20,000 spots * 1 layer).
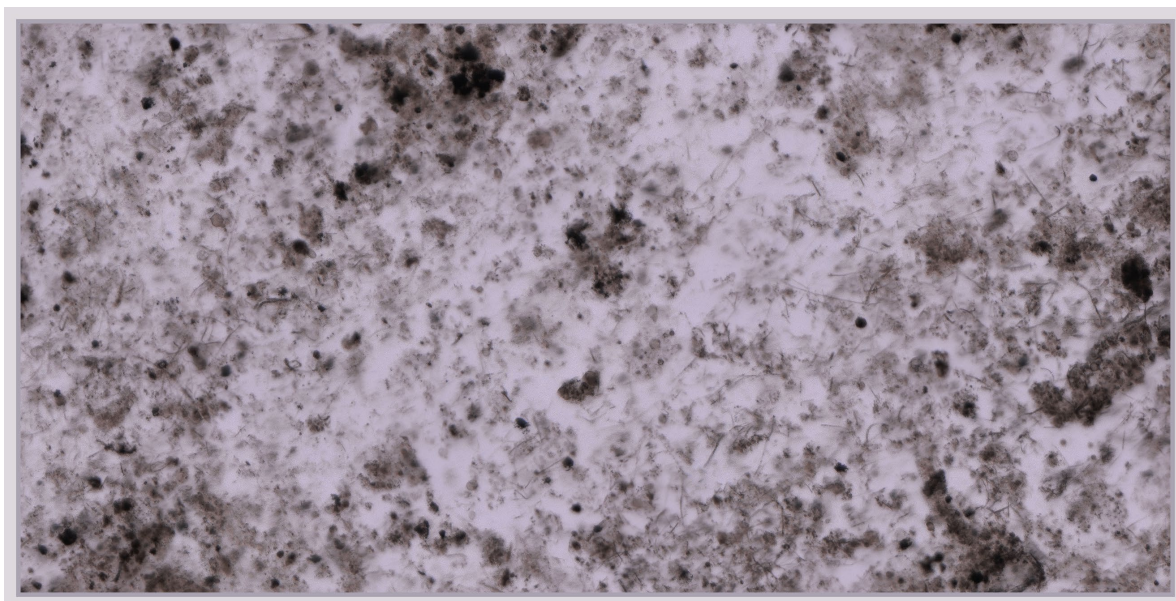

Fig. 1. Transmitted light microscope image of the area (0.5 mm by 1 mm) chosen for the volumetric imaging. Dark patches represent the agglomeration of individual microfossils. Microfossils of various states of preservation can be seen. Note, that the picture represents a multi-focus stack image, thus, not all microfossils are exposed to the surface. Various individual filamentous and spheroidal microfossils can be seen.

A spectral preprocessing routine was applied to every mass spectrum which includes baseline correction, denoising, and averaging (Meyer *et al.*, 2017). The single mass unit decomposition has been performed on the preprocessed mass spectra using Simpson integration of the mass peaks (Meyer *et al.*, 2017; Riedo *et al.*, 2013a). Overall, 260 single mass unit intensities (1-260 m/q) have been retrieved from the averaged mass spectra and 6 additional mass pairs have been integrated, to resolve some isobaric interferences, namely - $^{24}Mg/C_2$ (24 m/q), $^{52}Cr/C_4H_4$ (52 m/q), and $^{13}C/CH$ (13 m/q). After this step, mass spectra have been assigned with location indexes, forming a reduced volumetric grid – 266 mass intensities*40,000 pixels; and for an averaged image – 266 mass intensities *20,000 pixels. The SNR within the preprocessed mass spectra was identified to be ∼10^3, which is limited by the noise floor and the dynamic range of the acquisition card. Further, the dataset has been log-transformed and divided into subsets for imaging and low dimensional analysis. The data reserved for the low dimensional analysis (3D and 2D images) was z-score normalized and reduced with principal components analysis (PCA) down to the first 100 principal components.

Further, to improve the image quality and the readability, a volumetric dataset was interpolated from the original size (100 pixels *200 pixels*2 layers) up to 250 pixels *500 pixels*5 layers using inverse distance interpolation. Low-intensity zones of the imaged isotopes have been made translucent to improve the visibility of spatial heterogeneity. The inverse distance interpolation has been performed using the log-transformed data, thus, color changes present in the pictures are logarithmic (base 10). Overall, using the data preprocessing routine, we have calculated 8 volumetric maps of isotopes of interest. Each volumetric map is represented by 625'000 voxels, which provides enough resolution to see the small discrepancies within the analyzed

area. The depth of the formed craters has been estimated using optical microscopy, which has shown, that 5 single consecutive laser shots applied to a given ablation position created ablation pits that are below or close to the µm range in depth. Thus, volumetric maps characterize the uppermost layer (nm-scale) of the Gunflint sample. It has to be mentioned, however, that due to the sharp difference in the absorption properties of the investigated microfossils and the host material (quartz), differences can be observed in the amount of the ablated material.

The dataset prepared for the low dimensional analysis (normalized and PCA reduced to the first 100 components) was further analyzed using the Uniform Manifold Approximation and Projection (UMAP) method. The UMAP provides dimensionality reduction and groups similar spectra together. Here we outline some of the sub-steps of the method, for the full and thorough description we refer to the original UMAP manuscript (McInnes *et al.*, 2018).

As in many other manifold learning methods, the UMAP algorithm tries to preserve small distances, thus, recovering the local connectivity of data and it's intrinsic dimensionality. In the first stage of the algorithm, similarity graph construction is performed using the approximate nearest neighbor (NN) descent (Dong *et al.*, 2011), implemented within the UMAP (using small minimal distance of 0.1, 5 nearest neighbors, and cosine distance as a metric). Although the recall rate of nearest neighbors is reported to be high, it is not exact. The approximate NN search makes the algorithm fast, but it also includes the possibility of small mistakes in the determination of nearest neighbors. Even though exact solutions exist, for example, the GPU accelerated UMAP implementation provides such an option (Nolet *et al.*, 2020), approximate NN-search within the original UMAP-learn python implementation (McInnes *et al.*, 2018) provides enough accuracy for the typical fs-LIMS tasks. Further, in the second stage, the algorithm weighs the nearest neighbors and forms the weighted NN-graph using smoothing kernels that adapt to the local neighborhood. In the third and last stage, UMAP performs a graph optimization by running a stochastic gradient descent for a determined number of epochs, decreasing the loss between low-dimensional and high-dimensional distances. The end stage (stochastic gradient descent) essentially represents the force-directed type optimization of weighted neighborhood graph. The starting positions of the nodes are defined by Laplacian eigenmaps (spectral embedding) (Belkin & Niyogi, 2003). The usage of spectral embedding as an initialization step provides good means for an improved preservation of the global structure of the investigated manifold (Kobak & Linderman, 2019). The final step of the UMAP algorithm performs the embedding of the graph into a new coordinate system, which better reflects the similarity of n-dimensional vectors. Further, to define the classes from the low-dimensional embeddings, a density-based, hierarchical clustering method (HDBSCAN) (McInnes *et al.*, 2017) was used, which provides a clustering hierarchy and derives clusters of spectra that share significant degree of similarity. And lastly, for the derivation of reduced topological representations (in the form of a network) of the UMAP embedding, a Mapper algorithm (Singh *et al.*, 2007; Van Veen *et al.*, 2019) was used, which covers the original data with overlapping filter functions, that are used to form the network and capture local connectivity of the data.

Results

The mass spectrometric investigation of the Gunflint microfossils using the high-resolution fs-LIMS instrument yielded a large amount of data – 266 volumetric mass intensity maps. Since most of the known Gunflint microfossils are organically preserved, we have used ^{12}C as an indicator of microfossils. Figure 2 (upper panel) shows

the distribution of ^{12}C signal acquired from the Gunflint sample. The 3D map reveals a high degree of co-occurrence between ^{12}C intensities and the distribution of micro-fossils identified using optical microscopy (see fig.1). The middle panel in Figure 2 shows the distribution of ^{31}P intensities, which also reveal a relatively high degree of co-occurrence with ^{12}C and with the distribution of microfossils (from the optical picture of the Gunflint chert, see Figure 1 and lower panel in fig.2). However, ^{31}P is mostly present only within the areas of high ^{12}C concentrations, which indicates that for the most part, ^{31}P remained below the limit of detection.

The co-occurrence of these two elements (^{12}C and ^{31}P) indicates that the measured carbon is likely of biological origin. However, there is a large part of the sample where we have observed a noisy distribution of both elements, and therefore it is hard to definitively conclude where exactly areas with the organic carbon are located. More-over, the host mineral and the microfossils seem to be fused, from the analysis of the 3D maps. Similarly, the histogram of carbon intensities (not shown) could be charac-terized as a skewed distribution, where the spectral ^{12}C intensities are smoothly tran-sitioning from the noise level to the intensities with ^{12}C saturation. Similar distribu-tion could be observed for the ^{31}P and other elements, where we do not observe any clear distinction between two separate entities – host (quartz) and microfossils (or-ganic carbon). Thus, deterministic identification of organic spectra solely on the basis of the presence of biorelevant elements is possible for the high-intensity end mem-bers of the distributions, but localization of the exact boundary between two classes is very challenging.
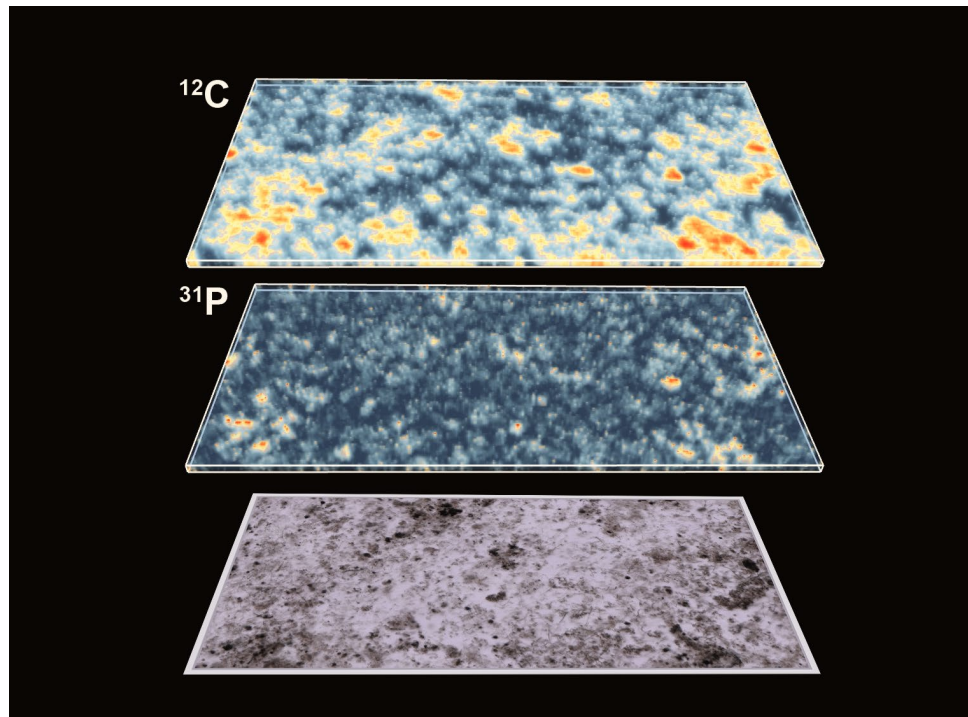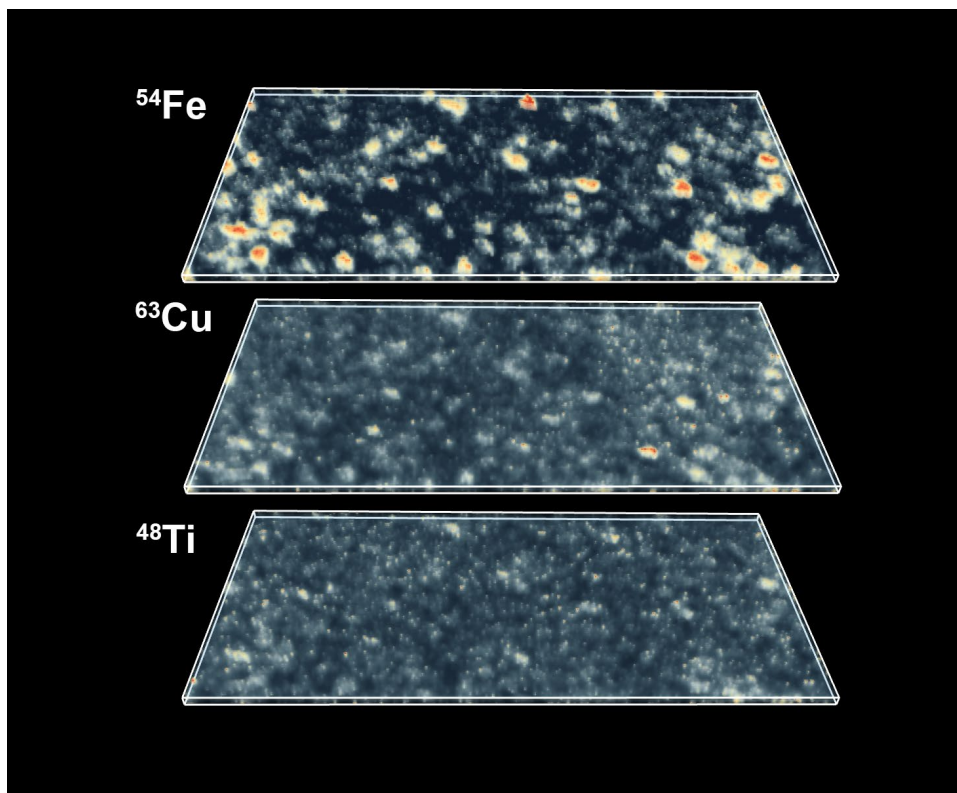

Fig.2. Volumetric isotope distribution maps from the Gunflint chert. Upper panel – Dis-tribution of ^{12}C (red color indicates higher concentrations). Middle panel – distribution of ^{31}P (red color indicates higher concentration). To compare maps with the optical im-age, see lower panel. Note that the distribution of microfossils and maps of ^{12}C and ^{31}P correlate with each other.

To identify locations of metallic inclusions we calculated three additional maps. Figure 3 shows the distribution of ^{54}Fe, ^{63}Cu, and ^{48}Ti volumetric intensities. The

distribution of ^{54}Fe, largely coincides with the distribution of ^{12}C (see fig.2), thus Fe can be associated with the microfossils. However, some of these locations can also be attributed to the siderite (FeCO$_3$). In contrast, localization of high ^{63}Cu and ^{48}Ti areas differ from the ^{54}Fe map and indicate the presence of chemical heterogeneity. One can see bright inclusion with high concentrations of ^{56}Fe in the upper middle part of the sample. Similarly, high ^{63}Cu micro-inclusions can be observed in the lower right part of the sample. The ^{48}Ti localization reveal presence of the small grains, dispersed throughout the sample. Overall, the metallicity present in the Gunflint chert shows relatively fine-grained mineralization which is disseminated throughout the sample.

Fig.3. Volumetric isotope distribution maps from the Gunflint chert. Upper panel – Distribution of ^{54}Fe (red color indicates higher concentrations). Middle panel – distribution of ^{63}Cu (red color indicates higher concentration). Lower panel – distribution of ^{48}Ti (red color indicates higher concentration). To compare maps with the optical image, see fig. 1.

Figure 4 shows three panels with different volumetric isotope distribution maps. The upper panel shows the distribution of ^{32}S isotope intensities. In the upper part of the volumetric map, we have identified a ^{32}S saturated inclusion, which was interpreted to be a pyritic (FeS$_2$) inclusion (see further in the text), intermixed with organic material from the microfossils (relatively bright spot in the ^{12}C map as well, see Figure 2). However, this position can also represent a small pod of pyritized microfossils previously reported in (Wacey *et al.*, 2013). In addition to the pyritic inclusion, we have observed another inclusion saturated in ^{32}S, which is located in the lower-left part of the sample. One can compare the location of this inclusion with the volumetric map of ^{107}Ag, shown in the middle panel (Figure 4). The exact spatial localization of ^{32}S and ^{107}Ag, indicates the presence of typical low-temperature hydrothermal sulfide inclusion – acanthite (Ag$_2$S), which again, seems to be sampled with organic remains of the microfossils. The third inclusion that could be observed from the ^{32}S

map located in the lower right side of the map, is highly correlated with the ^{63}Cu isotope (see fig.3 middle panel), indicating the presence of another hydrothermal sulfide mineral – chalcopyrite ($CuFeS_2$). In addition to the three sulfidic inclusions, one can observe a fine-grained distribution of S-containing material. In contrast to the localization of inclusions, the fine-grained lower intensity ^{32}S signal largely follows the structure outlined by the ^{12}C intensity map, so the observed S in that area is likely of organic origin.

The middle panel of Figure 4 shows the distribution of ^{107}Ag intensities. In contrast to the inclusions of S, silver mineralization appears to be relatively rare and localized in a few grains. For example, in addition to the main acanthite inclusion, one could observe a fine, enriched area in the lower-left part of the sample. However, since the intensities of ^{107}Ag are relatively small in those areas, silver might be present as an impurity in other sulfide inclusions. The third map shown in Figure 4 presents the distribution of ^{139}La intensities. The brightest spots in the lower right part of the volumetric map indicate the localization of inclusions rich in rare earth elements (REE) - ^{139}La and ^{141}Pr. The mineralogical association of the inclusions can be attributed to the phosphate minerals due to the co-occurrence of P and La (see fig.2). However, main constituents of the analyzed spots are Na and K (see further in the results section).
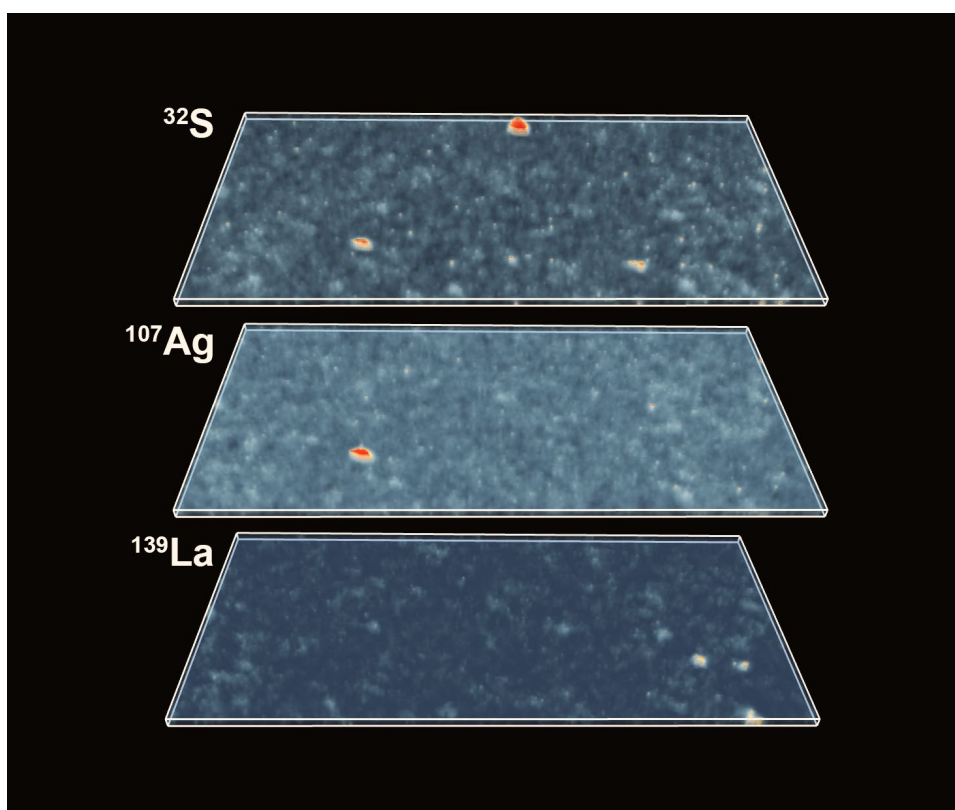
Fig. 4. Volumetric isotope distribution maps of ^{32}S, ^{107}Ag, and ^{139}La from the Gunflint chert. The upper panel – distribution of the ^{32}S indicates the presence of localized sulfide inclusions (red color indicates higher concentrations). Smaller intensities of ^{32}S co-occur with ^{12}C (see Figure 2). Middle panel – distribution of ^{107}Ag indicates the presence of Ag

sulfide (Acanthite). Lower panel – distribution of ^{139}La indicates the presence of REE mineral inclusions in the Gunflint chert.

The transitional structure of all investigated maps represents the gradual change from the host material to the elemental and isotopic composition of the inclusions or microfossils. It is reasonable to conclude that this effect present within the volumetric maps is due to interpolation effects (upsampling to the higher resolution), however, we see that the same effects are present within original data as well. The same patterns were observed to be present within other element maps – gradual skewed transitions. The analysis of the full volumetric dataset (to search for potentially missing inclusions) will result in a very high workload (260 maps for comparison) and likely will be counterproductive. Additionally, as mentioned before, it is not clear where to define the borderline within the intensity profiles which will outline the different minerals. However, the mass spectra can be considered as high-dimensional vectors, and latter can be compared in terms of distances. Assuming that similar minerals and compounds will yield a small pairwise distances, thus, the high similarity of ionization profiles. In order to find similar spectra within the dataset and approximate mineralogical composition, we performed the analysis of spectral neighborhood using the cosine metric and UMAP manifold learning method.

Figure 5A shows 40'000 (2 layers) fs-LIMS spectra sampled from the original 260-dimensional space and reduced down to three dimensions using UMAP. The colors assigned to the data points show the distribution of the ^{12}C intensities on the low dimensional manifold. The structure of the similarity network revealed by UMAP indicates the presence of three large groups – 1) The matrix mineralogy represented with quartz mineral. 2) The extended body of spectra that belong to the microfossils. 3) A smaller group of spectra with a more complicated similarity structure registered from various inclusions. The distribution of the ^{12}C intensities in the microfossils cluster indicate that microfossils are sampled with varying volume - some are highly enriched in ^{12}C, some are highly depleted. The cluster of inclusions outlined in the figure reveals a high degree of ^{12}C saturation (compared to quartz) and relatively high proximity to the microfossils cluster, which indicates that spectra registered from these locations are a mixture between microfossils and some additional chemistry (see further in the text).

The panel on the right, Figure 5B, reveals the same embedding as in the left panel but colored using the ^{24}Mg intensities. As was briefly presented in the introduction, it is hypothesized that if microfossils are indeed of photosynthetic origin, they might have preserved some traces of Mg within their bodies and collapsed cell walls (Mg is a crucial metal for the operation of the chlorophyll complexes and photosynthesis in general). Figure 5B indicates that ^{24}Mg (being ~3 orders of magnitude lower abundant than ^{12}C) is present in the spectra that are identified to be from microfossils and represents a trace element. The spectra identified from the cluster of inclusions are also showing a relatively high degree of ^{24}Mg ion yield. The full mass scale spectral convolution reveals that similarity diverges from the host mineralogy and inclusions. The so-called "transition line", see fig. 5B, indicates a detached progression of spectral similarity between the host and the microfossils. Interesting to note is that the spectral neighborhood shows a progression of the spectral similarity from poorly sampled microfossils (small parts of the microfossils are ablated and ionized) up to the point of ^{12}C saturation and subsequent hydrocarbon chain formation (higher percentage of the ablated and ionized material is from microfossils), which is reflected in ^{12}C intensities. Thus, the low dimensional representation of the volumetric imaging dataset

reflects not only the chemical differences but also provides additional insights into processes within the evolving plasma and ionization. The higher ^{12}C and ^{24}Mg concentrations also reflect the progression of the SNR, which indicates that pixels that are sampled with a good volumetric ratio (host/microfossils) show higher Mg intensities.
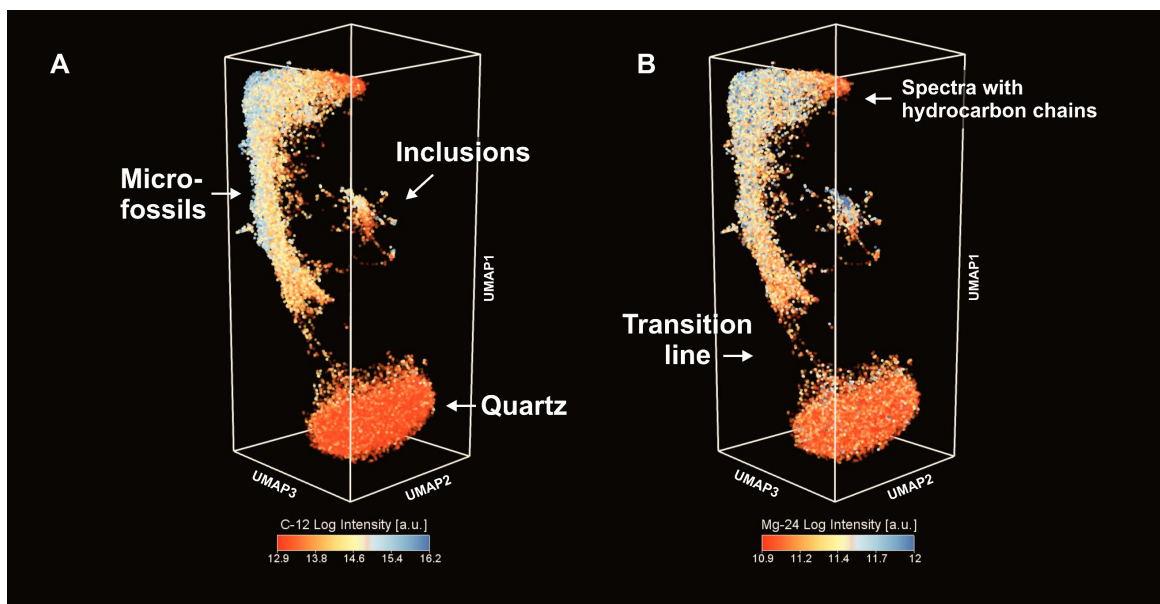

Fig. 5. The low dimensional proximity of fs-LIMS spectra revealed by UMAP. The individual data points represent a single mass spectrum. Full volumetric (40,000 mass spectra – 2 layers) data has been considered for the analysis. Note the presence of three large clusters. Figure 5A – The proximity of fs-LIMS spectra colored according to the ^{12}C intensities. Clusters associated with the microfossils and inclusions reveal a high level of ^{12}C saturation. Figure 5B – The proximity of fs-LIMS spectra colored according to the ^{24}Mg intensities. Clusters associated with the microfossils and inclusions reveal a high level of ^{24}Mg saturation. Note, that the most enriched in ^{24}Mg spectra are located in the inclusions cluster.

Figure 6 shows the distribution of ^{32}S (fig.6A) and ^{31}P (fig.6B) intensities within the UMAP embedding. The elevated concentrations of S and P coincide with the locations of high C and Mg signal. Relatively low intensities of these elements in the microfossils cluster indicate that P and S are present as trace elements. For example, the sulfides cluster identified in fig.6A, reveal high S concentrations, and reflects the positioning of sulfides in the embedding. Thus, the elongated cluster of microfossils (see figs.5 and 6) reveal multielement enrichment of C, P, S, and Mg within the same spectra and also shows a high degree of spectral similarity (calculated on the basis of full mass range) using the cosine metric. The cosine similarity represents the cosine of an angle between two compared mass spectra and determines whether two vectors are pointing in the same direction.
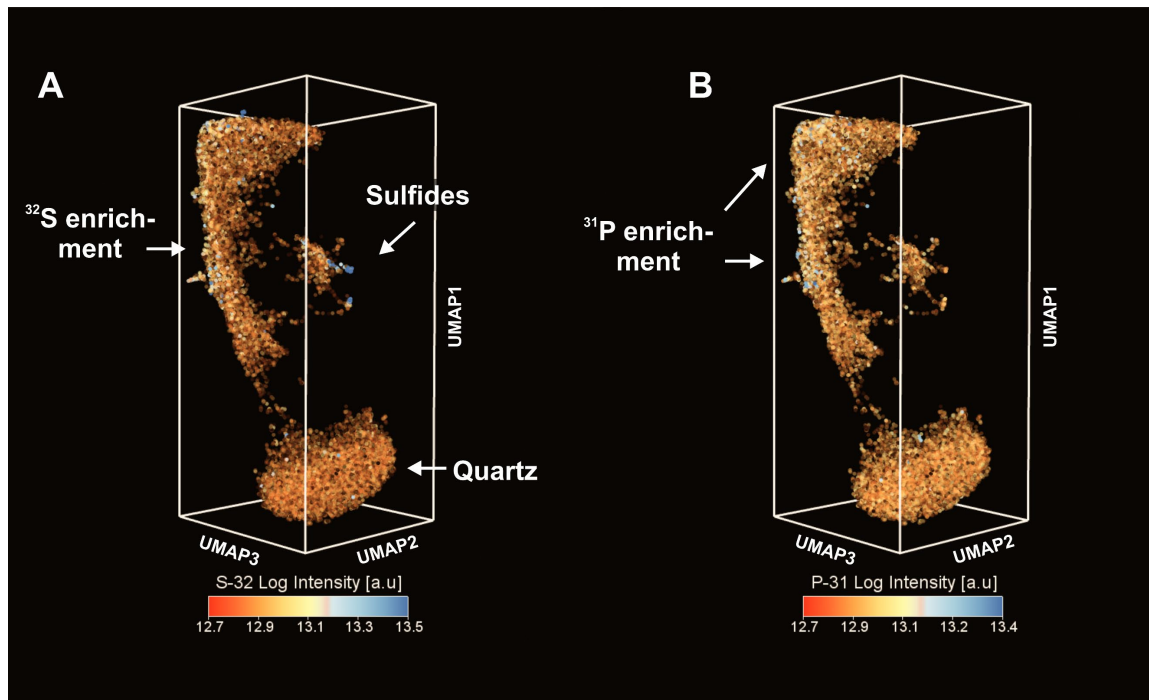
Fig.6. The low dimensional proximity of fs-LIMS spectra revealed by UMAP. Figure 6A – The proximity of fs-LIMS spectra colored according to the ^{32}S intensities. Cluster associated with the microfossils reveal higher concentrations of ^{32}S. Figure 6B – The proximity of fs-LIMS spectra colored according to the ^{31}P intensities. Cluster associated with the microfossils reveal high level of ^{31}P saturation. Note, that P and S enriched spectra also enriched in ^{24}Mg (see fig. 5B).

In general, the main clusters are easily separable. However, the smaller cluster of the inclusions shows a higher degree of complexity of the similarity structure. This complexity hints toward the presence of multiple inclusions of varying chemistry (see Figures 3 and 4). In order to define the boundaries between classes, a density-based, hierarchical clustering method (HDBSCAN) (Campello *et al.*, 2013; McInnes *et al.*, 2017) was used, which provides a clustering hierarchy, from which a simplified tree of significant clusters can be derived. Figure 7A shows the result of clustering of all 40'000 mass spectra registered from the Gunflint sample (2 layers). One can note that we have identified 4 clusters in the first iteration of clustering. These clusters correspond to the large body of spectra registered from the microfossils (dark blue data points), from the matrix (quartz – purple points), from the inclusions (light blue points), and the REE inclusions (brown data points). The apparent similarities in the spectra that are transitioning from the microfossils to the inclusion cluster indicate volumetrically mixed sampling of microfossils and portions of the inclusions. In order to improve the quality of embedding for the inclusions cluster, the original mass spectra were extracted and an additional UMAP embedding with a higher number of epochs (optimization iterations) using cosine metric and a small number of nearest neighbors (5-NN) was performed. The result of the embedding with interpretation is shown in Figure 7B. Since some of the clusters appeared to be fused into each other, but with distinguishable density profiles, kernel density estimates (KDE) for all identified clusters were calculated. The clusters identified in the upper part of the embedding represent a group of various sulfides that have been briefly introduced before, namely, silver sulfides (acanthite), iron sulfides (pyrite), and copper sulfides (chalcopyrite). The blue kernel (upper part) represents the pyrite inclusion (see Figures 3

and 4, the upper panels), the purple kernel indicates the chalcopyrite inclusions, and lastly, the orange kernel – spectra identified from the acanthite minerals. Further down in the embedding, the largest cluster of Fe-rich microfossils (orange kernels) could be identified. The topology of the similarity network indicates that there is a transition of Fe-dominated spectra towards the Cu-sulfides (chalcopyrite), which indicates the presence of Cu and Fe mixing line. On the right side of the cluster assigned to the Fe-rich microfossils a cluster of Ti-rich microfossils is located (blue kernels). The Ti-rich microfossils cluster also indicate the there is a gradual shift in similarity, which represents a mixing of Ti towards Fe. And lastly, a small cluster of Al-rich microfossils is located close to the cluster of Iron-rich microfossils. The presence of Al and Fe together can indicate the presence of Fe-Al-silicates (i.e. clay minerals) in association with the microfossils. Adding the previously described REE inclusions with quartz and microfossils we have identified 9 groups of spectra that are present in the investigated part of the Gunflint sample (0.5 mm by 1 mm) which reveal a distinct similarity measures.
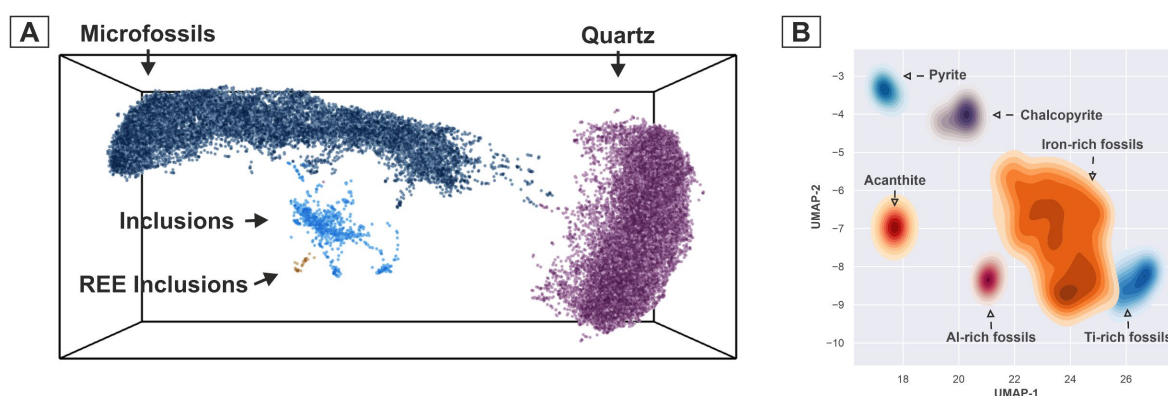
Fig. 7. The clustering results of the UMAP scores. A) The clustering of the dominant components. Data points are colored according to the assigned cluster. The individual data points represent a single mass spectrum. B) Additional sub clustering of spectra present in the group "Inclusions" as marked in Figure 7A. The clustering results reveal three groups of sulfide inclusions (Pyrite, Covellite, Acanthite) and three clusters of spectra with enhanced metallicity (Fe-rich, Al-rich, Ti-rich).

The analysis of the spectral neighborhood provides a means for unsupervised data-driven classifications; however, the understanding of the chemical discrepancies has to be done with the original spectra. For the accurate characterization of identified classes, the spectra of every given class were averaged into single mass spectra. Figure 8 reveals the averaged mass spectra from all identified classes. All spectra shown in the figure are normalized to the maximal peak intensity - from 0 to 100 [a.u.]. With the aim to deliver more details, spectral intensity profiles have been truncated down to the range from 0 to 50 [a.u.]. For most cases, the most intense line was ^{23}Na, followed by ^{39}K, and ^{28}Si as main constituents of the silica matrix. The large signals of Na and K can be attributed to the low ionization potential of the given elements. To deliver a better understanding of the spectral intensity profiles and results of our interpretation, the reader has to acknowledge that most of our spectra represent a volumetrically varying mixture of all components present in the ablated spot. Since the microfossils might contain nano mineralization, trace elements, and impurities, and they likely are sampled with some parts of the silica matrix, spectral

profiles after averaging might be rather complicated and represent the intermixed chemistry from various sources.

The first three spectra from the upper panel (fig. 8) reveal the chemical compositions of previously identified sulfides. The first cluster of Cu sulfides shown in the upper left panel is dominated by Cu with the minor contribution of Fe and Cr. The ^{32}S concentration is rather low, however, we have to consider the fact the mass spectrum presented in the figure is an average of ~100 pixels (200 laser shots), which provides means for the drift of original ratios. Additional to the peaks of sulfides, one can observe the presence of a relatively intense ^{12}C peak and a significantly lower amount of Si, Al, and high K and Na. The presence of intense C, O, chains of hydrocarbons, and the proximity of spectra to the microfossil cluster within the main embedding (Figures 5, 6 and 7) provides a piece of evidence that the spectra are registered from the mixture of organic material, silica matrix and a chalcopyrite inclusion. The next spectrum shown in the middle part of the upper panel is an Acanthite inclusion. One may see that the Ag, C, Na, Al, Si, and K lines are well above the 50 [a.u.] threshold. Relatively high concentrations of S, C, Cr, Fe, Ti, Cu, and Cl may be noted. On the right side of the upper panel the spectrum registered from the pyrite inclusion is shown. A high concentration of S and Fe with virtually no other significant metallic elements indicate that the spectra are registered from the pyrite. Among the trace elements within the pyritic inclusion, one could note the presence of Pb.

The second row of spectra in Figure 8 shows the classes that could be characterized as spectra from the microfossils with significant metallic content (see further in the discussion). First in the second row of spectra with significant metallic content is the spectra registered from the Fe-rich-microfossils – the largest group of all among the inclusions. Interesting to note that the spectrum shows relatively high concentrations of Cr. Moreover, the spectrum presents the full range of aforementioned metallic elements like Ti, Cu, Zn in minor concentrations. Another group identified from the embedding shown in the middle part of the second row is Ti-rich microfossils. The averaged spectrum reveals a high concentration of Ti as a main metallic constituent, followed by Fe, Cr, and Zn. The Al-rich spectra shown on the right side of the middle panel reveals a high concentration of Al, followed by significantly reduced Fe and Cr.

The last row in Figure 8, the left panel, outlines the composition of the REE inclusions – spectra with a relatively high concentration of ^{139}La and ^{141}Pr. Among the main constituents of these inclusions only Na, K, Ca can be noted. The minor concentrations of C, Al, Si, Cr, and Fe are also present in the spectrum. The penultimate spectrum present in our interpretation model represents the Quartz mineral from the silicate matrix of the Gunflint chert. The main element observed within the Quartz mineral is Si which is followed by K and Na, C, Ca, and minor levels of Si oxides. The spectrum represents an average of 100 pixels randomly sampled from the host (quartz) cluster. This was done to compare the spectrum with the equal number of spectra from the microfossils. This brings us to the last big cluster of microfossils – the spectrum presented in the last row and the right panel. As one may notice, the spectrum in the linear scale reveals a very similar chemical composition to the spectra presented before with one notable difference – C is the most abundant element within this group. We have intentionally averaged the first hundred highest ^{12}C intensity spectra since

they represent the best volumetric sampling of the microfossils and result in the best SNR.
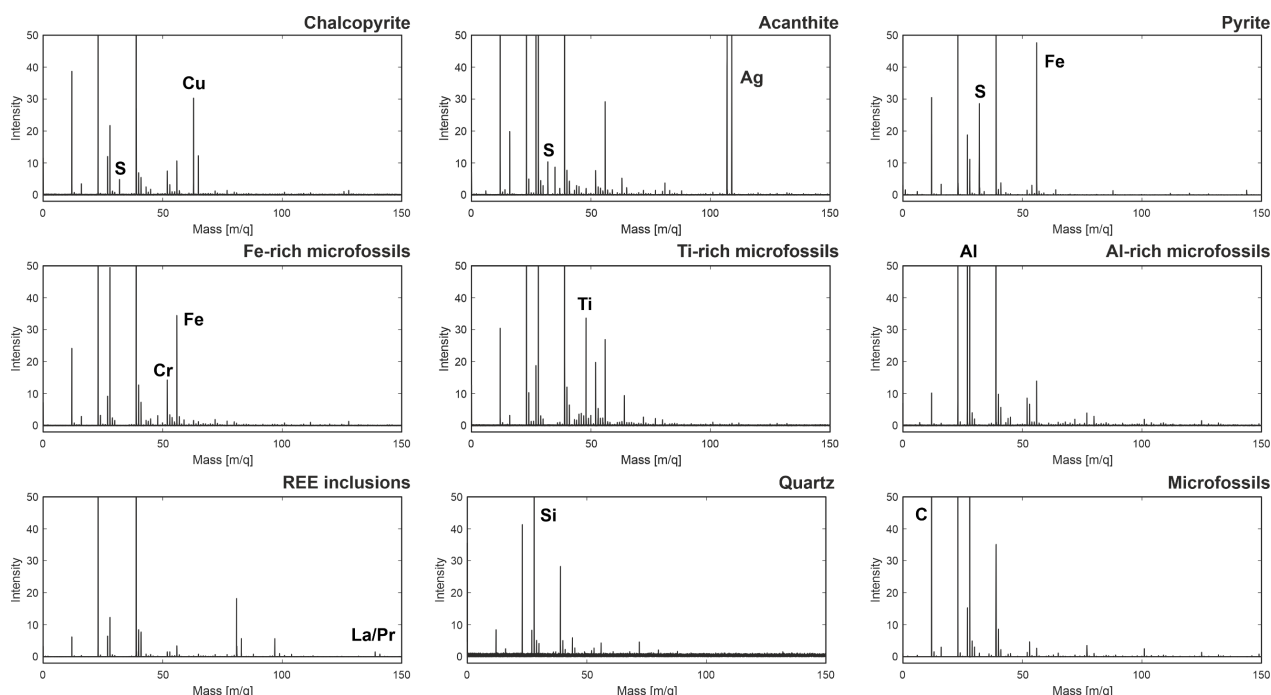

Fig. 8. The spectral classes identified from the UMAP embeddings. The spectra are averaged and normalized to the maximum peak intensity (0 to 100 [a.u.]). To enhance the visibility of the small peaks the intensity profiles are truncated to 50 [a.u.]. First row – spectral profiles of sulfide inclusions. Second row – spectral profiles of microfossils with enhanced metallicity. Third row – spectral profiles of REE inclusions, silicate matrix (quartz), and microfossils.

The logarithmic scale visualization of the same spectra from the host and the microfossils is shown in Figure 9. The SNR of the averaged spectrum from quartz is significantly lower in comparison to the spectrum registered from the microfossils (also visible in figure 8, bottom center panel). The SNR within quartz is only 10^2 in comparison to the microfossils – 10^4 for 200 laser shots averaged within each location. Lower signal recorded from quartz can be attributed to the lower absorption of the laser energy by quartz. Moreover, the complexity of the chemical composition and total amount of detected peaks is significantly different. The spectrum from the microfossils can be characterized as heavily dominated by hydrocarbon C_xH_x-type compounds and silicon oxide polyatomic species. The trace element composition reveals Mg, Cl, Ti, Cr, Cu, and Fe as being the most dominant ones. Among the more abundant elements, one can note the presence of elevated amounts of Ca, Na, K, and Si, whereas the latter (Si, K, Na) are signatures of the host mineralogy that is co-ablated with the organic inclusions. Due to the enhanced absorption of the fs-IR light by dark patches of microfossils, the ablation efficiency of the host mineral also increases. This phenomenon leads to the greater intermix between the host and the chemistry of the microfossils. Overall, Figure 9 offers a visual assessment of the spectral dissimilarity and provides insight and intuition into the mechanics of the spectral proximity. The interpretation of the UMAP embeddings, in short, follow the outlined logic - as soon as the laser hits the spot with the microfossils or spot with diluted organic remains, output

spectrum even with low volumetric sampling will form complex multipeak spectra, which in turn will yield a low similarity rank in comparison with the host spectra.
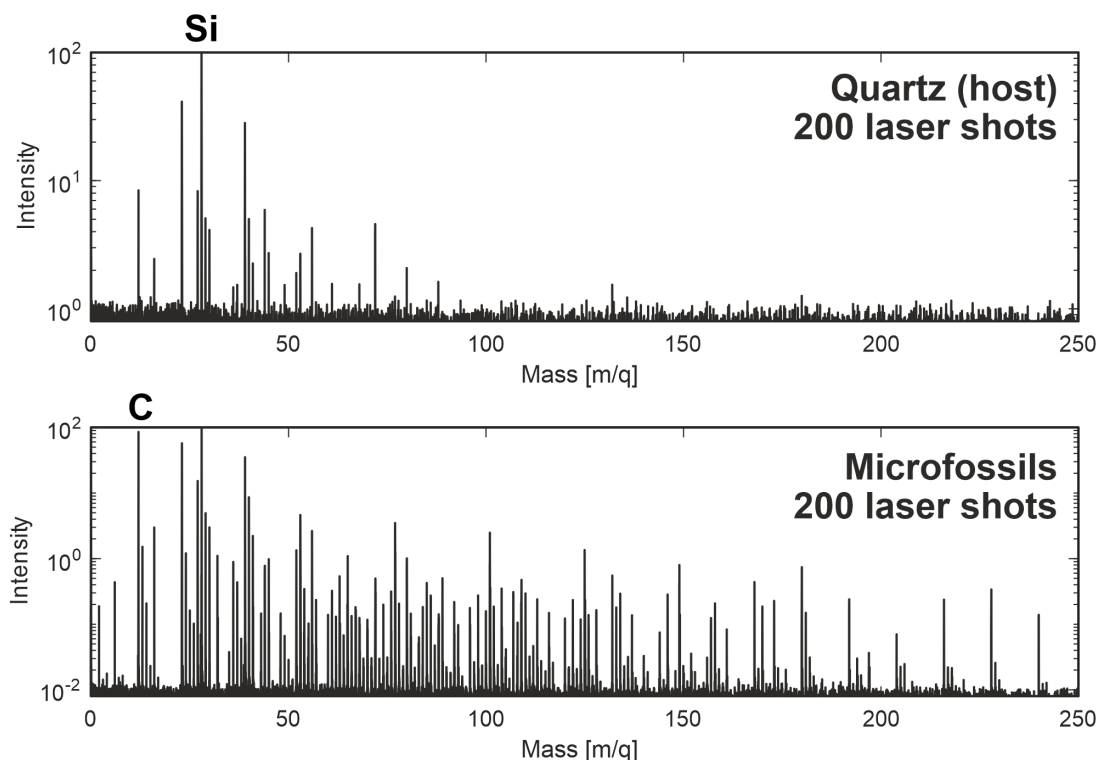

Fig. 9. The fs-LIMS mass spectra (averaged from 200 laser shots) sampled from quartz and microfossils. The mass spectra are presented in the logarithmic scale. Note the difference in the ion yield and changes in chemical composition.

Spatial interpretation of the volumetric maps.

In previous sections, the description of low dimensional embeddings of the 3D imaging dataset (2 layers) was provided as well as a description of identified clusters. It was shown that by using cosine spectral similarity it is possible to advance to the deterministic identification of minerals and entities present in the mass spectrometric image. In contrast to the "classical" data analysis, where one can compare distributions of the various ion yields and try to solve the classification problem using probabilistic approaches (i.e., using logistic regression), relational data analysis provides means to find more details and structure within complex datasets. For example, the popular ordination method - PCA does not provide any clear boundaries between different classes for the given dataset, contrary to the UMAP embedding results. However, the deterministic model and the interpretation that was built around it still lacks the spatial aspect. In order to understand how various clusters are spatially localized, the cluster map shown in Figure 10 can be utilized. The figure illustrates the distribution of the first three clusters. Laser ablation positions that are identified to be from the microfossils are colored with yellow nodes. The spectra clustered as inclusions with REE are colored orange. The grey nodes are sampled from the quartz matrix. To compare the distribution of the microfossils on the interpretation map and the optical image see Figure 1. As one may note, spatial localization of the microfossils largely follows the same structure captured in the optical image. The datapoints with dimmed colors represent the spatial difference in the class assignment. The

localization of the REE spots largely follows the initial interpretation made on the basis of the volumetric maps (see fig.4), with the addition of a few other pixels.
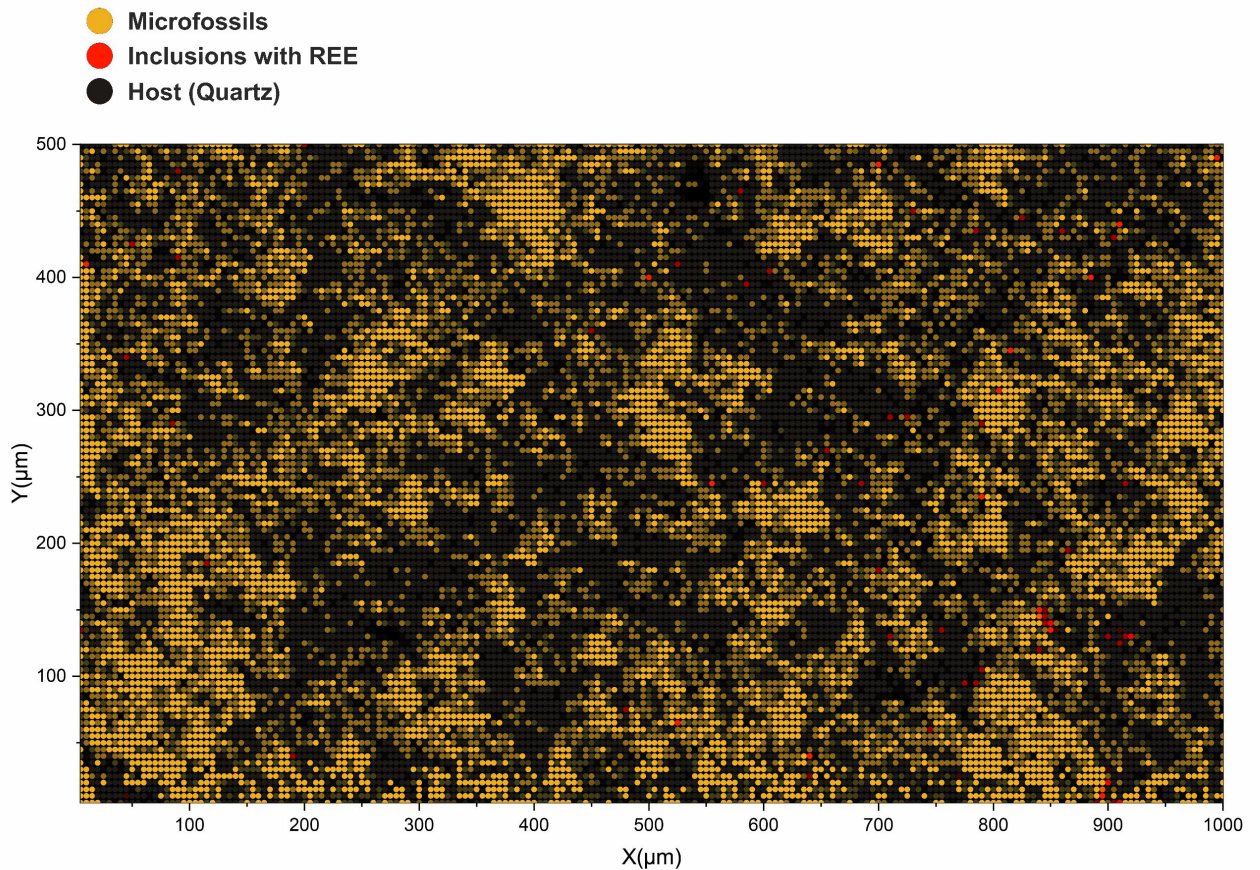

Fig. 10. The interpretation of volumetric imaging results using cluster assignments calculated from low dimensional embeddings. Yellow circles denote spectra registered from the organic remains of the Gunflint microbes. Grey circles denote spectra registered from the host mineral. Red circles denote locations of spectra registered from spectra with REE. Note, that nodes are translucent and changes in the cluster assignment reflect the volumetric inhomogeneity of the sample.

Figure 11 reveals the spatial localization of the six remaining clusters. The interpretation map presented in this figure shows only the second layer, due to the fact, that data from the surface contains very few metallic spectra and therefore will only clutter the figure. The yellow nodes disseminated through the picture represent the spatial localization of the Fe-rich fossils (cluster with high Fe and Cr). The red nodes mainly localized in the upper part of the picture represent the spectra registered from the pyritic inclusions. The blue nodes localized in the lower left part of the sample are registered from the silver sulfides (acanthite). The orange nodes - spectra from the Cu-sulfides (Covellite). The green nodes – Ti-rich microfossils and the purple nodes represent the spectra from the Al-rich spots. Note, that all previously described sulfide inclusions are present in the final interpretation map. Another noteworthy fact is that the spectra located on the upper layer, in most cases, belong to the group of microfossils. In general, the spatial distribution of the iron minerals reflects the fact that iron input was the most dominant among other metals.
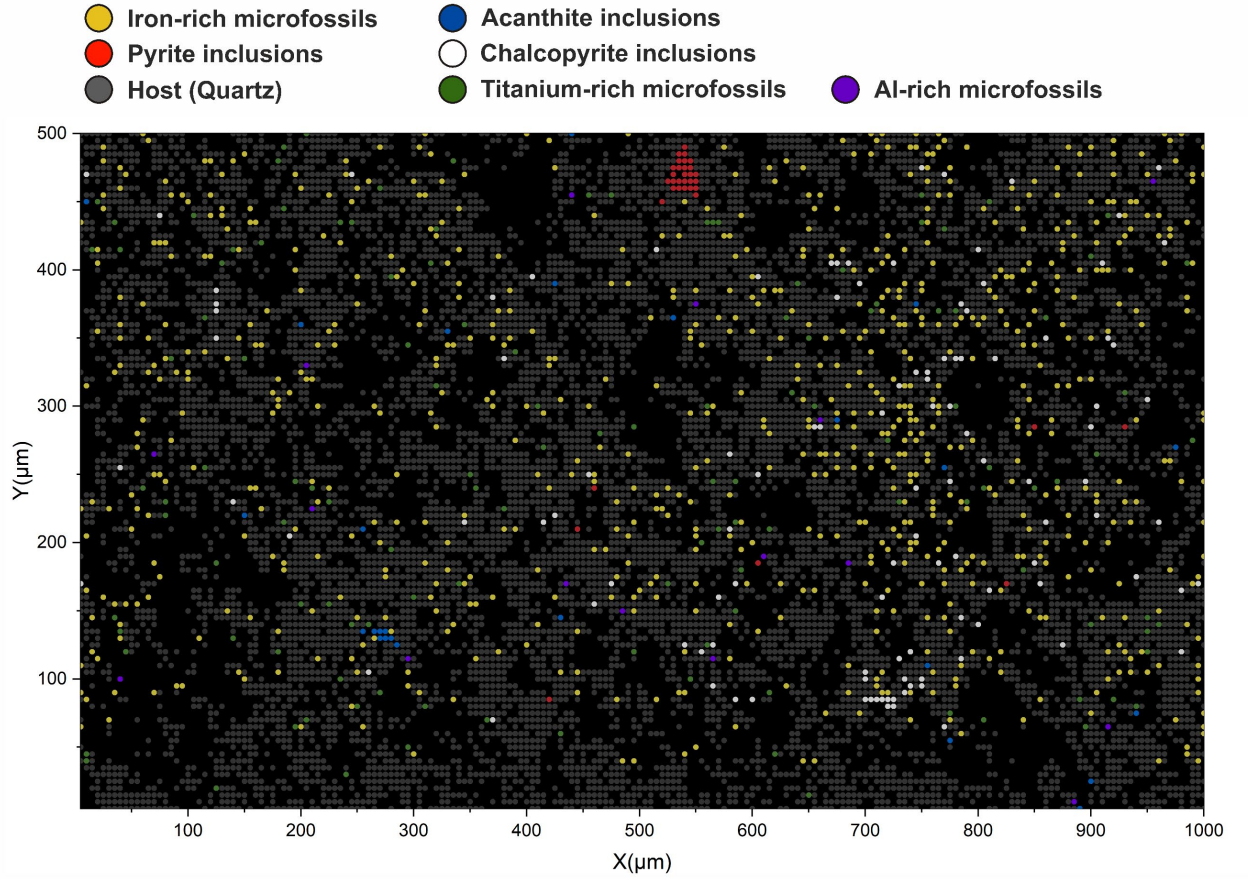
Fig. 11. The interpretation map of the bottom layer. Yellow circles denote spatial localization of spectra registered from the Fe-rich microfossils. Red circles denote spectra registered from the pyrite inclusions. Grey circles denote locations of spectra registered from the silicate matrix (quartz). The blue circles denote locations of spectra from acanthite minerals. The orange circles denote locations of spectra from chalcopyrite minerals. The green circles denote locations of spectra from Ti-rich microfossils. The purple circles denote locations of spectra from Al-rich microfossils. Note, that the majority of spectra assigned with high metallicity from the bottom layer are clustered as microfossils on the surface layer, thus, to declutter the interpretation map only the bottom layer is shown.

Averaging dynamics and secondary features derived from the topology of the spectral proximity networks.

Additionally, to the three-dimensional UMAP embedding of the image, we have applied the Mapper algorithm (Singh *et al.*, 2007) to capture the internal topology in the form of the network using the Python Kepler Mapper implementation (Van Veen *et al.*, 2019). Here we will provide only a short description of the algorithm, for the full account of the inner workings of the algorithm we refer the reader to the original publication (Singh *et al.*, 2007) and an introductory paper on topological data analysis (Carlsson, 2009). In general, construction of the Mapper networks often trails the following logic – first step, typically includes a calculation of low dimensional representation of original n-dimensional observations (by using neighborhood graphs or ordination methods). Further, reduced observations are often combined with other metrics that capture outliers, density and irregularities in the data (KDE etc.). The

objective variables (can be PCA or UMAP scores) are binned into a user-defined number of overlapping filter functions. The distribution of the data points within localized bins is further clustered using the clustering algorithm (specific choice is task-dependent (Pedregosa *et al.*, 2011)). The nodes that share same observations are connected with an edge and the output network is typically visualized using force-directed layouts. Further, the output network is colored according to the target variable, which provides the interpretation and hot-spot analysis. The transformation of the original UMAP scores to the network provides several additional benefits that can be useful in further downstream machine learning tasks. Depending on the chosen resolution, the Mapper networks are able to better capture the global topology and provide some degree of tolerance to the noise and flexibility to combine various metrics. Another beneficial side of the Mapper networks are in the fact that it is possible to visualize an arbitrary number of dimensions within the single complex network. Most useful applications are typically two to three dimensional, however, it is possible to visualize 4 to 5 dimensional datasets by using the 4 and 5-hypercubes as filter functions. Such networks can provide an additional level of detail or coarseness if needed. The challenge of using the high-dimensional filter functions is that they grow with power and typically form large graphs, which are not always convenient to work with.

The fs-LIMS, being a point-to-point chemical characterization method, provides the capability to perform depth profiles and volumetric estimates of any solid material. However, the ablation rate of investigated materials depending on the applied laser wavelength might vary, so that the characterization of very small inclusions (microfossils, micro- to nanograins) can be affected by the number of applied laser shots. For example, if a majority of the investigated material is ablated with the first laser shot, further averaging additional laser shots is counterproductive, because the target material is already removed from the sample. Using the assumption that some microfossils (and the mineralization related with them) might be of nm size in depth, we decided to recalculate the new embeddings for the averaged dataset (one layer data) and compare how the structure of the proximity networks will change. Figure 12 shows the spectral proximity network computed on the UMAP scores (using an increased number of epochs, 5 NN, and a cosine metric) on the 2-layer dataset (40,000 mass spectra, 260 single unit masses). The first three UMAP dimensions have been used as filter functions divided into 40 windows with 30 percent overlap and further clustered using the density-based clustering algorithm (DBSCAN) within the overlapping voxels (Ester *et al.*, 1996). The output undirected graph was further visualized in the open-source graph visualization platform Gephi (Bastian *et al.*, 2009) using the ForceAtlas2 layout algorithm (Jacomy *et al.*, 2014) and edges are colored according to degree of the nodes (blue – higher degree, yellow and red – smaller degree). The network in fig. 12 reveal structure of the cosine spectral proximities and indicate the complexity of that metric present in the dataset. The topology of the network shows disjoint clusters that consists of three main components: the quartz cluster shows a uniform radial structure, indicating that spectra from this cluster has less internal variability of intensity profiles. In contrast, the spectra from the cluster of microfossils indicate gradual change of the spectral proximity and thus, form more complex shape, which reflects the change in the volumetric ablation and ionization of the microfossils. And lastly, the cluster from inclusions - also shows complex internal topology, which indicates the complexity of the chemical compositions of these inclusions.

As mentioned in the methods section, additionally to the original volumetric dataset an averaged mass spectrometric image was calculated, which contains the 5

single laser shots per pixel, averaged for every given pixel forming a dataset that consists of 20'000 mass spectra (single layer). Figure 13 shows the spectral proximity network calculated from a single layer (fully averaged) mass spectrometric image, but visualized using the Yifan Hu graph layout algorithm (Hu, 2005) and calculated using the same hyperparameters. The change of the layout algorithm was motivated by artefactual visualization of microfossils cluster using the ForceAtlas2. Figure 13A shows the distribution of eigenvector centralities (Bonacich, 2007) calculated from the fully averaged mass spectrometric image. The network on the right (Figure 13B) shows the distribution of node degrees on the same network. As shown in fig. 13, the global structure of the similarity graphs remained very similar – two biggest components are easily separable. However, one can see that the transition line between inorganic and organic spectra (figs. 13A and 13B) appeared in the proximity structure after averaging more spectra (5 single laser shot spectra) and the fine structure of metallic inclusions was reduced to a single spike, which now shows the linkage to the transition line, and thus to the microfossils. An interpretation of this observation could be made in a following way - the pairwise spectral distances with an increased averaging are getting smaller, due to the volumetric domination of the silica matrix or microfossils (depending on the pixel location), thus, the spectra from different classes are becoming more fused into each other. It was also mentioned that the secondary metrics calculated from the spectral similarity networks might be of use in the downstream machine learning tasks (i.e., prediction of biogenicity of spectra from other Precambrian cherts). For example, Figure 13A shows three distinct clusters with easily identifiable eigenvector centrality measures, that can be further used as a predictive feature together with their UMAP scores. The eigenvector centralities have advantages over measures like degree (see Figure 12B), betweenness, and closeness centrality. The eigenvector centralities provide an easy visual assessment of data that are most representative of microfossils, or the host mineral. The topological structure of the graph itself also provides very important information, i.e., the eigenvalue spectrum of the normalized graph Laplacian describes the network's structure on a global level (de Lange *et al.*, 2014) just by using one metric, without referring to individual nodes or connections. For example, the characteristic "neck" of the transition structure from the host (quartz matrix) towards the microfossils (Figure 10A and 10B) has been observed in our previous work from the Gunflint microfossils using the space-type mass-spectrometer (Lukmanov et.al., 2021). Overall, the current datasets generated with the LMS-GT instrument can be used as a labeled library of spectral representations of truly biogenic Precambrian fossils, which can be of high utility, for example, in comparison with abiotic Fischer–Tropsch type process products or other carbonaceous matter of undefined origin commonly found in the Precambrian.
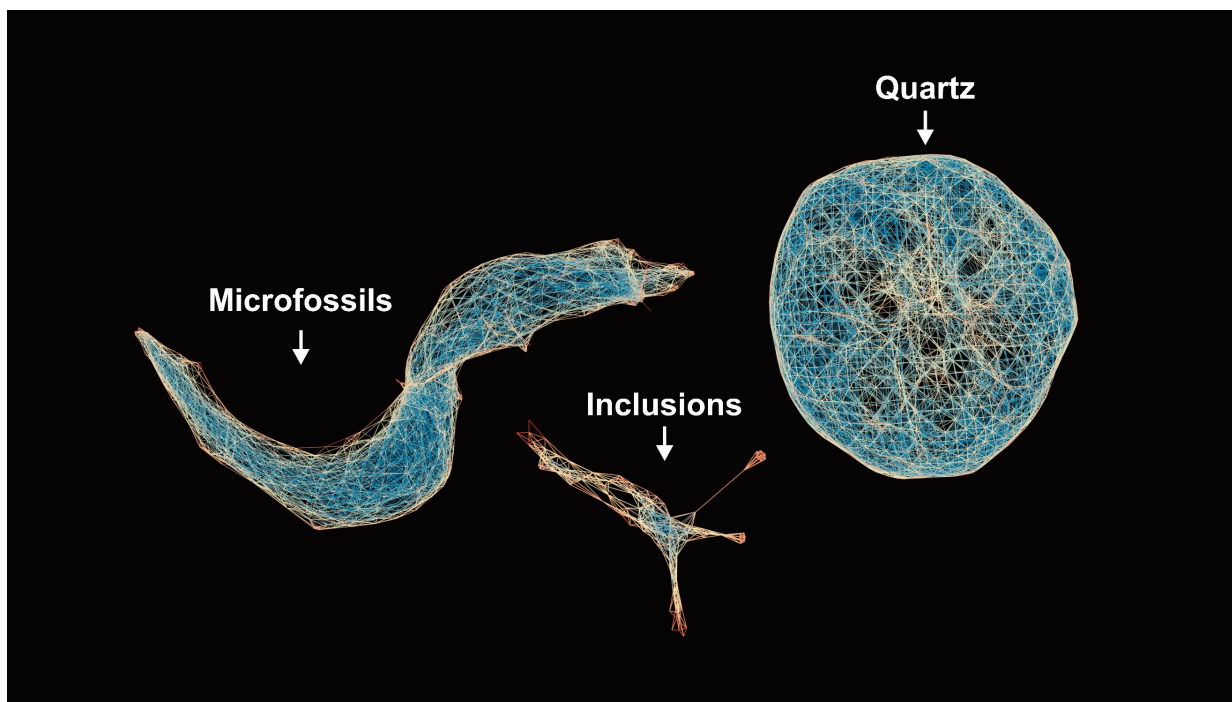
Fig.12. The spectral proximity network of the partially averaged mass spectrometric image (40'000 mass spectra – 260 single unit masses). The network colored according to the degree centrality of nodes in the network. Note, that the fine transition structure is not present and clusters are detached.
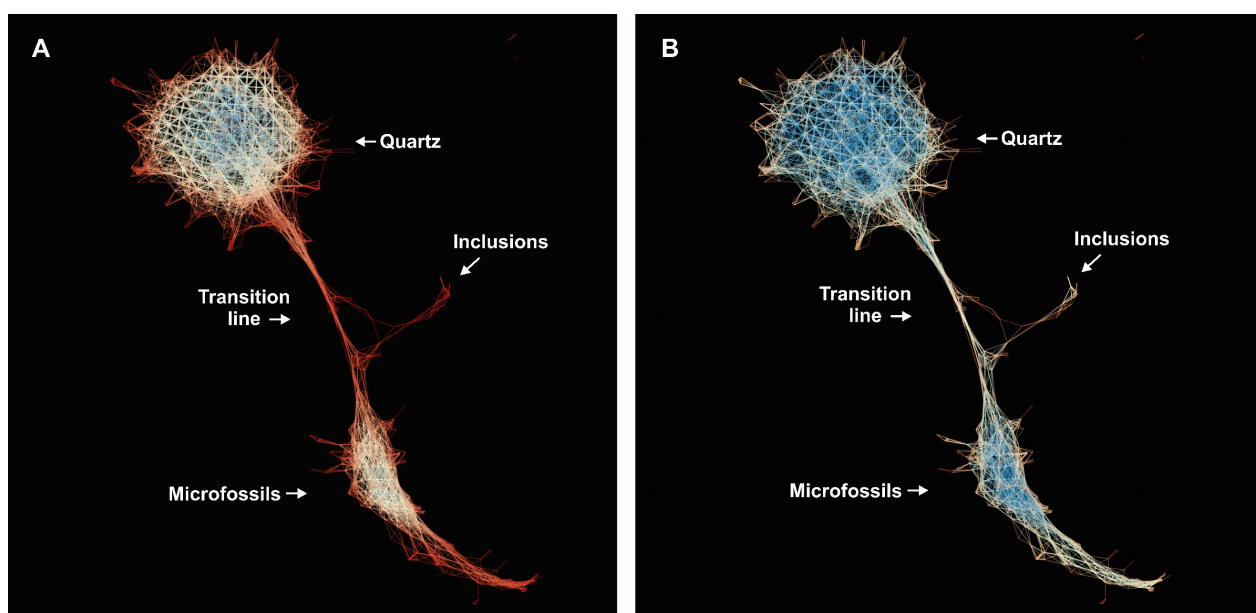

Fig. 13. The spectral proximity network of the fully averaged mass spectrometric image (20'000 mass spectra – 260 single unit masses). A) – The proximity network structure colored according to the eigenvector centrality of nodes in the network. The yellow and blue colors indicate more central nodes in the network. Fig.10B – The proximity network structure is colored according to the degree of nodes in the network. The blue colors indicate nodes with a higher degree. Note, that the fine transition structure gets thicker with the higher number of averaged laser shots. The inclusions cluster also lost its fine structure, which reflects the importance of good volumetric sampling.

Discussion

The chemical composition of the microfossils indicates that the majority of the microfossils were preserved as organic shells. The kerogenous matter which constitutes most of the organic remains was detected by measuring the abundant signal of ^{12}C, with smaller concentrations of N, O, P, S. Thus, the full range of biorelevant elements is present within the bodies of microfossils. An additional cosine similarity measure revealed that the microfossils form a distinct type of mass spectra. However, there are some challenges related to the elemental identifications: first of all, using the spectral decomposition and low dimensional embeddings, it was shown that metallic inclusions reveal a certain level of proximity to the microfossils. The most abundant group among inclusions identified as Fe-rich microfossils (see Figures 7 and 8) likely represent organically preserved cell walls with a set of intracellular Fe-rich minerals. The presence of greenalite nano-minerals recently was identified within specific morphospecies (Lepot *et al.*, 2017). However, Fe-mineralization could also be present by partially mineralized cell walls. Minerals such as greenalite can 'stick' to the exterior of cell walls. Our observations likely represent ionization of small portions of the microfossils, greenalite inclusions, and the surrounding silicate matrix. The abundant Fe, Si, and O signal observed in the spectrum (see Figure 8 – middle left panel) is supportive of that conclusion. Interestingly, high Fe concentrations within this group were also accompanied by relatively high Cr content. However, our observations show that the metallicity associated with microfossils is broader and reveals Cu, Ti, Al, and Ag mineralization. The range of identified metallicity indicates that the Gunflint populations experienced rather broad metallic hydrothermal load, which can be manifested as an increased toxicity during the Palaeoproterozoic. The presence of REE inclusions, mainly present with La and Pr, whereas other REE elements remained below the detection limit and thus where not detected. The analysis of the full spectral profile revealed an association with phosphates, which can be present in the shallow marine photic zone. By rapid enclosure in the newly formed silica gel, such inclusions can survive the dehydration stage of the chert formation and consequently end up being trapped in the chert.

The relatively widespread presence of Mg within the cluster of microfossils is supportive of the proposed hypothesis that photosynthetic organisms from the Gunflint chert are likely to contain trace concentrations of this element. Despite the absence of signatures of heavy alteration and secondary mineralization (this sample was collected from the Schreiber beach locality), Mg can be present as a small inclusions of magnesite or dolomite within the chert, which can be of inorganic origin (Lepot *et al.*, 2017). In general, biologically processed Mg is expected to be fractionated, moreover, different types of chlorophyll are expected to have different fractionation factors (Moynier & Fujii, 2017). Therefore, if Mg is indeed of biological origin in addition to the association to the organic matter it should display the expected fractionation factors. Further work will be needed to confirm the biogenic origin of Mg using isotope fractionation, nano-characterization of grains and their association with microfossils.

Here, we also have to report the caveats related to this work. The spectral profile of the quartz mineral shown in Figure 8 indicates that the total ion yield is a lot smaller in comparison to the spectra registered from the microfossils. This observation could be explained by the usage of the fs-IR laser. The clean quartz material is nearly transparent for the given wavelength and therefore more energy is required to be deposited on the surface of the sample. In contrast to quartz, microfossils are more absorptive of the given wavelength and can yield higher SNR even with smaller

energies. Unequal laser fluences required for the balanced ionization of both materials can be solved by using nonlinear beta barium borate crystals, that can multiply the output frequencies of the laser. Another potential pitfall concerns the surface quality of the sample. The orthogonal to the sample laser focusing implemented in our current setup provides a fixed position of the laser focus, thus changes in the topography of the sample can induce changes in the subsequent ionization. Such changes can affect the output image quality by forming an ion yield that is not related to the chemistry of the investigated sample.

The spectral decomposition used in this work covers only single unit masses, except for Mg and Cr lines that have been specifically chosen for the analysis. Thus, additional scripts are required for future campaigns to extract finer spectral information, resolving the majority of the isobaric interferences. The calculation of embeddings, as was mentioned before, was done by using the cosine metric, however, a great variety of other metrics are available and their impact on the aspects of spectral similarity needs to be assessed. Moreover, the number of neighbors (N) in the construction of the neighborhood graph has a profound effect on the structure of the embedding. By choosing a large N, one can approximate more of the global structure, or, alternatively, emphasize more of the local neighborhood by reducing N. In general, choosing the right N requires some trial and error, however, the reader has to keep in mind that provided embeddings are parameter dependent. Additionally, the proximity of nodes in the output embedding can have no physical meaning if there is no edge between them (e.g., if distances are not defined). Thus, the analytical assessment of the graph connectivity is helpful in the interpretation step of the UMAP embeddings. A similar situation affects the construction of the Mapper networks, where a variety of hyperparameters is present. For example – the number of filter function windows, percentage of overlap, hyperparameters of the clustering algorithms.

Conclusions

The conclusions of this work are summarized as follows. First, the chemical composition of Precambrian Gunflint chert identified with the LMS-GT instrument provides unambiguous identification of organically preserved microfossils. The volumetric images calculated for specific masses provide a clear view of the chemical diversity and spatial heterogeneity of the investigated area. Second, the relational data analysis applied to the large spectral dataset captures the complex structural details of spectral similarity. The low dimensional UMAP embeddings calculated from imaging cubes yielded 9 distinct clusters and a strong separation between the organic and inorganic spectra. The averaging dynamics of the imaging data provide an additional perspective on the preservation of the signal quality. Third, the co-occurrence of ^{12}C, ^{31}P, and ^{24}Mg provides an additional line of evidence towards the photosynthetic nature of investigated microfossils. Lastly, fs-laser ionization mass spectrometry combined with manifold learning techniques provides a powerful analytical framework and is capable of accelerating knowledge extraction and yielding new insights into chemical diversity of investigated materials.

Acknowledgments

References

Alleon J., Bernard S., Le Guillou C., Marin-Carbonne J., Pont S., Beyssac O., McKeegan K. D., and Robert F. (2017) Molecular preservation of 1.88 Ga Gunflint organic microfossils as a function of temperature and mineralogy. pp 16147-16147.

Awramik S. M., and Barghoorn E. S. (1977) The Gunflint microbiota. *Precambrian Research*, 5: 121-142.

Azov V. A., Mueller L., and Makarov A. A. (2020) Laser ionization mass spectrometry at 55: Quo Vadis? *Mass spectrometry reviews.*

Barghoorn E. S., and Tyler S. A. (1965) Microorganisms from the gunflint chert. *Science*, 147: 563-575.

Bastian M., Heymann S., and Jacomy M. (2009) Gephi: an open source software for exploring and manipulating networks.Proceedings of the International AAAI Conference on Web and Social Media.

Belkin M., and Niyogi P. (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15: 1373-1396.

Bonacich P. (2007) Some unique properties of eigenvector centrality. *Social networks*, 29: 555-564.

Brasier M. D., Green O. R., Jephcoat A. P., Kleppe A. K., Van Kranendonk M. J., Lindsay J. F., Steele A., and Grassineau N. V. (2002) Questioning the evidence for Earth's oldest fossils. *Nature*, 416: 76-81.

Brasier M. D., and Wacey D. (2012) Fossils and astrobiology: new protocols for cell evolution in deep time. *International Journal of Astrobiology*, 11: 217-228.

Campello R. J., Moulavi D., and Sander J. (2013) Density-based clustering based on hierarchical density estimates.Pacific-Asia conference on knowledge discovery and data mining. Springer.

Carlsson G. (2009) Topology and data. *Bulletin of the American Mathematical Society*, 46: 255-308.

Cloud P. E. (1965) Significance of the Gunflint (Precambrian) Microflora. *Science*, 148: 563-575.

Cui Y., Bhardwaj C., Milasinovic S., Carlson R. P., Gordon R. J., and Hanley L. (2013) Molecular imaging and depth profiling of biomaterials interfaces by femtosecond laser desorption postionization mass spectrometry. *ACS applied materials & interfaces*, 5: 9269-9275.

Cui Y., Moore J. F., Milasinovic S., Liu Y., Gordon R. J., and Hanley L. (2012) Depth profiling and imaging capabilities of an ultrashort pulse laser ablation time of flight mass spectrometer. *Review of Scientific Instruments*, 83: 093702.

de Lange S., de Reus M., and Van Den Heuvel M. (2014) The Laplacian spectrum of neural networks. *Frontiers in Computational Neuroscience*, 7.

Dong W., Moses C., and Li K. (2011) Efficient k-nearest neighbor graph construction for generic similarity measures.Proceedings of the 20th international conference on World wide web.

Ester M., Kriegel H.-P., Sander J., and Xu X. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise.Kdd.

Grimaudo V., Moreno-García P., Riedo A., López A. C., Tulej M., Wiesendanger R., Wurz P., and Broekmann P. (2019) Review—Laser Ablation Ionization Mass Spectrometry (LIMS) for Analysis of Electrodeposited Cu Interconnects. *Journal of The Electrochemical Society*, 166: D3190-D3199.

Grimaudo V., Tulej M., Riedo A., Lukmanov R., Ligterink N. F. W., de Koning C., and Wurz P. (2020) UV post-ionization laser ablation ionization mass spectrometry for improved nm-depth profiling resolution on Cr/Ni reference standard. *Rapid Communications in Mass Spectrometry*, 34: e8803.

House C. H., Schopf J. W., McKeegan K. D., Coath C. D., Harrison T. M., and Stetter K. O. (2000) Carbon isotopic composition of individual Precambrian microfossils. *Geology*, 28: 707-710.

Hu Y. (2005) Efficient, high-quality force-directed graph drawing. *Mathematica journal*, 10: 37-71.

Huang R., Yu Q., Li L., Lin Y., Hang W., He J., and Huang B. (2011) High irradiance laser ionization orthogonal time-of-flight mass spectrometry: A versatile tool for solid analysis. *Mass Spectrometry Reviews*, 30: 1256-1268.

Jacomy M., Venturini T., Heymann S., and Bastian M. (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9: e98679-e98679.

Kobak D., and Linderman G. C. (2019) UMAP does not preserve global structure any better than t-SNE when using the same initialization. *bioRxiv*: 2019.12.19.877522.

Lepot K., Addad A., Knoll A. H., Wang J., Troadec D., Béché A., and Javaux E. J. (2017) Iron minerals within specific microfossil morphospecies of the 1.88 Ga Gunflint Formation. *Nature communications*.

Liang Z., Zhang S., Li X., Wang T., Huang Y., Hang W., Yang Z., Li J., and Tian Z. (2017) Tip-enhanced ablation and ionization mass spectrometry for nanoscale chemical analysis. *Science Advances*, 3: 1059-1059.

Ligterink N. F. W., Grimaudo V., Moreno-García P., Lukmanov R., Tulej M., Leya I., Lindner R., Wurz P., Cockell C. S., Ehrenfreund P. and others. (2020) ORIGIN: a novel and compact Laser Desorption – Mass Spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces. *Scientific Reports*, 10: 9641.

McInnes L., Healy J., and Astels S. (2017) hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2: 205.

McInnes L., Healy J., and Melville J. (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Meyer S., Riedo A., Neuland M. B., Tulej M., and Wurz P. (2017) Fully automatic and precise data analysis developed for time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, 52: 580-590.

Moreno-García P., Grimaudo V., Riedo A., Tulej M., Wurz P., and Broekmann P. (2016) Towards matrix-free femtosecond-laser desorption mass spectrometry for in situ space research. *Rapid Communications in Mass Spectrometry*, 30: 1031-1036.

Moynier F., and Fujii T. (2017) Theoretical isotopic fractionation of magnesium between chlorophylls. *Scientific Reports*, 7: 6973.

Nolet C. J., Lafargue V., Raff E., Nanditale T., Oates T., Zedlewski J., and Patterson J. (2020) Bringing UMAP Closer to the Speed of Light with GPU Acceleration. *arXiv preprint arXiv:2008.00325*.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., and Dubourg V. (2011) Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12: 2825-2830.

Planavsky N., Rouxel O., Bekker A., Shapiro R., Fralick P., and Knudsen A. (2009) Iron-oxidizing microbial ecosystems thrived in late Paleoproterozoic redox-stratified oceans. *Earth and Planetary Science Letters*, 286: 230-242.

Riedo A., Meyer S., Heredia B., Neuland M. B., Bieler A., Tulej M., Leya I., Iakovleva M., Mezger K., and Wurz P. (2013a) Highly accurate isotope composition measurements by a miniature laser ablation mass spectrometer designed for in situ investigations on planetary surfaces. *Planetary and Space Science*, 87: 1-13.

Riedo A., Neuland M., Meyer S., Tulej M., and Wurz P. (2013b) Coupling of LMS with a fs-laser ablation ion source: Elemental and isotope composition measurements. *Journal of Analytical Atomic Spectrometry*, 28: 1256-1269.

Riedo A., Tulej M., Rohner U., and Wurz P. (2017) High-speed microstrip multi-anode multichannel plate detector system. *Review of Scientific Instruments*, 88: 045114-045114.

Schopf J. W., and Kudryavtsev A. B. (2012) Biogenicity of Earth's earliest fossils: A resolution of the controversy. pp 761-771.

Schopf J. W., Kudryavtsev A. B., Walter M. R., Van Kranendonk M. J., Williford K. H., Kozdon R., Valley J. W., Gallardo V. A., Espinoza C., and Flannery D. T. (2015) Sulfur-cycling fossil bacteria from the 1.8-Ga Duck Creek Formation provide promising evidence of evolution's null hypothesis. *Proceedings of the National Academy of Sciences*, 112: 2087-2092.

Singh G., Mémoli F., and Carlsson G. E. (2007) Topological methods for the analysis of high dimensional data sets and 3d object recognition. *SPBG*, 91: 100.

Tulej M., Ligterink N. F., de Koning C., Grimaudo V., Lukmanov R., Keresztes Schmidt P., Riedo A., and Wurz P. (2021a) Current Progress in Femtosecond Laser Ablation/Ionisation Time-of-Flight Mass Spectrometry. *Applied Sciences*, 11: 2562.

Tulej M., Lukmanov R., Grimaudo V., Riedo A., de Koning C., Ligterink N. F. W., Neubeck A., Ivarsson M., McMahon S., and Wurz P. (2021b) Determination of the microscopic mineralogy of inclusion in an amygdaloidal pillow basalt by fs-LIMS. *Journal of Analytical Atomic Spectrometry*, 36: 80-91.

Van Veen H. J., Saul N., Eargle D., and Mangham S. W. (2019) Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *Journal of Open Source Software*, 4: 1315.

Wacey D., McLoughlin N., Kilburn M. R., Saunders M., Cliff J. B., Kong C., Barley M. E., and Brasier M. D. (2013) Nanoscale analysis of pyritized microfossils reveals differential heterotrophic consumption in the ~1.9-Ga Gunflint chert. *PNAS*, 110: 1053-1060.

Wacey D., Menon S., Green L., Gerstmann D., Kong C., McLoughlin N., Saunders M., and Brasier M. (2012) Taphonomy of very ancient microfossils from the ~3400Ma Strelley Pool Formation and ~1900Ma Gunflint Formation: New insights using a focused ion beam. *Precambrian Research*, 220-221: 234-250.

Wacey D., Saunders M., Kong C., Brasier A., and Brasier M. (2016a) 3.46 Ga Apex chert 'microfossils' reinterpreted as mineral artefacts produced during phyllosilicate exfoliation. *Gondwana Research*, 36: 296-313.

Wacey D., Saunders M., Kong C., and Kilburn M. R. (2016b) A new occurrence of ambient inclusion trails from the ~1900-million-year-old Gunflint Formation, Ontario: nanocharacterization and testing of potential formation mechanisms. *Geobiology*, 14: 440-456.

Wiesendanger R., Grimaudo V., Tulej M., Riedo A., Lukmanov R., Ligterink N., Fausch R., Shea H., and Wurz P. (2019) The LMS-GT instrument–a new perspective for quantification with the LIMS-TOF measurement technique. *Journal of analytical atomic spectrometry*, 34: 2061-2073.

Wiesendanger R., Wacey D., Tulej M., Neubeck A., Ivarsson M., Grimaudo V., Moreno-García P., Cedeño-López A., Riedo A., Wurz P. and others. (2018) Chemical and Optical Identification of Micrometer-Sized 1.9 Billion-Year-Old Fossils by Combining a Miniature Laser Ablation Ionization Mass Spectrometry System with an Optical Microscope. *Astrobiology*, 18: 1071-1080.

Wurz P., Riedo A., Tulej M., Grimaudo V., and Thomas N. (2020) Investigation of the Surface Composition by Laser Ablation/Ionisation Mass Spectrometry. *LPI Contributions*, 2241: 5061.

3.3   Double pulse fs-LIMS

3.3.1. Improved plasma stoichiometry recorded by LIMS using a double-pulse femtosecond laser ablation ion source - (Rapid Communications in Mass Spectrometry)

Andreas Riedo[1], Rustam Lukmanov[1]. Valentine Grimaudo[1], Coenraad de Koning[1], Niels F. W. Ligterink[1], Marek Tulej[1], Peter Wurz[1];

[1]Space Research & Planetary Sciences (WP), Physics Institute, University of Bern, Bern, Switzerland;

RESEARCH ARTICLE

Improved plasma stoichiometry recorded by laser ablation ionization mass spectrometry using a double-pulse femtosecond laser ablation ion source

Andreas Riedo ⊕ | Rustam Lukmanov ⊕ | Valentine Grimaudo ⊕ |
Coenraad de Koning ⊕ | Niels F. W. Ligterink ⊕ | Marek Tulej ⊕ | Peter Wurz ⊕

Physics Institute, Space Research and
Planetary Sciences, University of Bern,
Sidlerstrasse 5, Bern, 3012, Switzerland

Correspondence
A. Riedo, Physics Institute, Space Research and
Planetary Sciences, University of Bern,
Sidlerstrasse 5, 3012 Bern, Switzerland.
Email: andreas.riedo@space.unibe.ch

Funding information
Swiss National Science Foundation; Swiss
National Science Foundation, Grant/Award
Number: 193453

Rationale: Femtosecond (fs) laser ablation ion sources have allowed for improved measurement capabilities and figures of merit of laser ablation based spectroscopic and mass spectrometric measurement techniques. However, in comparison to longer pulse laser systems, the ablation plume from fs lasers is observed to be colder, which favors the formation of polyatomic species. Such species can limit the analytical capabilities of a system due to isobaric interferences. In this contribution, a double-pulse femtosecond (DP-fs) laser ablation ion source is coupled to our miniature Laser Ablation Ionization Mass Spectrometry (LIMS) system and its impact on the recorded stoichiometry of the generated plasma is analyzed in detail.

Methods: A DP-fs laser ablation ion source (temporal delays of $+300$ to -300 ps between pulses) is connected to our miniature LIMS system. The first pulse is used for material removal from the sample surface and the second for post-ionization of the ablation plume. To characterize the performance, parametric double- and single-pulse studies (temporal delays, variation of the pulse energy, voltage applied on detector system) were conducted on three different NIST SRM alloy samples (SRM 661, 664 and 665).

Results: At optimal instrument settings for both the double-pulse laser ablation ion source and the detector voltage, relative sensitivity coefficients were observed to be closer (factor of ~2) to 1 compared with single-pulse measurements. Furthermore, the optimized settings worked for all three samples, meaning no further optimization was necessary when changing to another alloy sample material during this study.

Conclusions: The application of a double-pulse femtosecond laser ablation ion source resulted in the recording of improved stoichiometry of the generated plasma using our LIMS measurement technique. This is of great importance for the quantitative chemical analysis of more complex solid materials, e.g., geological samples or metal alloys, especially when aiming for standard-free quantification procedures for the determination of the chemical composition.

1 | INTRODUCTION

Sensitive measurement techniques using a pulsed laser to probe the chemical composition of solids, e.g. Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS), Laser-Induced Breakdown Spectroscopy (LIBS), or Laser Ablation Ionization Mass Spectrometry (LIMS), are applied in various fields in academic and industrial research to quantitatively investigate the chemical composition of solid samples. Applications range from, e.g., chemical imaging of highly heterogeneous geological and/or biological samples,[1–9] to the dating of solids and mineralogical analysis,[10,11] and to the chemical analysis of high-performance materials used in industry.[12–16] LA-ICP-MS is among the most widespread laser-based analytical technique due to its accurate quantification, high reproducibility, and low limits of detection. Today, various hybrid systems commercially exist that integrate both LA-ICP-MS and LIBS instrumentation to extend the measurement capabilities for the chemical analysis of solids.[17–20]

The application of a pulsed laser system to probe the analyte of interest has many advantages, as structures down to tens of nanometers can be targeted, layered materials can be chemically profiled by consecutive laser pulses, and sample contamination via handling can be minimized as no mechanical or chemical contact is required during preparation, among others.[9,21,22] With the replacement of, e.g., nanosecond or picosecond laser systems with powerful and stable ultra-fast pulsed femtosecond laser systems, the analytical capabilities and figures of merit of all laser-based measurement techniques have improved significantly, including, e.g., minimized matrix effects and element fractionation effects due to the absence of laser plasma plume interaction, improvement of spatial resolution due to reduced heat dissipation, increased reproducibility of measurements, enhanced ionization and stoichiometry, among others.[23–27] The application of femtosecond laser systems, however, shows drawbacks as well, at least for LIMS and LIBS systems. The resulting ablation plasma produced by such systems is colder and the lifetime is shorter than those induced by longer pulsed laser systems.[28] The colder plasma temperature favors the formation of polyatomic species in the ablation plume[28] and, depending on the analytical tool used for chemical quantification, the analytical capabilities may be limited if the mass resolving power is not high enough to discriminate against these isobaric interferences. For LIBS, the shorter lifetime of the plasma decreases the signal-to-noise ratio of detected signals and therefore limits the detection sensitivity.[29]

In this contribution, we quantify the analytical performance of a collinear double-pulse (DP) femtosecond (fs) laser ablation ion source that is coupled with our LIMS system.[30] The measurement will show that, at optimal instrument settings, in particular of the DP laser ablation/ionization source and the detector, the relative sensitivity coefficients (RSC), required for quantification of the chemical composition of solids, can be improved significantly; they scatter closer to 1. The application of a DP-fs laser ion source was introduced previously in LIBS and LA-ICP-MS instrumentation (see, e.g.,[29,31–34]). In LIBS, post-heating of the plasma by a second laser pulse allows for,

e.g., a longer-lived plasma and, as a result, improved detection limits. In LA-ICP-MS, the DP measurement scheme allowed, e.g., for the generation of a more fine-grained aerosol, which is highly favorable for the ICP. Recently, we demonstrated by studies on a Mg sample that the application of the DP mode allowed an increase in ion yield relative to the single-pulse (SP) mode by a factor of up to about 30.[30] In the current study, we apply longer than typically applied delays between the two laser pulses and analyze monoatomic ions produced in such a configuration. To the best of our knowledge, this is the first contribution reporting about the impact of a DP-fs system on the quantitative performance of a LIMS system with the second pulse used for post-ionization of atoms.

2 | EXPERIMENTAL

2.1 | Sample materials

For performance characterization, measurements on three different National Institute of Standards and Technology (NIST) standard reference material (SRM) alloys were conducted: SRM 661, 664, and 665. Prior to measurements, the top surface was removed mechanically with a diamond file and cleaned using an argon ion sputter gun (ion energy of about 3 keV, a 10 degree incident angle, and sputtering time of about 20 min).[23] The samples were selected for this study because many previous studies were conducted on these samples using our miniature LIMS system to which different laser systems were connected, from nanosecond[35,36] laser systems (operated at IR and UV wavelengths) to femtosecond laser systems.[23]

2.2 | Laser ablation ionization mass spectrometer

The measurement capabilities and principles of operation of the mass spectrometric system used for the performance characterization of the DP ablation ion source are described in detail in various scientific contributions.[3,23,36–38] Therefore, only a brief description of the measurement principles of the system, based on SP laser ablation, is given in the following.

The system used for recording mass spectra is a miniature (mass analyzer with $160 \times \varnothing\ 60$ mm) reflectron-type time-of-flight (TOF) mass spectrometer that was originally designed for the *in situ* analysis of the chemical composition of solids on planetary surfaces. The mass spectrometer is located within a vacuum chamber and a beam guiding system is used to guide the laser pulses towards the entrance window of the chamber. The laser pulses are focused by an optical lens system positioned directly above the mass analyzer, through the system, and towards the sample surface, which is positioned roughly 1 mm below the entrance ion optics of the mass spectrometer. In the current design, only positively charged species can enter the ion optical system. At the entrance of the system the ions are accelerated, focused, and confined into the field-free drift path of the mass analyzer. At the ion mirror, the ions are reflected towards the

multichannel (MCP) detector system[37] by passing the field-free drift path for a second time. The ions are separated in the field-free drift paths according to their mass-to-charge ratio, following the TOF measurement principle. The quadratic equation $m(t) = k_0(t - t_0)^2$ is used for the conversion of a TOF to a mass spectrum, where k_0 and t_0 are fitting constants. Note that the mass resolving power of this miniature mass analyzer allows for discrimination of different isotopes, but is not sufficient to resolve isobaric interferences, e.g., doubly charged ^{56}Fe cannot be distinguished from singly charged ^{28}Si. In this study, only two out of four anodes of the MCP detector system were used for the characterization of the DP ion source. A high negative voltage at the front, and a less negative voltage at the back of the MCP stack (chevron configuration), are applied to guide the released electrons towards the anodes. The back voltage was kept during the measurements at −300 V, while the front was varied in the measurements. The applied detector voltage mentioned in the following refers to the high negative one. Two high-speed ADC cards (each with two channels, channels were combined on each card resulting in a doubled sampling rate of 3.2 GS/s, 12-bit vertical resolution, with onboard processing capabilities) were used for the recording of TOF spectra. For characterization of the recorded mass spectra, in-house written software is used, which allows for, e.g., TOF to mass scale conversion, peak integration, and the calculation of RSC values.[39] In this study, we consider certified elements in the SRM samples until Fe for calculation of RSC values. Iron is the most abundant species in these samples and gain-loss was observed which affects the derived RSC values. Similar to our earlier studies, the RSC value of a certain element is calculated via measured abundance divided by the NIST certified abundance. The samples are positioned below the mass spectrometer using a three-dimensional (3D) micro-translation stage with micrometer positioning accuracy. The complete instrument setup, including laser system, acquisition card, sample stage and delay stage (discussed in the following), are fully controlled remotely by in-house written software.

The fs laser system used in this study is operated at a wavelength of $\lambda = 775$ nm and a laser pulse repetition rate of 1 kHz, and produces pulse widths of $\tau \sim 190$ fs. The mechanical realization of the collinear DP system was described in detail in Tulej et al,[30] and is described only briefly in the following. The fundamental laser beam (p-polarization) is first transformed to an s-polarized beam via a half-wave plate (denoted as WP in Figure 1) installed directly at the output of the laser system before being split into two beams. An s-polarization sensitive 50/50 beam splitter (denoted as BS1) was used for beam splitting. The pulses from one beam are guided to a static beam guiding system (defined as P1) that is composed of a static retroreflector system and a polarization-sensitive attenuator (denoted as A). In comparison to the first realization of the DP system described in Tulej et al,[30] this static retroreflector was added to the beam guiding system because it enables correction for small deviations between the travelling paths between both laser beams. Further, it allows the same beam characteristics, such as the polarization of both beams. It is known that the integration of retroreflector systems affects the polarization of the incoming beam;
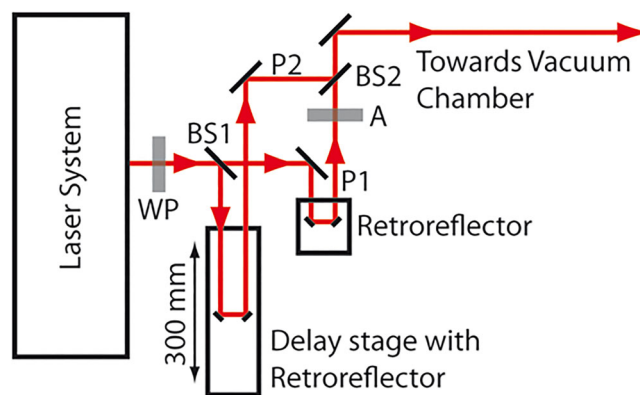

FIGURE 1 Schematics of the collinear DP-fs laser ablation ion source. The integrated retroreflector positioned on a remotely controlled linear stage allows for variation of the delay between both laser pulses with delay increments of ~300 fs. Further information is given in the text

see, e.g., the technical note from Thorlabs regarding the integration of a prism retroreflector or theoretical calculations.[40,41] Pulses from the second beam are guided to a dynamic beam guiding system (denoted as P2). The latter system is composed of a retroreflector that is positioned on a remotely controlled linear stage with a traveling distance of 300 mm. The dynamic guiding system is installed in such way that delays relative to P1 of +300 to −300 ps can be realized with time increments of ~300 fs. The energy of both laser pulses can be changed by either adjusting the output power of the main laser system or using the attenuator installed in the static beam guiding system. A second beam splitter (BS2) is used finally to recombine the beams of both beam-guiding systems. From that position, both pulses follow the same beam path towards the mass spectrometer.

2.3 | Measurement procedure

For the quantification of the effect of the DP ion source on the analytical performance of the system, multi-parametric studies are required (e.g., applied pulse energy, the delay between P1 and P2, etc.). First, a laser irradiance campaign (16 different pulse energies in the range of ~400 nJ to 7.48 μJ were applied) in SP mode was conducted on SRM 664. Each measurement of different pulse energy was performed on a new and unused sample position. A total of 100,000 laser shots were applied on each sample position for each pulse energy, resulting in 100,000 spectra, which were stored on the host computer in packages of 1000 histogrammed spectra, resulting in 100 saved TOF spectra. This campaign allowed to define, e.g., the laser ablation threshold, the maximum pulse energy before observing space charge effects, and, most importantly, the pulse energy conditions where stable ablation can be observed. Note that SP measurements were conducted using the static beam guiding system (P1, see Figure 1) because of the integrated beam attenuator that allows easy adjustments of the laser pulse energy. Subsequently,

dense raster campaigns were applied on NIST 664 to investigate the impact of the delay between P1 and P2 on the plasma stoichiometry. The raster campaign consisted of 2000 sample positions with a pitch of 30 μm and was repeated for three different pulse energies (1.8 μJ, 2.1 μJ, and 2.7 μJ) with both pulses set at the same energy. Each sample position in the raster campaign involved a different delay between P1 and P2 with a delay increment of ∼300 fs. For each delay, four files of 4000 spectra were recorded and saved on the host computer. These raster campaigns were required to define, e.g., the zero delay between both pulses from P1 and P2, the delay where the hottest plasma conditions can be observed (reduction of polyatomic species), or the regime where mostly post-ionization of the generated plasma occurs. With the knowledge of these regimes, measurement campaigns conducted with the single- and double-pulse scheme (where the total applied pulse energy in both campaigns was the same) were conducted on fresh and unused sample positions. At optimal measurement conditions (laser and detector settings), DP measurements were finally extended to SRM 661 and 665 to assess the broader applicability of the found measurement parameters.

3 | RESULTS AND DISCUSSION

Through the SP laser irradiance campaign, the laser ablation threshold (close to 1 μJ), as well as the regime where a stable ablation is obtained, was identified. Stable and continuous ablation occurred at about 1.8 μJ, whereas at lower pulse energies, only major abundant elements were identified in the recorded and histogrammed spectra (data not shown). The information of, e.g., the ablation threshold is of importance for the application of the DP ion source. DP studies near but above the ablation threshold energy (considering first laser pulse applied) have several advantages, including, e.g., reduced surface and space charge effects. These effects impact the spectral quality; peak distortion and spectral jitter are reduced, which in turn improves the mass resolution. Moreover, for sufficiently low plasma densities, less laser radiation shielding effects occur and the second laser pulse can penetrate through the plasma plume more efficiently.

In Figure 2, the recorded signals (integrated peak intensity) of $^{56}Fe^+$ (right panels) and $^{56}Fe_2^+$ (left panels) of two delay campaigns conducted at a total applied pulse energy of 2.1 μJ and 2.7 μJ are presented (the detector was operated at voltages of −1850 V and −1900 V for the 2.7 μJ and 2.1 μJ campaign, respectively). Note that each data point corresponds to the mean of the recorded signal of the four histogrammed TOF spectra (each file is a histogram of 4000 spectra). A moving average filter (window size of three data points) was applied for better visualization of the recorded data. Iron is the major abundant element of NIST 664 and, hence, the positively charged dimer $^{56}Fe_2^+$ can be formed and observed readily in the plume. The delay campaigns show three different regimes; R1 where a zero delay between both pules is achieved, R2 where conditions with the hottest plasma are reached, and R3 where mostly post-ionization takes place (see notations in the top left panel). A similar signal trend was reported in an earlier publication where measurements were conducted in DP mode on a Mg sample.[30] At a zero delay (indicated with a dashed line), which is virtually identical to the situation where one single laser pulse with double energy is applied (see discussion below for Figure 3), a small signal enhancement can be identified (see insert in the bottom left panel). Note that the raster campaigns were conducted with a delay accuracy of ∼300 fs for the identification of the zero delay. A slight imperfection at zero delay between both laser pulses may result in a slightly reduced signal due to potential interference effects. Right and left from this localized signal enhancement a smooth drop is recognizable, which is then followed by a large increase in the signal,
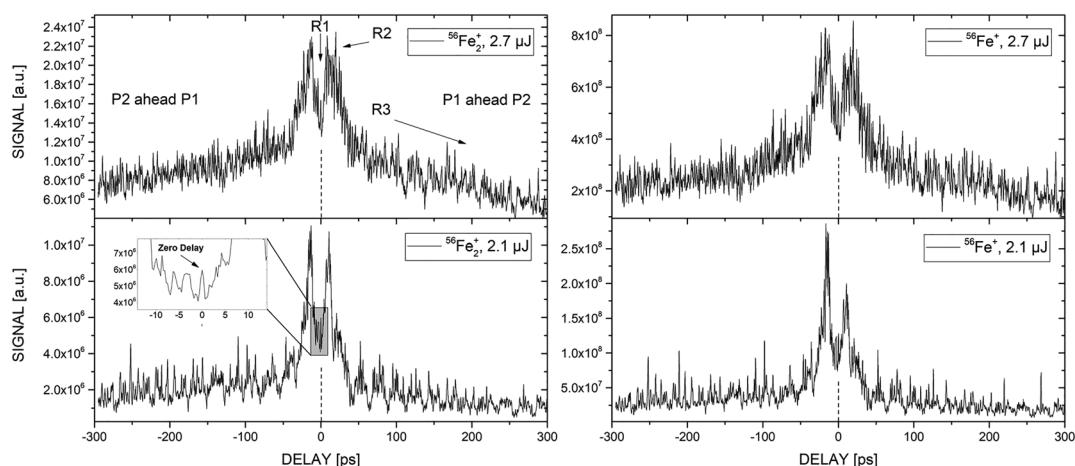

FIGURE 2 Two pulse delay campaigns conducted at two different total pulse energies are shown (2.1 μJ and 2.7 μJ). The time shift between both pulses was probed between −300 and +300 ps. In total, 2000 delays were investigated, with a delay increase of ∼300 fs between each step. For both campaigns, the recorded signal of $^{56}Fe_2^+$ (left panels) and of $^{56}Fe^+$ (right panels) are plotted. Three regimes can be identified (see top left panel), with R1 corresponding to the regime of zero delay, R2 to the hottest plasma condition, and R3 to the regime where mostly post-ionization of the neutral atoms in the ablated plume takes place
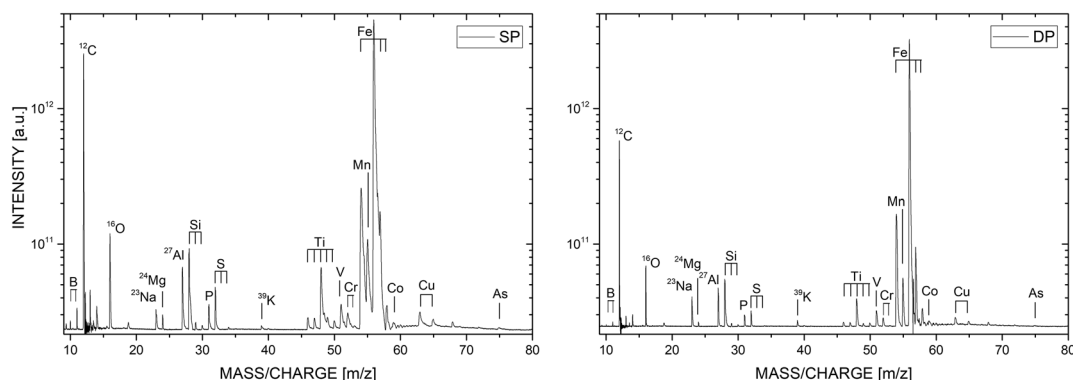
FIGURE 3 Mass spectra of NIST 664 recorded in SP mode (left) and DP mode (right), both recorded at the same total applied pulse energy of 3.5 µJ. The delay in the DP mode was zero (R1 regime in Figure 2). In comparison to the SP measurement, slightly lower intensities are observed for the DP measurement. Note that the scaling of both panels is identical

forming the regime of hottest plasma conditions at a delay of about 25 ps, comparable with our previous study conducted on a Mg sample.[30] At these conditions, neutral species that are produced during the ablation process are efficiently post-ionized, resulting in a significant increase in the overall recorded signal. Note that the increase in recorded signal for atomic species is higher than the production of, e.g., positively charged dimers at hottest plasma conditions, leading to a net reduction in such species contribution, as also determined by Tulej et al.[30,42] For the measurement conducted at a pulse energy of 2.7 µJ (see Figure 2), a net gain of positively charged ^{56}Fe relative to the dimer $^{56}Fe_2$ of ~10% was observed. Note that the effect of DP post-ionization is more efficient at lower applied pulse energies. This effect is expected because the post-ionization is more efficient having a lower density plasma produced.[30] Note that, before starting these delay campaigns, special attention was given to achieve equal pulse energy and comparable ablation profiles for both pulses. This explains why the signal evolution in Figure 2 is almost symmetrical around zero delay. After passing the hot plasma conditions, a steep drop in recorded signal is observed, which is comparable with our earlier study conducted on a Mg sample (see Tulej et al, fig. 4).[30]

For verification, i.e., if the DP mode with zero delay between the two pulses produces a comparable signal to SP mode, three measurements of both modes working at the same total energy of 3.5 µJ and detector voltage of −1850 V were conducted (see Figure 3). Note that, for this study, as shown in Figure 3, the pulse energy of 3.5 µJ was selected, because from SP measurements it is known that a pulse energy of 3.5 µJ allowed for stable and continuous material ablation (ablation threshold was observed at ~1 µJ). The shown measurements are an accumulation of 100,000 single mass spectra (100 files, each corresponding to a histogrammed spectrum of 1000 single spectra). Note that the y-axis scaling is identical for both panels and that the recorded signal is displayed in log-scale. In comparison to the SP measurements, a slightly lower recorded signal can be observed for the DP measurements at zero delay. This effect can be attributed to the accuracy of ~300 fs for the identification of the zero delay between both laser pulses. As discussed in the

previous paragraph, a slightly imperfect overlap between both pulses results immediately in a reduced signal. Interestingly, small differences between the conducted SP and DP ablation measurements were observed. In comparison to the SP measurements, in all three measurements conducted in DP mode lower intensities for Ti and S were observed. The SP measurements are in line with previous measurements; typically, a much higher ion yield of Ti relative to other monoatomic ions was observed. In the study of Neuland et al,[43] measurements were conducted on four geological standard materials and Ti showed consistently an elevated RSC value at the level of about 10 (fig. 5 in Neuland et al[43]). In the DP measurement, the measured signal of Ti is lower, which is closer to the actual stoichiometry of the ablated material. Interference effects in close vicinity of a perfect zero delay might be a possible source for this observation – the electrons can be heated for a slightly longer time which might improve the ionization of the ablated material. In addition to this observation, a better resolved mass spectrum is observed for the DP measurement, which may be accounted to, e.g., reduced surface charge effects. This observation cannot be generalized for all DP measurements. For instance the DP measurement displayed and discussed in the following Figure 4 shows a comparable mass resolution to the SP measurement (see mass-to-charge region of the Fe isotopes). Note that the ion optical system of the employed TOF mass spectrometer can typically handle low intensity plasmas better than high intensity ones.

A similar observation of reduced ion yield of Ti was found in DP measurements conducted at hottest plasma conditions (see regime R2 in Figure 2). The left panel of Figure 4 shows a section of the simulated mass spectrum of NIST 664 around Ti, whereas, in the middle and right panels, the actual SP and DP measurements are shown. Again, both the SP and DP measurement campaigns were conducted at almost identical measurement settings. The applied pulse energy was ~3.4 µJ and ~3.5 µJ for the DP and SP measurement campaign, respectively. Both measurements were performed with the same detector voltage of −1850 V. The shown parts of the SP and DP spectra correspond to histogrammed spectra of 30,000 spectra each. All three spectra are normalized to $^{54}Fe^+$. The
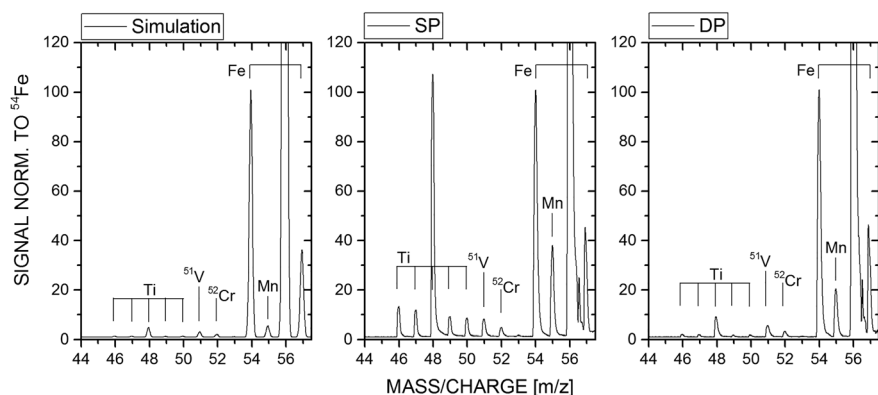
FIGURE 4 Left: Simulated mass spectrum of NIST 664 around Ti; middle: single-pulse (SP) measurement; right: DP measurement at hottest plasma conditions (R2, see Figure 2). For comparison, the simulated and recorded signals were normalized to $^{54}Fe^+$. The total applied pulse energy was almost identical in the SP ($\sim$3.5 μJ) and DP ($\sim$3.4 μJ) measurement mode
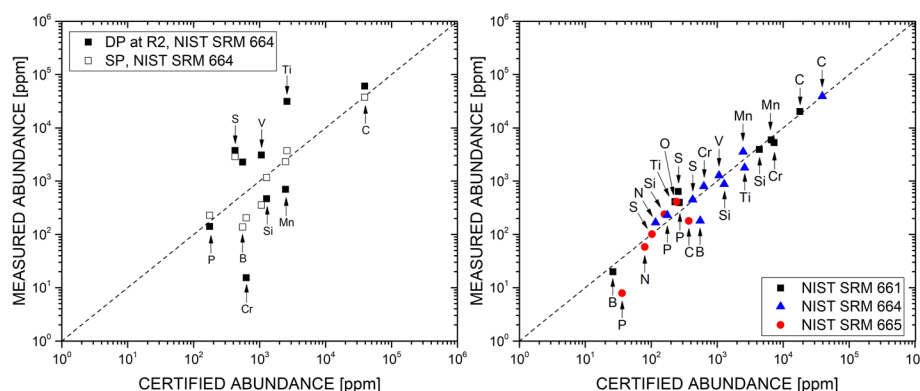

FIGURE 5 Left: RSC values calculated for the case of DP measurements at hottest plasma conditions (R2) and SP measurements at the same total pulse energy. Right: RSC values derived for best instrument settings in the DP mode, i.e., pulse energy, pulse delay, and detector gain. In comparison with the RSC values displayed on the left side, the RSC factors scatter much closer around 1 in a fully optimized system operated in DP mode at the optimal delay

simulated spectrum (all elements have RSC = 1) was generated using our in-house designed analysis software that provides the option to simulate spectra with, e.g., user-defined noise, observed mass resolution, or element/isotope abundances of a reference sample.[39] It can be clearly asserted that the Ti abundances recorded in the DP measurements match much better the known composition as shown in the simulated spectrum, while the relative recorded intensity of Ti in SP mode is significantly too high. This indicates that post-heating and post-ionization of the generated plasma with the second pulse improves the quantitative measurement at these specific instrument settings and pulse delay (hottest plasma condition).

Optimal measurement settings in view of element RSC could be derived by tuning (i) the delay of the second pulse in DP mode, (ii) total pulse energy, and (iii) applied detector voltage (an increase of ion signal, e.g., 30 times, can lead to ion detector saturation effects). Concerning the pulse delay, optimal measurement conditions were identified for measurements conducted in regime R3 (delay of $\sim$300 ps), with a total applied pulse energy of $\sim$ 5.8 μJ. Furthermore, the optimization included the reduction of the applied voltage at the MCP stack detector. A higher total applied voltage results in higher gain but also increases the probability to induce gain losses. Optimal detector settings were found at moderate amplification conditions to avoid such saturation effects and gain losses. A total potential difference of about 1450 V was applied over the MCP stack, which represents a compromise between signal amplification[37] and detector saturation and gain loss. Nevertheless, a sensitivity drop was observed at the arrival of ^{54}Fe ions, which is the second-highest abundant isotope of Fe, with iron being the most abundant element in

all three investigated samples. To stay away from the detector limitations, we limited the analysis of RSC values for certified elements arriving earlier than ^{56}Fe, more precisely until ^{55}Mn.

In Figure 5, the RSC values of three different measurement conditions are shown. Note that all the measurements were conducted at the same detector voltage and applied total pulse energy of $\sim$5.8 μJ. The left panel contains RSC values for NIST 664 derived from SP and DP measurements (R2, hottest plasma conditions). The right-hand panel presents the RSC values derived from the three NIST SRMs 661, 664, and 665 samples at the best instrument settings (delay $\sim$300 ps, pulse energy of $\sim$5.8 μJ, and applied detector voltage of $\sim$ −1750 V). Note that the detector voltage and the total applied pulse energy were kept the same for the individual measurements inside the same panel.

The RSC values for SP measurements (Figure 5, left panel) scatter around the trend line of 1 (dashed line), which is comparable to previous measurements. Surprisingly, the RSC values for the DP measurements conducted at the hottest plasma regime scatter slightly more than the SP measurements. A possible explanation could be that the produced number of ions cannot be accounted for correctly by the ion optical and detector system of our miniature LIMS system. The situation is very different for measurements conducted at optimal conditions (pulse delay and energy, and detector voltage). As can be observed in the right panel of Figure 5, the scatter of the RSC values is much less, values are closer to 1. Note that Al was not considered for calculation of RSC values because of potential isobaric interferences with doubly charged ^{54}Fe. Note that for measurements conducted on SRM 665, the RSC values of N and O are shown as well

and match very well with the trend line representing RSC = 1. These are the only two elements shown for which NIST provides a non-certified estimate abundance based on a single instrument technique, rather than a certified abundance as for all other shown elements. In addition to the graphical representation of the RSC values derived for the measurements show in Figure 5, the improvement can be quantified using the following two equations:

$$a = \frac{1}{n}\sum_{i}^{n}|\log(RSC_i)|; \text{ and } b = (1-a)*100;$$

where RSC_i denotes the RSC value of element i and n the number of RSC values available for the quantification. The absolute value of the logarithmic value of the RSC is required because in this way both RSC values of 0.1 and 10 have the same weight for the quantification of the sum (both are a factor of 10 off). The closer a is to 0, and b is to 100, the better the quantification of the elements by the applied instrument setting. Note that for this calculation, only the RSC values derived for SRM 664 are considered (SP measurement, DP measurement conducted in hottest plasma conditions, and DP at optimal conditions, see Figure 5). For the SP measurement, we derive for $a = 0.31$ and $b = 69$, for the DP hottest plasma conditions $a = 0.67$, $b = 33$, and for DP at optimal conditions $a = 0.14$ and $b = 86$. This quantification shows that the DP measurements conducted at optimal conditions are by a factor of about 2 closer to the optimal situation (14% off for DP measurements at best settings, and 31% for the SP measurements). As mentioned before from visual inspection of Figure 5, the quantification shows as well that the DP measurements conducted at hottest plasma conditions were less optimal than the SP measurements. Tulej et al[30] showed that the post-ionization at hot plasma conditions is very effective. However, in this study, such conditions might lead to a too dense plasma of positively charged species, which may exceed the capabilities of the ion optical system of our miniature TOF-MS system, which in turn would lead to limited resolution in the TOF spectrum, limiting the accuracy of, e.g., peak integration.

At the optimal conditions, lower signal intensity was typically observed as well, which may reduce the detection sensitivity of the applied instrument. However, such conditions may still be of high importance when a matrix-matched standard is not available for the analysis of an unknown sample. Important to note is that the larger RSC scatter (e.g., left panel of Figure 5) is not a limiting factor for the mass spectrometric quantification of solids using LIMS. In case a matrix-matched standard is available, the RSC values derived from the standard can be applied to an unknown sample, even if the applied measurement conditions lead to a larger variation in RSC values between different elements.

4 | CONCLUSIONS

Previous studies have shown that the coupling of femtosecond (fs) laser systems to the LIMS technique allows for the improvement of various figures of merits, ranging from clean and reliable chemical depth profiling to improved stoichiometry of the generated plasma. In this contribution, we demonstrate that the application of a DP-fs laser ablation ion source leads to an improved stoichiometry of the generated plasma. The second laser pulse supports the ionization of the atoms generated during the ablation as a result of the first pulse, which overall improves the stoichiometry of the recorded plasma. The optimization of both laser settings (delay and pulse energy) and detector system allowed the identification of the most suitable measurement conditions. Calibration factors, RSC values, for the investigated NIST steel alloys SRM 661, 664, and 665 are observed to be close to 1, which is of interest to measurement campaigns on unknown samples where the availability of matrix-matched reference samples is limited.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Andreas Riedo https://orcid.org/0000-0001-9007-5791
Rustam Lukmanov https://orcid.org/0000-0001-9257-7410
Valentine Grimaudo https://orcid.org/0000-0002-7010-5903
Coenraad de Koning https://orcid.org/0000-0002-2540-7689
Niels F. W. Ligterink https://orcid.org/0000-0002-8385-9149
Marek Tulej https://orcid.org/0000-0001-9823-6510
Peter Wurz https://orcid.org/0000-0002-2603-1169

REFERENCES

1. Neubeck A, Tulej M, Ivarsson M, et al. Mineralogical determination in situ of a highly heterogeneous material using a miniaturized laser ablation mass spectrometer with high spatial resolution. *Int J Astrobiol*. 2015;15:133-146.
2. Tulej M, Neubeck A, Ivarsson M, et al. Chemical composition of micrometer-sized filaments in an aragonite host by a miniature laser ablation/ionization mass spectrometer. *Astrobiology*. 2015;15(8): 669-682.
3. Wiesendanger R, Wacey D, Tulej M, et al. Chemical and optical identification of micrometer-sized 1.9 billion-year-old fossils by combining a miniature laser ablation ionization mass spectrometry system with an optical microscope. *Astrobiology*. 2018;18:1071-1080.
4. Yan W, Casey JF. A new concordia age for the 'forearc' bay of islands ophiolite complex, Western Newfoundland utilizing spatially-resolved LA-ICP-MS U-Pb analyses of zircon. *Gondw Res*. 2020;86:1-22.
5. Chen G, Gao J, Lu J, Zhang R. In situ LA-ICP-MS analyses of mica and wolframite from the Maoping tungsten deposit, southern Jiangxi, China. *Acta Geochim*. 2020;39(6):811-829.
6. Ye Y, Wang H, Wang X, Zhai L, Wu C, Zhang S. Elemental geochemistry of lower Cambrian phosphate nodules in Guizhou Province, South China: An integrated study by LA-ICP-MS mapping and solution ICP-MS. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2020;538: 109459.
7. Grijalba N, Legrand A, Holler V, Bouvier-Capely C. A novel calibration strategy based on internal standard–spiked gelatine for quantitative bio-imaging by LA-ICP-MS: Application to renal localization and quantification of uranium. *Anal Bioanal Chem*. 2020;412(13): 3113-3122.

8. Vogt DS, Schröder S, Rammelkamp K, Hansen PB, Kubitza S, Hübers HW. CaCl and CaF emission in LIBS under simulated Martian conditions. *Icarus*. 2020;335:113393.

9. Green T, Kuznetsov I, Willingham D, et al. Characterization of extreme ultraviolet laser ablation mass spectrometry for actinide trace analysis and nanoscale isotopic imaging. *J Anal At Spectrom*. 2017;32(6):1092-1100.

10. Liu Y, Hu Z, Gao S, et al. In situ analysis of major and trace elements of anhydrous minerals by LA-ICP-MS without applying an internal standard. *Chem Geol*. 2008;257(1-2):34-43.

11. Chang Z, Vervoort JD, McClelland WC, Knaack C. U-Pb dating of zircon by LA-ICP-MS. *Geochem Geophys Geosyst*. 2006;7(5):1-14.

12. Moreno-García P, Grimaudo V, Riedo A, et al. Towards structural analysis of polymeric contaminants in electrodeposited cu films. *Electrochim Acta*. 2016;199:394-402.

13. Moreno-García P, Grimaudo V, Riedo A, Tulej M, Wurz P, Broekmann P. Towards matrix-free femtosecond-laser desorption mass spectrometry for in situ space research. *Rapid Commun Mass Spectrom*. 2016;30(8):1031-1036.

14. Lee S, Gonzalez JJ, Yoo JH, Chirinos JR, Russo RE, Jeong S. Application of femtosecond laser ablation inductively coupled plasma mass spectrometry for quantitative analysis of thin Cu(In,Ga)Se$_2$ solar cell films. *Thin Solid Films*. 2015;577:82-87.

15. Hou H, Cheng L, Richardson T, et al. Three-dimensional elemental imaging of Li-ion solid-state electrolytes using fs-laser induced breakdown spectroscopy (LIBS). *J Anal At Spectrom*. 2015;30(11):2295-2302.

16. Gutierrez-Gonzalez A, Gonzalez-Gago C, Pisonero J, et al. Capabilities and limitations of LA-ICP-MS for depth resolved analysis of CdTe photovoltaic devices. *J Anal At Spectrom*. 2015;30(1):191-197.

17. Dong M, Wei L, González JJ, et al. Coal discrimination analysis using tandem laser-induced breakdown spectroscopy and laser ablation inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem*. 2020;92(10):7003-7010.

18. Syta O, Wagner B, Bulska E, et al. Elemental imaging of heterogeneous inorganic archaeological samples by means of simultaneous laser induced breakdown spectroscopy and laser ablation inductively coupled plasma mass spectrometry measurements. *Talanta*. 2018;179:784-791.

19. Bonta M, Gonzalez JJ, Quarles CD, Russo RE, Hegedus B, Limbeck A. Elemental mapping of biological samples by the combined use of LIBS and LA-ICP-MS. *J Anal At Spectrom*. 2016;31(1):252-258.

20. Bonta M, Limbeck A. Metal analysis in polymers using tandem LA-ICP-MS/LIBS: Eliminating matrix effects using multivariate calibration. *J Anal At Spectrom*. 2018;33(10):1631-1637.

21. Grimaudo V, Tulej M, Riedo A, et al. UV post-ionization laser ablation ionization mass spectrometry for improved nm-depth profiling resolution on Cr/Ni reference standard. *Rapid Commun Mass Spectrom*. 2020;34:e8803.

22. Käser D, Hendriks L, Koch J, Günther D. Depth profile analyses with sub 100-nm depth resolution of a metal thin film by femtosecond - laser ablation - inductively coupled plasma - time-of-flight mass spectrometry. *Spectrochim Acta Part B: At Spectrosc*. 2018;149:176-183.

23. Riedo A, Neuland M, Meyer S, Tulej M, Wurz P. Coupling of LMS with a fs-laser ablation ion source: Elemental and isotope composition measurements. *J Anal At Spectrom*. 2013;28(8):1256.

24. Zhang B, He M, Hang W, Huang B. Minimizing matrix effect by femtosecond laser ablation and ionization in elemental determination. *Anal Chem*. 2013;85(9):4507-4511.

25. Hergenröder R, Samek O, Hommes V. Femtosecond laser ablation elemental mass spectrometry. *Mass Spectrom Rev*. 2006;25(4):551-572.

26. Russo RE, Mao X, Gonzalez JJ, Mao SS. Femtosecond laser ablation ICP-MS. *J Anal At Spectrom*. 2002;17(9):1072-1075.

27. Zeng X, Mao XL, Greif R, Russo RE. Experimental investigation of ablation efficiency and plasma expansion during femtosecond and nanosecond laser ablation of silicon. *Appl Phys A*. 2005;80(2):237-241.

28. Hou H, Mao X, Zorba V, Russo RE. Laser ablation molecular isotopic spectrometry for molecules formation chemistry in femtosecond-laser ablated plasmas. *Anal Chem*. 2017;89(14):7750-7757.

29. Labutin TA, Lednev VN, Ilyin AA, Popov AM. Femtosecond laser-induced breakdown spectroscopy. *J Anal At Spectrom*. 2016;31(1):90-118.

30. Tulej M, Wiesendanger R, Riedo A, Knopp G, Wurz P. Mass spectrometric analysis of the mg plasma produced by double-pulse femtosecond laser irradiation. *J Anal At Spectrom*. 2018;33(8):1292-1303.

31. Piñon V, Fotakis C, Nicolas G, Anglos D. Double pulse laser-induced breakdown spectroscopy with femtosecond laser pulses. *Spectrochim Acta Part B: At Spectrosc*. 2008;63(10):1006-1010.

32. Stratis DN, Eland KL, Angel SM. Dual-pulse LIBS using a pre-ablation spark for enhanced ablation and emission. *Appl Spectrosc*. 2000;54(9):1270-1274.

33. González J, Liu C, Yoo J, Mao X, Russo RE. Double-pulse laser ablation inductively coupled plasma mass spectrometry. *Spectrochim Acta Part B: At Spectrosc*. 2005;60(1):27-31.

34. Grassi R, Grifoni E, Gufoni S, et al. Three-dimensional compositional mapping using double-pulse micro-laser-induced breakdown spectroscopy technique. *Spectrochim Acta Part B: At Spectrosc*. 2017;127:1-6.

35. Tulej M, Iakovleva M, Leya I, Wurz P. A miniature mass analyser for in-situ elemental analysis of planetary material–performance studies. *Anal Bioanal Chem*. 2011;399(6):2185-2200.

36. Riedo A, Bieler A, Neuland M, Tulej M, Wurz P. Performance evaluation of a miniature laser ablation time-of-flight mass spectrometer designed for in situ investigations in planetary space research. *J Mass Spectrom*. 2013;48(1):1-15.

37. Riedo A, Tulej M, Rohner U, Wurz P. High-speed microstrip multi-anode multichannel plate detector system. *Rev Sci Instrum*. 2017;88(4):045114.

38. Grimaudo V, Moreno-García P, Riedo A, et al. High-resolution chemical depth profiling of solid material using a miniature laser ablation/ionization mass spectrometer. *Anal Chem*. 2015;87(4):2037-2041.

39. Meyer S, Riedo A, Neuland MB, Tulej M, Wurz P. Fully automatic and precise data analysis developed for time-of-flight mass spectrometry. *J Mass Spectrom*. 2017;52(9):580-590.

40. He W, Fu Y, Zheng Y, et al. Polarization properties of a corner-cube retroreflector with three-dimensional polarization ray-tracing calculus. *Appl Optics*. 2013;52(19):4527-4535.

41. Liu J, Azzam RMA. Polarization properties of corner-cube retroreflectors: Theory and experiment. *Appl Optics*. 1997;36(7):1553-1559.

42. Tulej M, Neubeck A, Riedo A, et al. Isotope abundance ratio measurements using femtosecond laser ablation ionization mass spectrometry. *J Mass Spectrom*. 2020;55(12):e4660.

43. Neuland MB, Grimaudo V, Mezger K, et al. Quantitative measurement of the chemical composition of geological standards with a miniature laser ablation/ionization mass spectrometer designed for in situ application in space research. *Meas Sci Technol*. 2016;27(3):035904.

3.4 Molecular desorption using LDMS

3.4.1. ORIGIN: a novel and compact Laser Desorption – Mass Spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces

Niels F.W. Ligterink[1], ValentineGrimaudo[2], Pavel Moreno-García[3], Rustam Lukmanov[2], Marek Tulej[2], Ingo Leya[2], Robert Lindner[4], PeterWurz[2], Charles S. Cockell[5], Pascale Ehrenfreund[6,7] & Andreas Riedo[2]

[1]Center for Space and Habitability, University of Bern, Bern, Switzerland.
[2]Space Research and Planetary Sciences, Physics Institute, University of Bern, Bern, Switzerland.
[3]Interfacial Electrochemistry Group, Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland.
[4]Life Support and Physical Sciences Instrumentation Section, European Space Agency, ESTEC, Bern, The Netherlands.
[5]School of Physics and Astronomy, UK Centre for Astrobiology, University of Edinburgh, Edinburgh, United Kingdom.
[6]Laboratory for Astrophysics, Leiden Observatory, Leiden University, Leiden, The Netherlands.
[7]Space Policy Institute, George Washington University, Washington, DC, USA

SCIENTIFIC REPORTS

natureresearch

OPEN

ORIGIN: a novel and compact Laser Desorption – Mass Spectrometry system for sensitive *in situ* detection of amino acids on extraterrestrial surfaces

Niels F. W. Ligterink[1] ✉, Valentine Grimaudo[2], Pavel Moreno-García[3], Rustam Lukmanov[2], Marek Tulej[2], Ingo Leya[2], Robert Lindner[4], Peter Wurz[2], Charles S. Cockell[5], Pascale Ehrenfreund[6,7] & Andreas Riedo[6]

For the last four decades space exploration missions have searched for molecular life on planetary surfaces beyond Earth. Often pyrolysis gas chromatography mass spectrometry has been used as payload on such space exploration missions. These instruments have relatively low detection sensitivity and their measurements are often undermined by the presence of chloride salts and minerals. Currently, ocean worlds in the outer Solar System, such as the icy moons Europa and Enceladus, represent potentially habitable environments and are therefore prime targets for the search for biosignatures. For future space exploration missions, novel measurement concepts, capable of detecting low concentrations of biomolecules with significantly improved sensitivity and specificity are required. Here we report on a novel analytical technique for the detection of extremely low concentrations of amino acids using ORIGIN, a compact and lightweight laser desorption ionization – mass spectrometer designed and developed for *in situ* space exploration missions. The identified unique mass fragmentation patterns of amino acids coupled to a multi-position laser scan, allows for a robust identification and quantification of amino acids. With a detection limit of a few fmol mm^{-2}, and the possibility for sub-fmol detection sensitivity, this measurement technique excels current space exploration systems by three orders of magnitude. Moreover, our detection method is not affected by chemical alterations through surface minerals and/or salts, such as NaCl that is expected to be present at the percent level on ocean worlds. Our results demonstrate that ORIGIN is a promising instrument for the detection of signatures of life and ready for upcoming space missions, such as the Europa Lander.

The detection of signatures of life, past or present, on Solar System objects beyond Earth is of major importance for a better understanding on the presence of life in the universe and how it emerges. Habitability can be traced through several parameters[1,2], but in particular the detection of biomolecules such as amino acids, lipids and nucleobases on the surfaces of planets and moons are promising indicators for the presence of life. However, their unambiguous detection is extremely challenging and depends on various parameters. Past space exploration missions have focused on detecting biomolecules on Mars and Saturn's moon Titan[3–5], albeit without success. Under the premise that life exists or has existed on extraterrestrial bodies, current instruments, such as pyrolysis gas chromatography–mass spectrometry (pyr GC-MS), struggle with the detection of the biosignatures, partly due to

[1]Center for Space and Habitability, University of Bern, Bern, Switzerland. [2]Space Research and Planetary Sciences, Physics Institute, University of Bern, Bern, Switzerland. [3]Interfacial Electrochemistry Group, Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland. [4]Life Support and Physical Sciences Instrumentation Section, European Space Agency, ESTEC, Bern, The Netherlands. [5]School of Physics and Astronomy, UK Centre for Astrobiology, University of Edinburgh, Edinburgh, United Kingdom. [6]Laboratory for Astrophysics, Leiden Observatory, Leiden University, Leiden, The Netherlands. [7]Space Policy Institute, George Washington University, 20052, Washington, DC, USA. ✉e-mail: niels.ligterink@csh.unibe.ch

the presence of (chloride) salts and minerals[6–11]. To continue the search for molecular biosignatures[1,2], novel and robust life detection instruments and measurement techniques are important.

Two main candidates for the search of biosignatures in our Solar System are Europa and Enceladus, moons of Jupiter and Saturn, respectively. Previous studies found that these objects have mineral- and organic-rich subsurface oceans[12–15]. On their sea beds, hydrothermal vents could be present[16]. Similar to Earth, such environments represent promising habitats for life, since all ingredients exist for life to emerge[17]. Therefore, these so-called ocean-worlds are prime targets for future space exploration missions devoted to the search for extraterrestrial life[18–20]. Landers are the most promising spacecraft that are able to investigate their surfaces for the presence of biosignatures, in particular biomolecules[20,21]. Such molecules can be brought up from the subsurface oceans and survive within the icy surfaces for millions of years[22].

Laser Desorption Ionization Mass Spectrometry (LDI-MS) is a powerful tool for the analysis of molecules and allows measurements in various ways[23,24]. In its simplest form, LDI-MS instruments desorb material directly from a surface and ionize it with a single laser pulse, followed by the detection of the ions using a mass analyzer. LDI-MS has a number of advantages over existing space instruments. Due to its high sensitivity, only very low concentrations of analyte are required for a successful detection. LDI-MS instruments can be more compact, requires no consumables, such as carrier gases, and can operate without an extraction furnace, which all affect the dimensions, weight, and power consumption of the instrument; all stringent requirements for space instruments. Furthermore, we show that the detection of biomolecules is not affected by contaminants, such as chloride salts, which in contrast seriously limits the performance of current pyr GC-MS systems[8]. In recent years, various groups started working on LDI detection methods for space instrumentation[25–29] and a LDI- Quadrupole Mass Spectrometer (LDI-QMS) is part of the Mars Organic Molecule Analyzer (MOMA) suite on the upcoming ExoMars rover[30,31].

In this paper we present a novel measurement protocol using a compact LDI-MS system, called ORIGIN (ORganics Information Gathering INstrument) for the sensitive detection, identification, and quantification of amino acids. The system consists of a miniature time-of-flight mass spectrometer, developed for *in situ* detection of biomolecules on Solar System bodies, and a nanosecond laser system that desorbs and ionizes surface material with a single laser pulse. Laser desorption studies conducted on pure amino acids allowed the identification of their unique fragmentation patterns under the applied laser desorption conditions. This not only enables the identification of amino acids in more complex mixtures, but also allows their quantification. By careful monitoring specific major and minor biomolecule fragments, surface concentrations as low as a few fmol mm^{-2} can be detected. The results presented here are discussed in light of the requirements for upcoming space missions from ESA and NASA for the investigation of ocean worlds, such as Jupiter's moon Europa, planned to be realised beyond 2020.

Results

The ORIGIN system consists of a miniature reflectron time-of-flight mass spectrometer (RToF-MS, 160 mm × Ø 60 mm, m/Δm ≈ 1'000)[32] that has a nanosecond pulsed laser system as an ion source. The mass analyzer is axis symmetric with a central hole at the entrance and exit. The nanosecond laser pulses ($\tau \approx 3$ ns, $\lambda = 266$ nm) are guided through a beam expander and via various mirrors to the focusing lens. The laser pulses are guided through the mass analyzer to the sample surface, which is positioned below the mass analyzer. Importantly, the sample is positioned slightly out of focus, i.e., the focal point is about a millimeter below the exit. Below the exit, i.e., below the mass analyzer, a steel sample holder on a X,Z translation stage is placed, where the laser desorbs and ionizes the material. Only cations can enter the ion optical system of the mass analyzer and are guided to the micro channel plate (MCP) detector system[33]. For every laser pulse a full ToF spectrum is recorded, which is converted to a mass spectrum (see Sect. 4.1 for a full description of the system). The sample, i.e., organic films of amino acids, are prepared in polished cavities (Ø 3 mm) by dropcasting 1 µL amino acid solution with concentrations ranging from 100–1 µM, resulting in average surface concentrations of 14–0.14 pmol mm^{-2} (see Sect. 4.2 for the sample preparation procedure). Each of these produced cavities is sampled at 40 positions, linearly spaced by ~50 µm, with 100 laser shots at each position (see Sect. 4.3 for the measurement protocol). The resulting mass spectra are filtered based on peak signal-to-noise ratios (SNRs) to remove spectra without signal. For the data handling and data analysis we use in-house developed software (see Sect. 4.4 for the analysis protocol)[34].

Amino acid detection and identification. The first objective of this study was to show that ORIGIN can detect and identify amino acids and salts placed on a steel surface. Therefore, twenty samples of pure proteinogenic and abiotic amino acid solutions at a concentration of 14 pmol mm^{-2} and an equal parts NaCl/KCl mixture at 0.7 µg mm^{-2} were measured (the full list of amino acids is given in Sect. 4.2). The lowest pulse energy at which a signal was identified was used for the measurement and thus differs for each molecule. The resulting mass spectra are presented in Fig. 1 (top). Except for lysine, all amino acids are detected and display sparse mass fragmentation patterns. Importantly, for the majority of amino acids unique mass fragmentation patterns were observed, although isomers and enantiomers are in some cases difficult to differentiate (e.g. (iso)leucine and L/R-AABA). Figure 1 (bottom) shows the mass fragmentation contributions of the amino acids, divided into parent peak (green), amino acids without the carboxyl (−COOH, 45 amu, red) group, amino acid side chains (amino acid minus 74 amu, blue) and "other" (purple). In the first three groups also (de)protonated fragments are included. It can be seen that the fragmentation patterns are dominated by -COOH stripped and side chain masses. The signal of the NaCl/KCl mixture is dominated by Na$^+$ and K$^+$ ions and clusters of X$^+$(YCl), where X and Y can be Na or K.

Multi-position scan and quantification of organic film material. In Fig. 2, measurements of the amino acids methionine and histidine are highlighted. In panel (A) the intensity of two fragment masses at each
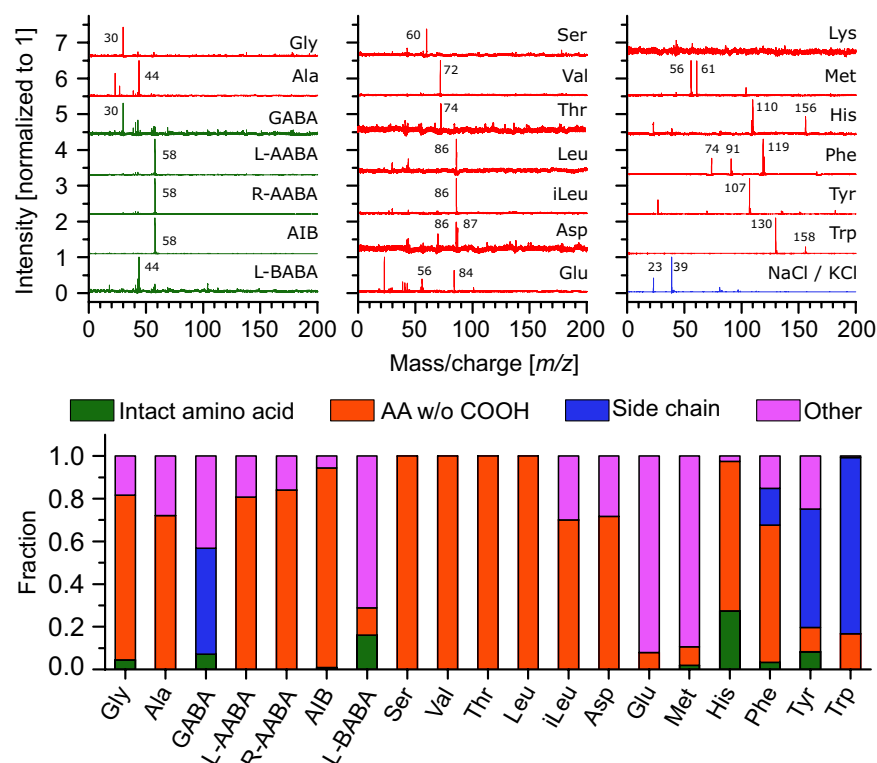
Figure 1. Top: Mass spectra of various biotic (red) and abiotic (green) amino acids (14 pmol mm^{-2} average surface concentration), and of a mixture of NaCl/KCl salt (blue, ~0.7 µg mm^{-2} average surface concentration). See Sect. 4.2 for a list of the abbreviations used for the amino acids. The primary fragments are labelled according to their mass. Bottom: Fragmentation pattern of the measured amino acid spectra (lysine omitted), grouped by masses corresponding to the parent (i.e. intact) molecule (green), the amino acid without its −COOH group (red), the side chain of the amino acid (blue), and other contributions (purple). Signals from −COOH-stripped and side chain masses dominate the amino acid spectra.

of the forty analysed positions is shown, for an average surface concentration of 14 pmol mm^{-2}. Signals that are below a SNR of six are excluded. Intensity variations up to an order of magnitude are seen from position to position, which is due to concentration gradients in, and non-uniformity of, the organic film. This clearly demonstrates that the quantification of dried amino acid films using only one spot analysis is highly unreliable. The multi-position measurement protocol circumvents this problem by accumulating signal over the entire diameter of the sample cavity, covering low and high organic film concentrations and therefore effectively averaging the signal. In panel B) the ratios of the two amino acid fragments at each position are displayed and compared to the average ratio obtained from co-added data. Most of the individual ratios are within the 1σ standard deviation obtained from the accumulated data, showing that there is minimal deviation from spot to spot and demonstrating that the fragmentation pattern is quite uniform. In panel C) the accumulated intensities of a single mass peak as a function of the average surface concentration, which ranged from 0.14–14 pmol mm^{-2}, are displayed. The data show a fitted linear correlation between signal intensity and surface concentration. From the obtained linear correlation one can conclude that the accumulation of spectra from multiple positions represents a simple and robust procedure for the quantification of amino acid concentrations. The bottom panels depict simulated spectra using the established fragmentation patterns of methionine and histidine, which agree well with the measured data and reproduce the mass spectra well. Similar to spectroscopic measurements where e.g., molecules are identified through their corresponding unique spectroscopic patterns, such a software routine is of high interest to current experiments as amino acids may be identified and quantified in more complex mixtures, including additional contaminates.

Analysis of amino acid mixture. To test the performance of the system on complex samples, measurements were performed using mixtures containing all twenty amino acids (0.7 pmol mm^{-2} each) with and without added NaCl (~0.7 µg mm^{-2}). NaCl is added to simulate measurement conditions on icy moons, which are known to have large amounts of salt embedded in the ice. Figure 3 shows the results for scans at two applied pulse energies (1.4 and 2.6 µJ). The signals are clearly visible and the signal size increases with increasing pulse energy. Except for lysine, mass fragments of all amino acids are clearly visible in the 2.6 µJ measurement. However, not all signals can be unequivocally assigned to a single amino acid, since some of the mass fragments overlap. For example, this hinders the unambiguous identification of alanine just based on the $m/z = 44$ signal. When NaCl is added, the amino acid can still be identified via the mass fragments, however the signal intensities are observed
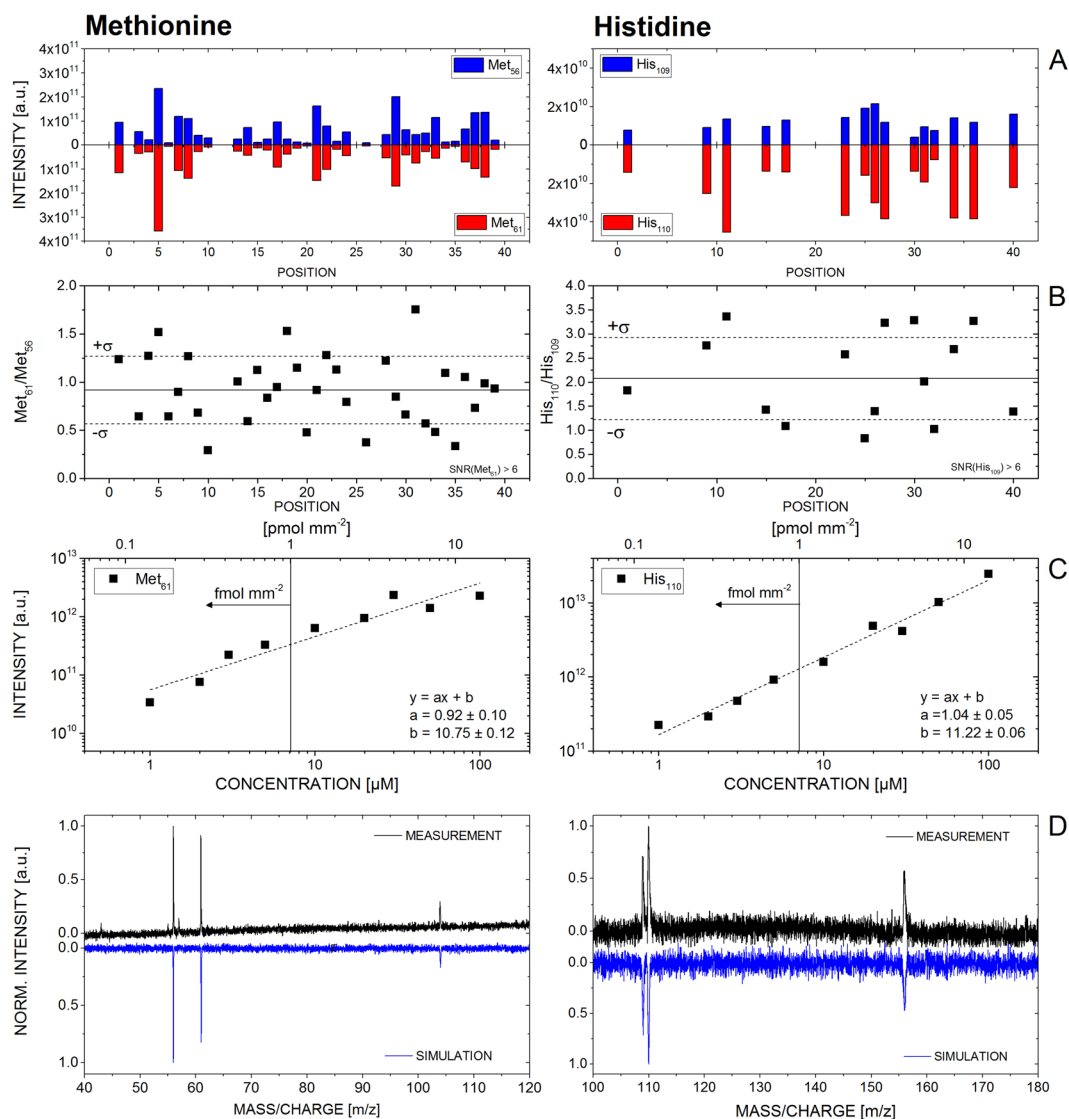
Figure 2. Results of measurements of the amino acids methionine (left) and histidine (right). (**A**) Measured signal intensities of mass 56 and 61 of methionine and 109 and 110 of histidine, at each of the forty positions per scan. Empty positions have signals below a signal-to-noise ratio of six, which are omitted. (**B**) Ratios between the corresponding mass fragments at each of the forty positions. The solid line indicates the mean ratio from the co-added data of all positions. The dashed lines give the 1σ standard deviation of the mean of the measured data. (**C**) Intensities measured as a function of average surface concentration for concentrations in the range of 0.14–14 pmol mm^{-2} (1–$100\,\mu$M solution concentration), including linear regression coefficients. (**D**) Synthetic mass spectra (blue) compared to the measured data (black).

to be lower. Sodium chloride is mostly UV transparent and thus the laser pulse will interact with the amino acids. Therefore, it is likely that the salt crust reduces desorption and/or ionization of the amino acids and higher laser pulse energies may be required. The bottom panel of Fig. 3 shows for the amino acids detected in the $2.6\,\mu$J pulse campaign the 3σ limit of detection (LOD$_{3\sigma}$, see Sect. 4.5). The high SNR of the mass fragments makes it possible to detect many of the amino acids at surface concentrations below 100 fmol mm^{-2}, and in the case of tryptophan even down to 1 fmol mm^{-2}.

Discussion

In this work, we demonstrate that the compact and lightweight ORIGIN system has the measurement capabilities for the identification and quantification of amino acids. With a LOD$_{3\sigma}$ down to a few fmol mm^{-2} and limit of quantification (LOQ) in the order of 100 fmol mm^{-2}, it excels the LDI-QMS instrument (LOD$_{3\sigma} \leq 1$ pmol mm^{-2}), which is part of the MOMA suite[30] on the ExoMars rover, by 1–3 orders of magnitude. Amino acid identification in mixtures is possible, although for some species a deconvolution procedure for studying the fragmentation pattern is needed to distinguish overlapping mass contributions. The need for a deconvolution step will be even more important for real samples, which may contain other organic molecules. It is important to note that the
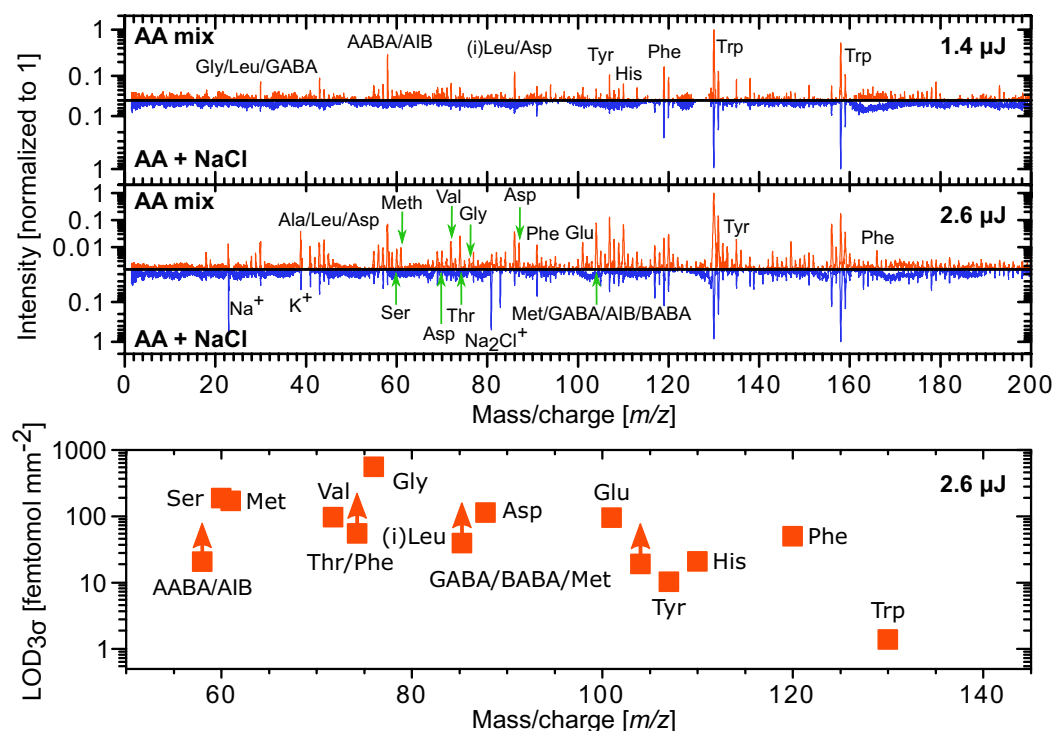
Figure 3. Top: Measured spectra of a mixture of 20 amino acids (0.7 pmol mm^{-2} per amino acid, red) and the same mixture with added NaCl (~0.7 µg mm^{-2}, blue). The mixtures were measured with pulse energies of 1.4 (upper panel) and 2.6 µJ (middle panel). Out of the 20 amino acids, eleven can directly be identified from their unique mass features. Bottom panel: 3σ limits of detection (LOD$_{3σ}$) calculated from the 2.6 µJ measurement. Lower limits are given for signals that are clearly detected but have contributions from multiple amino acids.

detection sensitivity of our system is currently not set up to maximum sensitivity. For example, the sensitivity can still be increased by ramping the voltage potential over the MCP detector stack[33,35]. A voltage increase over the MCP stack allows the generation of more electrons per single incoming ion and therefore produces an increased recorded signal. In the current setup, the detector gain can be increased by a factor of 32. Furthermore, the multi-position scan method used for the quantification of the organic film material can also be used to increase overall sensitivity by studying more positions on the sample, similar to previous studies conducting ablation campaigns of solids samples[36]. Scanning more positions allows for the direct increase of the SNR and consequently the detection sensitivity. With both methods, the detection sensitivity of the system can be significantly enhanced and is expected to reach sub-fmol mm^{-2} levels.

Besides its high detection sensitivity, ORIGIN has a number of advantages over existing systems for the detection of amino acids in space. For example, with ORIGIN we see no significant chemical reactions between amino acids and chloride-containing compounds, as reported in pyr GC-MS systems[8]. A decrease in amino acid signal intensities is seen, probably caused by NaCl reducing desorption and/or ionization, but in contrast to pyr GC-MS measurements, amino acids remain detectable with ORIGIN. In addition, the presence of NaCl or any other species at high concentration does not saturate the detector thanks to its large dynamic range of up to eight orders of magnitude[33]. The close proximity (~1 mm) of the ToF-MS ion entrance to the sample holder results in a very efficient ion coupling from the desorption plume to the mass spectrometer. Together with a high ion transmission of the time-of-flight analyzer system, low laser pulse energies of only a few µJ can be used in the LDI process, which minimizes molecule fragmentation, enabling a clear identification of biomolecules even in complex mixtures.

A future application of the ORIGIN system could be as payload on a Europa Lander mission, where it can serve as part of the Organic Compositional Analyzer (OCA) to investigate the biomolecular and amino acid content in the icy surface material of Europa. According to the Europa Lander Science Definition report[20], the OCA needs to be able to detect species at 1 nM in a 1 gram (=1 ml) surface sample. Melting a 1-gram sample and producing an organic film on a surface area similar to the surface area of the cavities used in this work (7.1 mm²), results in an average surface concentration of 141 fmol mm^{-2}; a concentration easily detectable by ORIGIN. A further requirement in the Europa Lander Science Definition report[20] is the detection of eight of the following amino acids: Ala, Asp, Glu, His, Leu, Ser, Val, Iva, Gly, β-Ala, GABA, and AIB, ten of which have already been successfully measured by ORIGIN (β-Ala and Iva were not included in the sample). However, securely identifying isomers and enantiomers is currently challenging. Finally, we have demonstrated that the ORIGIN system permits also detecting Na and K and the salts NaCl and KCl. It is very likely that other salts, silicates and metals can also be identified. The results demonstrate that ORIGIN is a promising instrument for the detection of life in upcoming space missions, specifically for the Europa Lander.
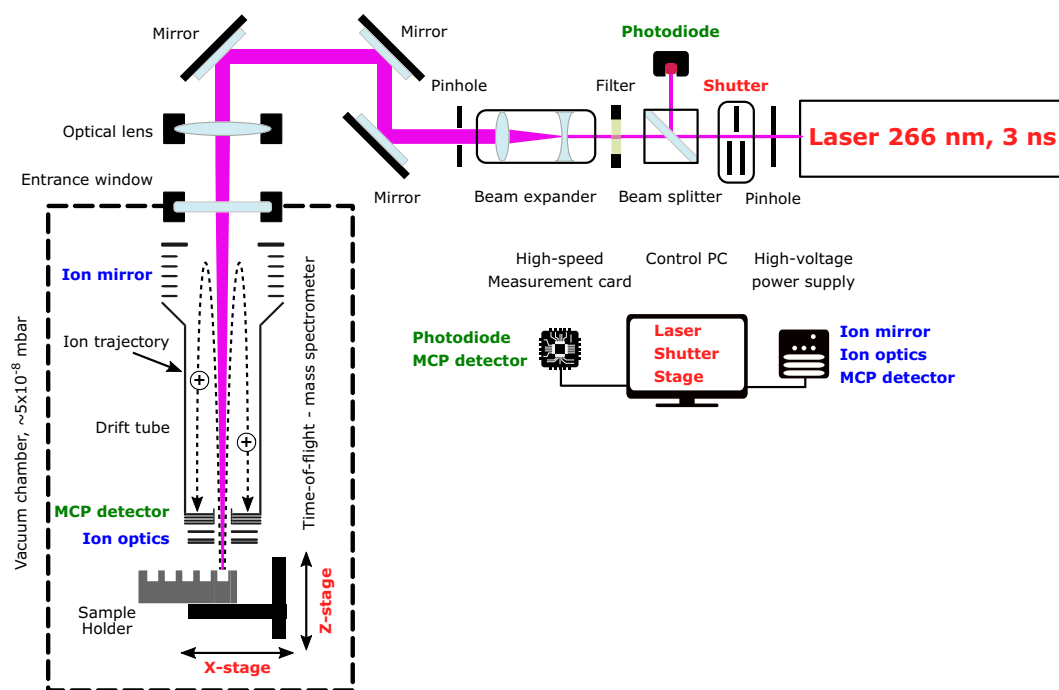
Figure 4. Schematic diagram of the ORIGIN system (not to scale) and of the wiring of the individual components. The mirrors, optical lens, and the beam expander are all placed on stages and manipulators to adjust the laser path.

Methods

Laser desorption mass spectrometry set-up.   To facilitate the detection of biomolecules on surfaces of Solar System objects, a novel and compact Laser Desorption/Ionization – Mass Spectrometry (LDI-MS) instrument has been designed and constructed (a schematic depiction is shown in Fig. 4), based on our experiences with existing systems and components[32,37–43]. This system is called ORIGIN (ORganics Information Gathering INstrument). The design is simple and robust, which establishes LDI-MS as a rugged instrument for *in situ* space exploration missions, while at the same time being lightweight and compact.

A Q-switched Nd:YAG laser (Quantel Brio, France) is used for desorption and ionization of sample material (pulse width $\tau \approx 3$ ns, wavelength $\lambda = 266$ nm, pulse repetition rate $= 20$ Hz). The laser system is equipped with a beam attenuator module for laser pulse energy adjustment. The system guiding the beam to the mass analyzer, the latter is located in a vacuum chamber, consists of a shutter system, attenuator filters, beam expander, high reflective silver mirrors, and a lens system. The shutter system (SH05, Thorlabs) is integrated directly after the laser system and controls the number of laser pulses for each measurement campaign. Following the shutter, an additional set of attenuation filters is placed in the path of the laser beam to (i) operate the laser system at best performance (the stability of the pulse energy is better at higher laser output) and to (ii) reflect laser light to an ultrafast Si-photodiode (Alphalas, Germany) that triggers the data acquisition system (see below). Laser pulses pass through a beam expander (TECHSPEC Vega, Edmund Optics), which expands the beam to about 15 mm in diameter. An iris installed after the expander system is used to cut out the homogenous inner part of the laser pulse (in terms of energy). Via high reflectivity silver mirrors the cleaned laser pulses are guided towards the focusing lens (f = 300 mm), which is mounted on a Z-stage and which allows, first, to keep the sample at an optimal distance relative to the mass analyzer (around 1 mm) and, second, to adjust the position of the laser focus. Subsequently, the laser light passes through a vacuum entrance window into the vacuum chamber where the miniature mass analyzer is located. Typically, laser pulses with energies ranging from 1.4–3.0 µJ, in increasing steps of 0.4 µJ, are applied in the course of our measuring campaigns. With an optical microscope the circular laser ablation spot size was measured to be around 20 µm. For the desorption measurements the laser is slightly out of focus and based on a simplified geometry we calculate the desorption spot to be about 30 µm in diameter. This method likely underestimates the spot size. With a spot size of 30 µm, an energy density of 0.20–2.35 J cm^{-2} or laser irradiances of 66–142 MW cm^{-2} is achieved in the measurements. Just below the laser entrance window, a miniature (160 mm × Ø 60 mm) reflectron-type time-of-flight mass spectrometer (RToF-MS)[32] is housed. This RToF-MS system is specifically designed for *in situ* space exploration and capable of achieving a mass resolution at desorption conditions of m/Δm $\geq$ 1'000 with a dynamic range of up to eight orders of magnitude[33]. Voltages of the ion optical system and the detector system (multichannel plate system, chevron configuration) are set remotely using a high voltage power supply. The detector signal is read out by a high-speed analog-to-digital-converter card (U1084A, Agilent, up to 4 GS s^{-1}). A high-speed Si-photodiode is used to trigger the data acquisition. The laser pulses pass along the central axis of the mass analyzer. Underneath the RToF-MS, a sample holder is placed on a X,Z-translation stage (Agilis, Newport). With this stage, samples can be moved relative to the laser spot and

Amino acid	Laser pulse energy	Amino acid	Laser pulse energy
Glycine	1.8 μJ	Leucine	1.4 μJ
Alanine	1.8 μJ	Isoleucine	1.4 μJ
γ-aminobutryic acid	1.8 μJ	Aspartic acid	1.4 μJ
R-α-aminobutryic acid	1.8 μJ	Glutamine	2.2 μJ
L-α-aminobutyric acid	1.8 μJ	Lysine	3.4 μJ (not detected)
α-aminoisobutyric acid	1.8 μJ	Methionine	1.4 μJ
L-β-aminobutanoic acid	1.8 μJ	Histidine	1.8 μJ
Serine	1.4 μJ	Phenylalanine	1.4 μJ
Valine	1.4 μJ	Tyrosine	1.8 μJ
Threonine	1.8 μJ	Tryptophan	1.4 μJ

Table 1. Applied laser pulse energies for the detection of various amino acids.

distinct positions along the X-axis can be probed, in both cases at micrometer precision. The sample holders are introduced into the vacuum chamber and onto the stage via an entrance port. The entire set-up is operated using in-house written Python software, giving control over the laser system, shutter, and sample stage. Fully automated routines make it possible to scan samples at multiple positions, for a certain number of laser pulses and at various laser powers.

The vacuum chamber is evacuated by a small turbomolecular pump (Pfeiffer TMU 071 P, 80 L N_2) that is backed by an oil free membrane pump, which results in a typical base pressure of about 5×10^{-8} mbar. The vacuum chamber is vented using high purity nitrogen (Alphagas 2, Garbagas) that avoids contaminations during sample exchange. Between the vacuum chamber and pumping system, an electronic shutter valve is placed (VAT), which is closed before changing the sample.

Sample preparation. To measure amino acids and salts at low concentrations and without contaminants, clean working conditions and materials are mandatory. All samples were prepared in a sterilized flow hood. A stainless steel (416 L) sample holder (Ø 27 mm), which has five equally spaced cavities (0.2 x Ø 3 mm), is used to prepare the samples on and introduce them into the LDI-MS setup. Before being used, the sample holder is carefully cleaned by washing it with isopropanol in an ultrasound bath for 15 minutes. Extensive rinsing with Milli-Q grade water (Milli-Q Gradient, TOC < 5 ppb) is followed by a one-hour bake-out at 80 °C to minimize the water content on the sample holder. Finally, the surface of the sample holder is flame sterilized by a propane/butane burning flame at T > 500 °C.

The following fifteen proteinogenic amino-acids were used in this work: Glycine (Gly), L-Alanine (Ala), L-Serine (Ser), L-Valine (Val), L-Threonine (Thr), L-Leucine (Leu), L-isoLeucine (iLeu), L-Aspartic acid (Asp), L-Glutamine (Glu), L-Lysine (Lys), L-Methionine (Met), L-Histidine (His), L-Phenylalanine (Phe), L-Tyrosine (Tyr) and L-Tryptophan (Trp). In addition, five abiotic amino acids were measured: γ-Aminobutyric acid (GABA), L/R-α-Aminobutyric acid (L/R-AABA), α-aminoisobutyric acid (AIB) and L-β-Aminobutyric acid (L-BABA). All amino acids were purchased from Sigma-Aldrich with a purity of >99%. As salts, NaCl (Roth, >99.5% purity) and KCl (Hanseler, >99.5% purity) were used. For each amino acid, a stock solution of 100 μM in Milli-Q grade water was prepared, which was further diluted to concentrations of 50, 30, 20, 10, 5, 3, 2, and 1 μM. The salt solutions were made at a concentration of 1%wt, matching current measurements of the surface of the icy moon Enceladus[13] and assumed to be similar for Europa, where Na and K have been detected to originate from the surface[44,45]. During the preparation of the solutions, sterilized Eppendorf tubes (Eppendorf BIOPUR Safe-lock tubes) and pipet tips were used (Eppendorf BIOPUR epT.I.P.S.). To prepare the samples on the sample holder, 1 μL of an amino acid or salt solution is dropcast into a cavity and left to dry, producing an organic film and/or a salt crust. Assuming a uniform distribution, amino acid surface concentrations of 0.14–14 pmol mm^{-2} are achieved ([concentration mol L^{-1} $\times 1 \times 10^{-6}$ L]/[$\pi \times 1.5^2$ mm^2]). Assuming an average amino acid size of 0.5 by 0.5 nm, this converts to a uniform coverage of just 0.2–21 molecular layers of amino acids. However, in reality, the organic film and salt crust are not uniformly distributed and concentration gradients within the cavity are possible.

Measurement protocol. After sample preparation, the sample holder is introduced into the set-up and the sample chamber is evacuated overnight to reach a pressure of ~5×10^{-8} mbar. Before measurements, the data acquisition system, ion optics including detector system, power supply, and laser are given about 45 minutes to warm up and stabilize. The sample holder is moved so that a sample cavity is just outside the laser focus, at ~1 mm distance relative to the RToF entrance electrode.

For each of the studied amino acids, scans with increasing laser power were made to find the lowest laser irradiance at which desorption and ionization of the molecule of interest takes place. This ensures limited fragmentation of the amino acids. The pulse energy used for the detection of each amino acid is given in Table 1.

The non-uniform distribution of sample material in the cavity can result in misleading data when just a single position is sampled. To minimize this effect, a linear scan that consists of 40 positions with a pitch of 50 μm is conducted, which increase our measurement statistics. Subsequently the data are co-added, which effectively averages

out variations due to variable sample material concentrations. Doing so, the measured mass signal intensity corresponds linearly to a specific average surface concentration (see Fig. 2c). After having measured the five sample cavities, mass spectra of the clean sample holder (area between cavities) were taken at similar measurement conditions for the purpose of mass calibration and to identify contamination from the sample holder, if present.

Data analysis. All data are analysed with in-house developed Matlab software[34]. Time-of-flight spectra are selected for peaks with a specified SNR (usually ≥ 6) or for peaks within a certain time window ($=m/z$ window). Subsequently, time-of-flight spectra of each single measurement position are co-added, where after spectra of the 40 positions can be co-added to a single ToF spectrum. The software corrects for time-of-flight shifts caused by surface charging, which can occur when the organic film is thick enough to act as an insulator. The interaction of the laser with the film positively charges the latter, which repels the produced cations and makes them arrive slightly earlier at the detector system. The correction is performed by an autocorrelation correction by selecting a peak, which is present in all the acquired time-of-flight spectra. One spectrum is used as a reference spectrum and all the other spectra are shifted so the peak positions of the selected peaks align. Measurements of the steel sample holder are used for the time-of-flight to mass-to-charge conversion by identifying specific ions ablated from the steel holder. The software also generates various statistics from a mass spectrum, such as peak position, integrated peak area, peak full width at half maximum (FHWM), and peak SNR by dividing the peak area by the integrated noise in a region where no signal appears. Blank measurements of the steel sample holder are conducted in each measurement run and used to determine that masses assigned to an amino acid have no or negligible contributions from steel surface contaminants. Simulated, synthetic mass spectra of measured mass spectra are created with our in-house developed software[34].

Limit of detection. The limit of detection (LOD) can be given in various ways. The simplest way is to use surface concentration, as we elaborate in Sect. **4.2** by giving $C_{surface} = 0.14$–14 pmol mm^{-2}. This procedure is only applicable if a signal is seen for a particular surface concentration $C_{surface}$. If the SNR of a detected amino acid mass fragment is sufficiently high, the limit of detection can be lower. Taking a 3σ SNR as the limit for detection, the LOD for a surface concentration is given as $\mathrm{LOD}_{3\sigma} = (3/[SNR]_{fragment}) \times C_{surface}$, where $[SNR]_{fragment}$ is the fragment signal-to-noise ratio and $C_{surface}$ is the surface concentration of the measurement in question. For example, in the measurements of the mixture of 20 amino acids the $C_{surface} = 0.7$ pmol mm^{-2} per amino acid. For tryptophan, the $[SNR]_{130} = 1731$ for the $m/z = 130$ fragment in the 2.6 µJ scan (see Fig. 3), resulting in $\mathrm{LOD}_{3\sigma} = 1.2$ fmol mm^{-2}.

For a meaningful comparison with the LOD requirements given in the Europa Lander Science definition report (1 nM material in a 1 g, or 1 ml, sample, which corresponds to 1 pmol of material), these numbers need to be converted to average surface coverage. Sublimating the water of this sample and distributing the remaining molecules over the same area as the sample holder cavities, i.e., 7.1 mm^2, results in a surface concentration of 141 fmol mm^{-2}.

It remains important to stress that the sensitivity of the system can easily be improved. For example, sampling 400 positions, instead of 40, of the same surface results in a tenfold increase in signal and would already push the $\mathrm{LOD}_{3\sigma}$ of the system into the sub-fmol mm^{-2} regime. In addition, the gain of the MCP detector can still be increased by a factor of 32.

References

1. Aerts, J., Röling, W., Elsaesser, A. & Ehrenfreund, P. Biota and Biomolecules in Extreme Environments on Earth: Implications for Life Detection on Mars. *Life* **4**, 535 (2014).
2. Hays, L. E. *et al.* Biosignature Preservation and Detection in Mars Analog Environments. *Astrobiology* **17**, 363–400, https://doi.org/10.1089/ast.2016.1627 (2017).
3. Biemann, K. *et al.* Search for organic and volatile inorganic compounds in two surface samples from the Chryse Planitia region of Mars. *Science* **194**, 72–76 (1976).
4. Niemann, H. *et al.* The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438**, 779 (2005).
5. Leshin, L. *et al.* Volatile, isotope, and organic analysis of martian fines with the Mars Curiosity rover. *Science* **341**, 1238937 (2013).
6. Benner, S. A., Devine, K. G., Matveeva, L. N. & Powell, D. H. The missing organic molecules on Mars. *Proceedings of the National Academy of Sciences* **97**, 2425–2430 (2000).
7. Navarro-González, R. *et al.* The limitations on organic detection in Mars-like soils by thermal volatilization–gas chromatography–MS and their implications for the Viking results. *Proceedings of the National Academy of Sciences* **103**, 16089–16094 (2006).
8. Royle, S. H. *et al.* Perchlorate-Driven Combustion of Organic Matter During Pyrolysis-Gas Chromatography-Mass Spectrometry: Implications for Organic Matter Detection on Earth and Mars. *Journal of Geophysical Research: Planets* **123**, 1901–1909 (2018).
9. Montgomery, W. *et al.* Effects of Oxygen-Containing Salts on the Detection of Organic Biomarkers on Mars and in Terrestrial Analog Soils. *Astrobiology* (2019).
10. François, P. *et al.* Magnesium sulfate as a key mineral for the detection of organic molecules on Mars using pyrolysis. *Journal of Geophysical Research: Planets* **121**, 61–74 (2016).
11. Lasne, J. *et al.* Oxidants at the surface of mars: a review in light of recent exploration results. *Astrobiology* **16**, 977–996 (2016).
12. Khurana, K. *et al.* Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* **395**, 777 (1998).
13. Postberg, F. *et al.* Sodium salts in E-ring ice grains from an ocean below the surface of Enceladus. *Nature* **459**, 1098 (2009).
14. Postberg, F. *et al.* Macromolecular organic compounds from the depths of Enceladus. *Nature* **558**, 564 (2018).
15. Roth, L. *et al.* Transient water vapor at Europa's south pole. *Science* **343**, 171–174 (2014).
16. Waite, J. H. *et al.* Cassini finds molecular hydrogen in the Enceladus plume: evidence for hydrothermal processes. *Science* **356**, 155–159 (2017).
17. Westall, F. *et al.* A Hydrothermal-Sedimentary Context for the Origin of Life. *Astrobiology* **18**, 259–293, https://doi.org/10.1089/ast.2017.1680 (2018).

18. Grasset, O. *et al.* JUpiter ICy moons Explorer (JUICE): An ESA mission to orbit Ganymede and to characterise the Jupiter system. *Planetary and Space Science* **78**, 1–21 (2013).
19. Phillips, C. B. & Pappalardo, R. T. Europa Clipper mission concept: exploring Jupiter's ocean moon. *Eos, Transactions American Geophysical Union* **95**, 165–167 (2014).
20. Hand, K. P. *et al.* Report of the European Lander Science Defintion Team. (2017).
21. Pappalardo, R. *et al.* Science potential from a Europa lander. *Astrobiology* **13**, 740–773 (2013).
22. Nordheim, T., Hand, K. & Paranicas, C. Preservation of potential biosignatures in the shallow subsurface of Europa. *Nature Astronomy* **2**, 673 (2018).
23. Posthumus, M., Kistemaker, P., Meuzelaar, H. & Ten Noever de Brauw, M. Laser desorption-mass spectrometry of polar nonvolatile bio-organic molecules. *Analytical chemistry* **50**, 985–991 (1978).
24. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **60**, 2299–2301 (1988).
25. Getty, S. A., Brinckerhoff, W. B., Cornish, T., Ecelberger, S. & Floyd, M. Compact two-step laser time-of-flight mass spectrometer for *in situ* analyses of aromatic organics on planetary missions. *Rapid Communications in Mass Spectrometry* **26**, 2786–2790 (2012).
26. Uckert, K. *et al.* IR resonance-enhanced organic detection with two-step laser desorption time-of-flight mass spectrometry. *Icarus* **299**, 15–21 (2018).
27. Arevalo, R. Jr. *et al.* An Orbitrap-based laser desorption/ablation mass spectrometer designed for spaceflight. *Rapid Communications in Mass Spectrometry* **32**, 1875–1886 (2018).
28. Briois, C. *et al.* Orbitrap mass analyser for *in situ* characterisation of planetary environments: Performance evaluation of a laboratory prototype. *Planetary and Space Science* **131**, 33–45 (2016).
29. Selliez, L. *et al.* Identification of organic molecules with a laboratory prototype based on the Laser Ablation-CosmOrbitrap. *Planetary and Space Science* **170**, 42–51 (2019).
30. Goesmann, F. *et al.* The Mars Organic Molecule Analyzer (MOMA) instrument: characterization of organic material in martian sediments. *Astrobiology* **17**, 655–685 (2017).
31. Li, X. *et al.* Mars Organic Molecule Analyzer (MOMA) laser desorption/ionization source design and performance characterization. *International Journal of Mass Spectrometry* **422**, 177–187 (2017).
32. Rohner, U., Whitby, J. A. & Wurz, P. A miniature laser ablation time-of-flight mass spectrometer for *in situ* planetary exploration. *Measurement Science and Technology* **14**, 2159 (2003).
33. Riedo, A., Tulej, M., Rohner, U. & Wurz, P. High-speed microstrip multi-anode multichannel plate detector system. *Review of Scientific Instruments* **88**, 045114, https://doi.org/10.1063/1.4981813 (2017).
34. Meyer, S., Riedo, A., Neuland, M. B., Tulej, M. & Wurz, P. Fully automatic and precise data analysis developed for time-of-flight mass spectrometry. *Journal of Mass Spectrometry* **52**, 580–590, https://doi.org/10.1002/jms.3964 (2017).
35. Riedo, A., Rout, S., Wiesendanger, R., Wurz, P. & Leya, I. EGT—A sensitive time-of-flight mass spectrometer for multielement isotope gas analysis. *Journal of mass spectrometry* **53**, 1036–1045 (2018).
36. Riedo, A. *et al.* Novel 2D binning approach for advanced LIMS depth profiling analysis. *Journal of Analytical Atomic Spectrometry* (2019).
37. Moreno-García, P. *et al.* Towards matrix-free femtosecond-laser desorption mass spectrometry for *in situ* space research. *Rapid Communications in Mass Spectrometry* **30**, 1031–1036, https://doi.org/10.1002/rcm.7533 (2016).
38. Neubeck, A. *et al.* Mineralogical determination *in situ* of a highly heterogeneous material using a miniaturized laser ablation mass spectrometer with high spatial resolution. *International Journal of Astrobiology* **15**, 133–146, https://doi.org/10.1017/S1473550415000269 (2015).
39. Neuland, M. B. *et al.* Probing the Allende meteorite with a miniature laser-ablation mass analyser for space application. *Planetary and Space Science* **101**, 196–209, https://doi.org/10.1016/j.pss.2014.03.009 (2014).
40. Riedo, A., Bieler, A., Neuland, M., Tulej, M. & Wurz, P. Performance evaluation of a miniature laser ablation time-of-flight mass spectrometer designed for *in situ* investigations in planetary space research. *Journal of Mass Spectrometry* **48**, 1–15, https://doi.org/10.1002/jms.3104 (2013).
41. Riedo, A. *et al.* Highly accurate isotope composition measurements by a miniature laser ablation mass spectrometer designed for *in situ* investigations on planetary surfaces. *Planetary and Space Science* **87**, 1–13, https://doi.org/10.1016/j.pss.2013.09.007 (2013).
42. Riedo, A., Neuland, M., Meyer, S., Tulej, M. & Wurz, P. Coupling of LMS with a fs-laser ablation ion source: elemental and isotope composition measurements. *Journal of Analytical Atomic Spectrometry* **28**, 1256–1269, https://doi.org/10.1039/C3JA50117E (2013).
43. Tulej, M. *et al.* Chemical Composition of Micrometer-Sized Filaments in an Aragonite Host by a Miniature Laser Ablation/Ionization Mass Spectrometer. *Astrobiology* **15**, 669–682, https://doi.org/10.1089/ast.2015.1304 (2015).
44. Carlson, R. *et al.* Europa's surface composition. *Europa*, 283–327 (2009).
45. Johnson, R. *et al.* Composition and detection of Europa's sputter-induced atmosphere. *Europa, University of Arizona Press, Tucson*, 507–527 (2009).

Acknowledgements

Author contributions

A.R. designed and constructed the experimental set-up, while N.L. wrote its control software. I.L. and P.W. provided equipment and facilities for the construction of ORIGIN. P.M-G. and V.G. aided in sample holder cleaning and sample preparation. R.Li. provided molecular samples. M.T. and R.Lu. aided in setting up the experiments. N.L. and A.R. performed the experiments and data analysis. N.L. and A.R. wrote the main manuscript text with input from C.S.C. and P.E. for the astrobiological context. All authors reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.F.W.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

180

3.5 The scoring board of biosignatures and scientific outlook

Here, I summarize potential biosignatures that can be detected and characterized with an integrated LIMS/LDMS system (mass spectrometer combined with an ion source and a microscope). I will outline the broad applicability of the LIMS technology in identifying a wide range of biosignatures and speculate that LIMS has the potential to further improve its' measuring capabilities.

Searches for signs of extinct or extant life on Mars require sensitive techniques and modern analytical approaches capable of delivering high-quality chemical information. More generally, the chemical analysis on the surfaces of planetary bodies is a challenging endeavor that requires a set of specific parameters to comply with for onboard instrumentation. This applies to the quality of information gathered from a wide range of samples, the size and power consumption of the instrument, and the ability to operate within harsh conditions. Therefore, only a small subset of analytical methods can qualify as space applicable.
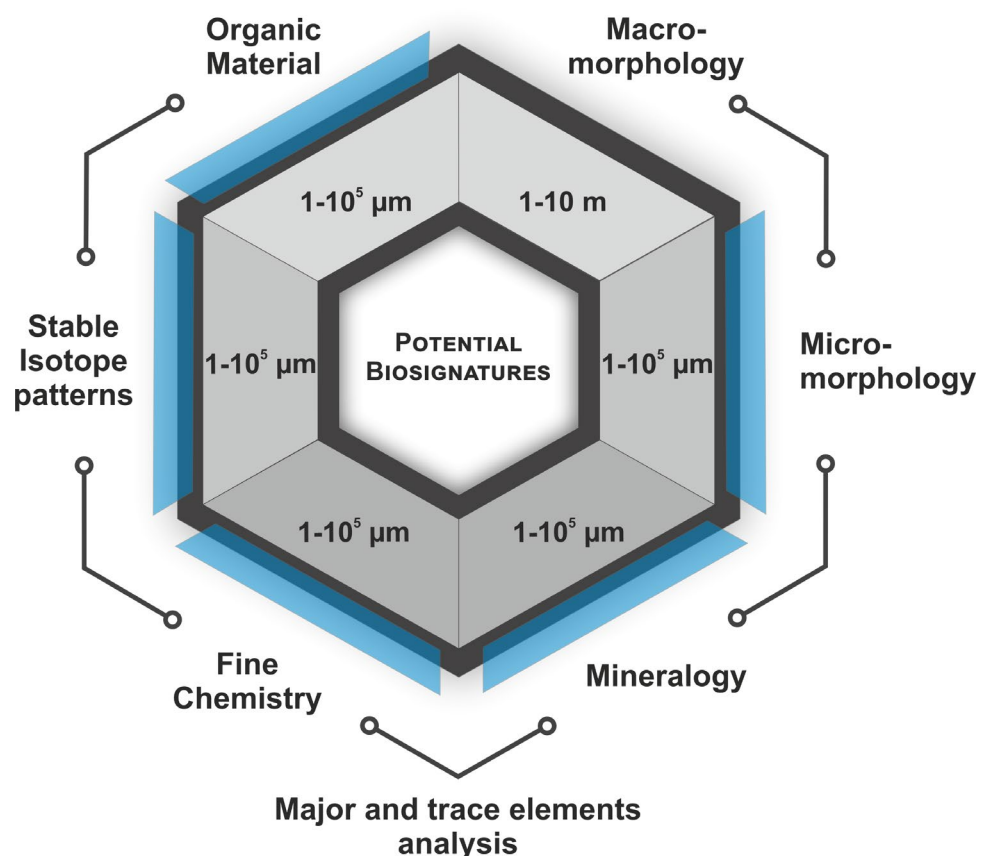
Figure 21. Applicability of the LIMS/LDMS combined with microscopy system in identi-fication of potential biosignatures. The blue sections indicate the applicability of the LIMS instrument to a given set of biosignatures. The figure is adapted from (Mustard et al., 2013).

The impact of the analytical technique can be scored on the generalized hexagon of potential biosignatures proposed by the Mars 2020 science definition team

(Mustard *et al.*, 2013) (see Fig.21). The binary scoring board presents six types of potential biosignatures and identifies the scale of applied methods. Broadly, the board can be divided into two big sections – morphological biosignatures and chemical biosignatures. The first group deals with different types of imagers that are required for the identification of - 1) macroscopic-scale structures (such as stromatolites), and 2) imaging of microscopic-scale structures (i.e., individual microfossils and bio-lamination sites). Although the first group primarily addresses cameras working in the visible range, it also includes methods like ground penetration radar and hyperspectral imaging.

The second large group of potential biosignatures goes into the detailed characterization of the chemistry of samples. Naturally, the first subgroup of the chemical biosignatures addresses the determination of mineralogical compositions. Considering the overall diversity of minerals, this subgroup places a hard constraint on analytical methods, meaning that method should be applicable to the wide range of minerals. The second subgroup of chemical biosignatures requires the determination of fine chemistries, i.e., rare earth element concentrations, which is also a very tough measure to deal with for most analytical methods. The third subgroup outlines the importance of measuring the isotopic fractionation factors. The systematic isotope fractionation on the microscale level, if measured correctly, can strongly indicate biological processing. However, among the available groups of potential biosignatures, identification of the isotopic ordering on the microscale places the harshest requirement for a measurement. And lastly, the fourth subgroup of chemical biosignatures presents the importance of molecular characterization of investigated materials.

Imaging of microscopic morphology

Interpretation of the geologic record, i.e., the origin of rocks found on planetary surfaces, is determined in the first place by the appearance of investigated rocks and formations. Therefore, it is important to image the sample, revealing its texture and morphology, which sometimes can be enough to identify minerals and major components building the rock. However, in most cases, it is crucial to locate the sample within the area of interest so that it can be analyzed with dedicated instruments. The spatial correlation of chemical, textural and mineralogical data can be very informative in understanding sample's origin and history of its alteration.

Accurately prepared samples (such as thin sections) are not likely to be part of upcoming space exploration programs due to the complexity of sample preparation procedures, at least in the nearest future. Nevertheless, there are developments in this field, e.g. (Dreyer *et al.*, 2013) presented the rock cutter that was developed as part of a broader effort to develop an in situ automated thin section (IS-ARTS) preparation instrument. However, until such instruments are in place, the investigated samples are expected to present rough surfaces. Typical investigation of the microscopic features present in the sample with LIMS system requires precise sample positioning under the instrument so that the analyzed spot will be located within the focal plane of the laser. Hence, understanding of the sample's complex topography is crucial for precise targeting of microscopic features.

The microscopes with sufficiently shallow depth of field can return resolved images only within the camera's focal plane. The images collected outside of the focus (in front of or behind the focal plane) can be significantly blurred. However, by collecting the set of images with different sample positions with respect to the focal point, it is possible to utilize focal stacking approach. Focal staking is an image

processing method that selectively compounds images and preserves only parts that are in focus. Therefore, the resulting image can have a significantly larger depth of field than any individual image collected at different distances. Currently, Mars Science Laboratory contains a Mars Hand Lens Imager (MAHLI) instrument (Edgett *et al.*, 2012) designed to collect images that can be later focus stacked. The MAHLI instrument uses a 2-megapixel color camera with an adjustable focus macro lens, demonstrating the utility of the method.
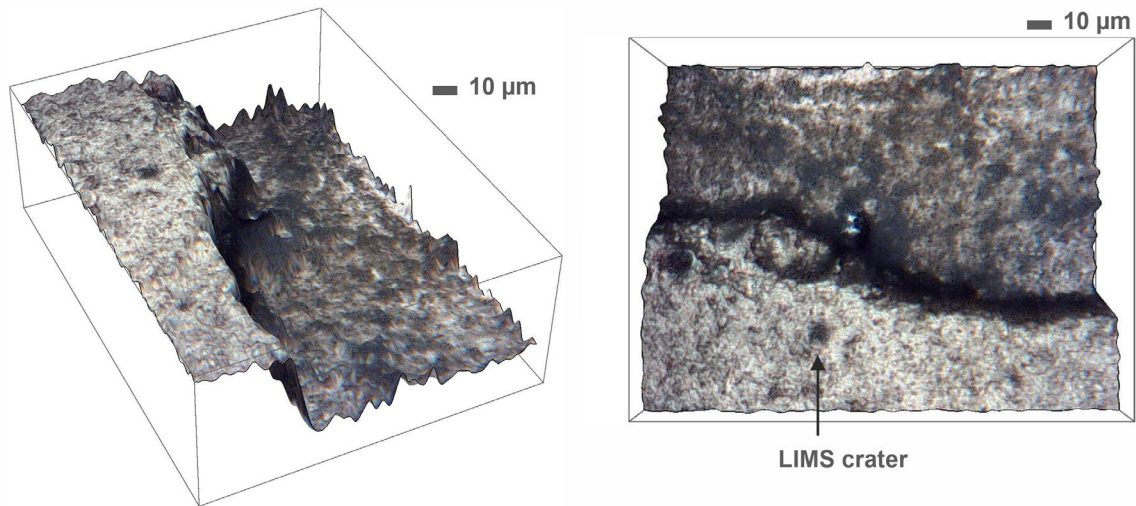
Figure 22. An exemplary 3D reconstruction using a focus stacking approach. The LIMS microscope pictures were used in the reconstruction of the sample's topography. The size of the step in the middle of the model is 50 μm, the scale bar is 10 μm.

Aside from improving the depth of field of images, it is possible to derive a depth map from parts of images that are preserved in focus. If the increment of the focal shift is known for the set of images, it is possible to derive a depth value for a given pixel, which in combination with the focus stacked image provides the 3D reconstruction of the sample under the microscope. The output depth reconstruction can be slightly noisy. Yet, the Gaussian smoothing kernel can be applied to the depth map, fixing the artefactual appearance of noisy pixels.

Figure 22 illustrates the 3D shape reconstruction using the focus stacking approach. The step-like sample topography is restored using the set of LIMS microscope pictures (16 images) collected at different focal positions. The sample represents a topographically challenging environment, which was successfully analyzed with LIMS system using the 3D elevation model (see the arrow in the lower right part of the figure). Currently used in the laboratory microscopy system, was developed to have a resolution of ~1 μm, which is desirable to resolve individual microbial cells. Detailed information on the microscopy system can be found in the publication by (Wiesendanger *et al.*, 2018). Need to mention that the LIMS microscope utilized in this thesis does not represent a space-ready prototype. However, a separate LIMS prototype integrated with a microscope-camera system was proposed by (Tulej *et al.*, 2014), which represents a compact solution, that can be placed onboard the rover or lander and showcase similar technical characteristics to the microscope utilized in the laboratory.

Mineralogy and fine chemistry

Identification and characterization of chemistries from heterogeneous samples with microscopic inclusions is the core competence of the LIMS technology. Accordingly, a significant part of the results section describes potential application pathways and efficient data analysis approaches that can directly link the identified chemistries with mineralogical assignments. As it was previously described, the LIMS provides very sensitive measurements on the microscale and captures local inhomogeneities of investigated materials. This might lead to the acquisition of different "flavors" of the same minerals, representing the full complexity of composition change. Moreover, the investigation of solid materials can change in the scale – from tens of millimeters down to the range of single micrometers. Thus, the LIMS technology can operate in the bulk-like regime, or if needed, provide an analysis in the microscopic regime, collecting data from a single position.

The inherent quality of the acquired mass spectra depends on the quality of ablation and ionization. If laser power fluences are not applied at the level of optimal ionization, the output stoichiometry and sensitivity can be significantly reduced. Therefore, laser power is an essential parameter to consider prior to the mass spectra acquisition campaigns. Detection of ppm (weight) level trace elements depends on the laser conditions and precise targeting of enriched inclusions. For example, the Ce and La enriched concentrations can be identified in specific minerals, i.e., monazite or xenotime. In the context of space exploration, where gravitational separation of minerals is not feasible, large-scale data collection campaigns can be performed, improving the probability of finding informative inclusions. The complexity of the data processing routines allows for automated data compressions to be made (see (Meyer *et al.*, 2017)) on the rover before transmission to Earth. For example, the extraction of peak intensities allows compressing the imaging data set consisting of 40'000 mass spectra down to a matrix with a comfortable size of ~70 Mbyte. Considering low connection bandwidth between Mars and the Earth, the data rate varies from about 500 bits per second to 32,000 bits per second, such compressions can be made onboard of the rover, and if needed, specific data can be retrieved separately.

As for data acquisition speeds, most of the current state delays come from the sample translation stage, which requires ~2-3 seconds to position the sample under the instrument accurately. The mass spectrum acquisition requires ~20 μs, and the speed of the histogramming of mass spectra is tied to the firing rate of the utilized laser (which is 1 kHz in the current ion source coupled to the mass spectrometer). Therefore, with simultaneous improvements in the speed of the stage positioning, laser-firing intervals, and the speed of the mass spectra archiving, it should be possible to reach the imaging speeds of the current commercial matrix-assisted desorption/ionization mass spectrometers (MALDI). For example, a contribution by (Zavalin *et al.*, 2015) reports MALDI TOF MS imaging in the continuous raster mode (stage continuously moves during data acquisition) at a rate of 40 pixels/second using the 2000 Hz laser repetition rate and 50 shots per pixel. With such speeds, imaging of the 2*2 mm area covered with 200 by 200 pixels will result in the acquisition times ~ 16 minutes. Of course, the imaging quality with such high acquisition rates will be significantly lower compared to step-wise imaging. However, the lower-quality images of the sample can be used to target interesting inclusions in a follow-up high-resolution imaging characterization campaign. Such discussion might seem premature due to the lack of in-situ sample preparation techniques (i.e., high-quality sample polishing). However, available improvements in the acquisition speed should not be discounted.

With the current rate of developments in electronics, engineering of precision stages, and laser technologies, it is reasonable to expect that in the 10-year horizon, such developments can be implemented for in-situ space instruments as well.

Stable isotope fractionation factors

Accurate determination of stable isotope patterns is of high interest in astrobiological missions since life fractionates a wide range of stable isotopes of elements involved in the metabolic pathways. For example, sulfur isotopes are known to significantly fractionate in a wide range of metabolic processes (Kaplan & Rittenberg, 1964). Carbon isotopes are also known to show significant isotopic heterogeneity (Blair *et al.*, 1985), thought, biogenic isotopic signatures can overlap, sometimes quite significantly with those of abiotic origin. Additionally, stable magnesium isotopes were shown to fractionate between different chlorophyll types (Moynier & Fujii, 2017), although not significantly. Therefore, identification of the isotope fractionation consistent with known metabolic pathways may indicate biological processing. However, in LIMS characterization of microscopic inclusions, the limited amount of sample and the instrument's limitations can affect the quality of isotope measurements. For example, the small molecular compounds formed in the plume can lead to unresolved isobaric interferences, limiting the peak abundance quantification capabilities. Another possible complication can come in the form of the limited SNR of the investigated element isotopes.

Although accurate isotope ratio calculations are generally hard, they are not out of the range of the instrument's capabilities. For instance, a contribution from (Tulej *et al.*, 2020) report measurements that demonstrate with the present LIMS instrument that element isotope ratios can be measured with accuracy and precision at the permille level, exemplified by the analysis of B, Mg, and Ni isotopes measured from the trevorite sample. Another contribution from (Wiesendanger *et al.*, 2019) reports the development of the spectral scoring approach for programmatic selection of highly resolved data. According to the quality criteria, the unsupervised selection of mass spectra recorded from the trevorite mineral illustrated that isotope ratios could be significantly improved. An alternative method improving the quality of mass spectra was proposed by (Tulej et al. 2018), addressing the spectral overlap of isobaric masses. The isobaric interferences that can commonly occur within specific mass ranges (mostly first oxides of abundant species) were reported to be significantly reduced by using the double-pulse ionization scheme. Overall, the accurate identification of isotope ratios is possible, however, the quality of the output data can differ.

Characterization of organic compounds using direct laser desorption.

The last group of remaining biosignatures in which LIMS/LDMS can operate represents the identification and characterization of pristine organic compounds without significant sample preprocessing and assistance of the matrix. As presented before, the Laser Desorption Mass Spectrometry operates in the significantly lower energy domain, compared to the Laser Ionization Mass Spectrometry. The direct desorption of molecules present on the sample's surface can be achieved by applying gentle laser irradiations. Hence, the fs-laser beams are not particularly suited for desorption studies since they provide more energetic radiation and may induce excessive fragmentation of parent molecules in comparison to the ns-laser radiations.

As it was reported in (Tulej et al., 2014), a space-ready LIMS/LDMS prototype was found to be capable of measuring amino acids, recording the identifiable fragmentation patterns. The LDMS measurements require applied laser energies to be below the ionization threshold, leading to the regime where ion desorption can occur. The increase in the applied laser energies can lead to the increased fragmentation of investigated molecules. In general, depending on the laser energies, a continuum between ion desorption and complete ionization can be observed. The desorption with increased fragmentation smoothly transitions to the quasidesorption and finally reaches the ionization regime, where elements can experience recombination processes. Therefore an accurate assessment of the ionization threshold should be done prior to the desorption measurements. The contribution by (Moreno-García *et al.*, 2016) reports the combination of the LIMS with LDMS in the characterization of organic constituents incorporated in the electrochemical Cu deposits upon growth. Moreover, several other groups report on the applicability of LDMS in the structural characterization of organic compounds. For example, contributions by (Selliez *et al.*, 2019; Selliez *et al.*, 2020) reveal the use of the Lab-CosmOrbitrap instrument in the identification of complex organic molecules (tholins). Additionally, a number of authors reported on the application of combined LDMS/LIMS systems for space applications (Arevalo Jr *et al.*, 2018; Goesmann *et al.*, 2017; Li *et al.*, 2017).

Concluding remarks

As demonstrated in the previous chapters and in this section, the integrated LIMS/LDMS system provides powerful analytical capabilities that cover 5 out of 6 groups of potential biosignatures (see Figure 21). Keeping in mind that it is possible to combine the LIMS/LDMS measurement results with complementary methods like Raman spectroscopy (Veneranda *et al.*, 2020) and utilize data-fusion approaches (Manrique-Martinez *et al.*, 2020), the output result can be synergetically strengthened. Bringing together scientific results from various scales (combining all available context knowledge) and various analytical methods (for example using the LIMS data together with Raman analysis (Rull et al., 2017 Astrobiology). can lead to descriptive and definitive scientific conclusions on the origin of investigated materials.
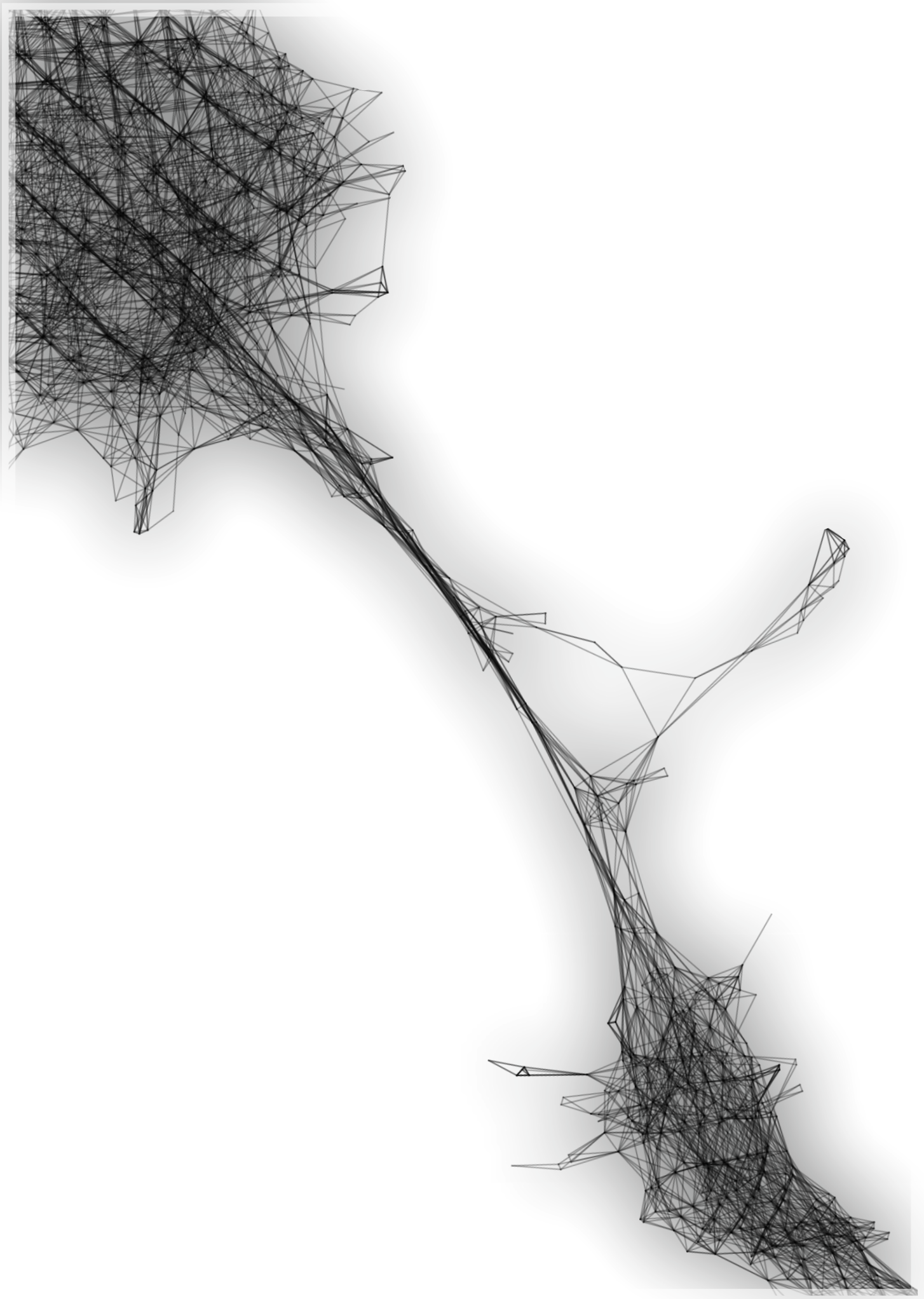
Conclusions

In summary, this thesis presents the results of analytical work completed on a large body of mass spectrometric observations. Overall, more than 200'000 mass spectra were recorded and analyzed from the Gunflint chert sample. The data recorded using the miniature time-of-flight mass spectrometer provides the crux of unambiguous identification of the bio-organic chemistries registered from the Gunflint chert sample. In this thesis, I argue that large-scale mass spectrometric observations allow covering the complete range of elemental variations, thus identifying bio-organic and inorganic chemistries of the Gunflint sample in a deterministic way, deriving the compositional structure solely from intensity values of mass peaks extracted from the mass spectra.

The thesis describes ways of improving the signal quality, including hardware modifications and computational approaches, and summarizes future development directions. The multi-color fs laser investigations revealed a gradual improvement in the quality of mass spectrometric data. The application of the far fs-UV wavelength was found to be preferential for the characterization of non-absorptive samples such as glasses and cherts, resulting in an equalized ionization efficiencies for heterogeneous materials. I also outline the different types of measurements that can be completed using the LIMS/LDMS system. The varying laser intensities applied to the sample can lead to the dual characterization of matter, providing the element and isotope composition and the structural characterization of bio-organic or inorganic molecules. In the context of space exploration, the broad applicability of the instrument makes it a favorable choice for future explorative missions, promising to enhance the scientific return.

Consequently, I've shown computational approaches and methods that can be applied to other mineralogically complex samples, underpinning the essential data patterns extraction process and improving the throughput of analysis. Application of graph-theoretic methods for the characterization of large mass spectrometric data uncovers natural clusters and groups of spectra. It also allows the formation of new secondary measures to be defined for specific compounds (i.e., carbon centralities for bio-organic inclusions) that arise from the structure of mass spectrometric networks. Proposed methods increase interpretability and speed of knowledge extraction, allowing the accurate identification of minerals and compounds. By mapping the complete shape of chemical compositions present in the Gunflint microfossils, it was shown that it is possible to build binary machine learning models that can be applied to identify bio-organic chemistries from other Precambrian cherts.

Overall, I believe that LIMS as a solid samples characterization technology best accounts for the needs of planetary space exploration and offers a promising solution for investigating chemical compositions of extraterrestrial bodies, whether it be an investigation of the lunar surface or Martian exploration program.

List of publications and proceedings:

1) **Lukmanov** R.A., Tulej M., Wiesendanger R., Riedo A., Grimaudo V., Ligterink N., de Koning C., Neubeck A., Wacey D., and Wurz P., "Multi-wavelength ablation/ionisation and mass spectrometric analysis of 1.88 Ga Gunflint Chert" Astrobiology (2021), accepted.

2) **Lukmanov** R. A., Tulej M., W. Ligterink N. F., De Koning C., Riedo A., Grimaudo V., Neubeck A., Wacey D., and Peter Wurz, "Chemical identification of microfossils from the 1.88 Ga Gunflint chert. Towards empirical biosignatures using LIMS" Chemometrics (2020)

3) **Lukmanov** R. A., Riedo A., Wacey D., Ligterink N. F. W., Grimaudo V., Tulej M., De Koning C., Neubeck A., and Peter Wurz, "On topological analysis of fs-LIMS data. Implications for in situ planetary mass spectrometry", Frontiers in Machine Learning and Artificial Intelligence (2021)

4) **Lukmanov** R. A., de Koning C., Schmidt P. K., Wacey D., Ligterink N. F.W., Gruchola S., Grimaudo V., Neubeck A., Riedo A., Tulej M., and Peter Wurz, "High-Resolution fs-LIMS 3D imaging and manifold learning reveal insight into chemical diversity of the Gunflint chert (1.88 Ga)", Frontiers in Astrobiology, (2021), submitted.

5) **Lukmanov** R. A., Tulej M., and Peter Wurz, "Aufspüren extraterrestrischer Lebenszeichen im Sonnensystem mittels lasergestützter Massenspektrometrie", VetsuisseNews, (2021)

6) Ligterink, N.F., Grimaudo, V., Moreno-García, P., **Lukmanov**, R., Tulej, M., Leya, I., Lindner, R., Wurz, P., Cockell, C.S., Ehrenfreund, P. and Riedo, A., 2020. ORIGIN: a novel and compact Laser Desorption–Mass Spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces. Scientific reports, 10(1), pp.1-10.

7) Wiesendanger, R., Grimaudo, V., Tulej, M., Riedo, A., **Lukmanov**, R., Ligterink, N., Fausch, R., Shea, H. and Wurz, P., 2019. The LMS-GT instrument–a new perspective for quantification with the LIMS-TOF measurement technique. Journal of analytical atomic spectrometry, 34(10), pp.2061-2073.

8) Grimaudo, V., Tulej, M., Riedo, A., **Lukmanov**, R., Ligterink, N.F., de Koning, C. and Wurz, P., 2020. UV post-ionization laser ablation ionization mass spectrometry for improved nm-depth profiling resolution on Cr/Ni reference standard. Rapid communications in mass spectrometry, 34(14), p.e8803.

9) Riedo, A., **Lukmanov**, R., Grimaudo, V., de Koning, C., Ligterink, N.F., Tulej, M. and Wurz, P., 2021. Improved plasma stoichiometry recorded by LIMS by using a double-pulse femtosecond laser ablation ion source. Rapid Communications in Mass Spectrometry, p.e9094.

10) Tulej, M., Neubeck, A., Riedo, A., **Lukmanov**, R., Grimaudo, V., Ligterink, N.F., Ivarsson, M., Bach, W., de Koning, C. and Wurz, P., 2020. Isotope abundance ratio measurements using femtosecond laser ablation ionization mass spectrometry. Journal of Mass Spectrometry, 55(12), p.e4660.

11) Cedeño López, A., Grimaudo, V., Riedo, A., Tulej, M., Wiesendanger, R., **Lukmanov**, R., Moreno-Garcia, P., Lortscher, E., Wurz, P. and Broekmann, P., 2019. Three-Dimensional Composition Analysis of SnAg Solder Bumps Using Ultraviolet Femtosecond Laser Ablation Ionization Mass Spectrometry. Analytical chemistry, 92(1), pp.1355-1362.

12) Tulej, M., **Lukmanov**, R., Grimaudo, V., Riedo, A., de Koning, C., Ligterink, N.F., Neubeck, A., Ivarsson, M., McMahon, S. and Wurz, P., 2020. Determination of the microscopic mineralogy of inclusion in an amygdaloidal pillow basalt by fs-LIMS. Journal of Analytical Atomic Spectrometry.

13) Wiesendanger, R., Tulej, M., Grimaudo, V., Cedeño-López, A., **Lukmanov**, R., Riedo, A. and Wurz, P., 2019. A method for improvement of mass resolution

and isotope accuracy for laser ablation time-of-flight mass spectrometers. Journal of chemometrics, 33(1), p.e3081.

14) López, A.C., Grimaudo, V., Riedo, A., Tulej, M., Wiesendanger, R., **Lukmanov**, R., Moreno-García, P., Lörtscher, E., Wurz, P. and Broekmann, P., 2021. Characterization of femtosecond laser ablation processes on as-deposited SnAg solder alloy using laser ablation ionization mass spectrometry. Spectrochimica Acta Part B: Atomic Spectroscopy, 180, p.106145.

15) Tulej, M., Ligterink, N.F., de Koning, C., Grimaudo, V., **Lukmanov**, R., Keresztes Schmidt, P., Riedo, A. and Wurz, P., 2021. Current Progress in Femtosecond Laser Ablation/Ionisation Time-of-Flight Mass Spectrometry. Applied Sciences, 11(6), p.2562.

16) Ligterink, N.F., Riedo, A., Tulej, M., **Lukmanov**, R., Grimaudo, V., de Koning, C., Wurz, P., Briois, C., Carrasco, N., Arevalo Jr, R. and Brinckerhoff, W.B., 2020. Detecting the elemental and molecular signatures of life: Laser-based mass spectrometry technologies. arXiv preprint arXiv:2009.02521.

17) de Koning C., Riedo A, Wiesendanger R., Grimaudo V., **Lukmanov** R., Ligterink N., Tulej M., and Wurz P., 2021. "Quantitative Elemental Analysis with the LMS-GT; a Next-Generation LIMS-TOF instrument", Journal of Mass Spectrometry, in press.

18) Wurz P., Tulej M., Riedo A., Grimaudo V., **Lukmanov** R., and N. Thomas, "Investigation of the Surface Composition by Laser Ablation/Ionisation Mass Spectrometry," IEEE Aerospace Conference Big Sky, MT, USA, (2021), in press.

19) Riedo A., de Koning C., Grimaudo V., Ligterink N.F.W., Schmidt P., **Lukmanov** R., Schwander L., Tulej M. and P. Wurz, "Detection of biomarkers using compact and sensitive laser ablation/desorption ionization mass spectrometry," proceedings of the 52nd Lunar and Planetary Science Conference, 15-19 March 2021, (2021), LPSC Abstract Nr. 1940.

20) Riedo A., Grimaudo V., Aerts J., **Lukmanov** R., Tulej M., Broekmann P., Lindner R., Wurz P., and Ehrenfreund P., "Laser Ablation Ionization Mass Spectrometry: A space prototype system for in situ Sulphur isotope fractionation analysis on planetary surfaces", Front. Astron. Space Sci. – Astrobiology.

List of conferences:

1) European cooperation in science and technology (COST) Action Grant – 400 EUR, 2018.
2) RED'18 – rencontres exobiologiques pour doctorants, summer school, 4-10 March 2018, Parc Ornithologique du Teich, France.
3) Moscow International Solar System Symposium, 11-15 October 2018, Moscow, Russia.
4) The European Planetary Science Congress, 19 to 24 September 2018, Berlin, Germany
5) "The chemical characterization of 1.88Ga microfossils using LIMS technology," invited talk at the Veterinary department of University of Bern, 23 October 2018, Bern, Switzerland.
6) Joint congress of The European Planetary Science (EPSC) and Division for Planetary Sciences (DPS) of the American Astronomical Society (ASS), 19 to 24 September 2019, Geneva
7) ESA LAB and Swiss space center (SSC) initiative – IGLUNA project. An international students competition on technology for space exploration. The P-14 project – LDMS for LIFE (building a laser desorption mass spectrometer for life detection with remote operation) 2019-2020, École polytechnique fédérale de Lausanne - EPFL, Lausanne, Switzerland.
8) Goldschmidt international conference on geochemistry, 18 to 23 August 2019, Barcelona
9) The European Geosciences Union (EGU) General Assembly 2020, 25 to 30 April 2020, (online) Vienna.

List of figures:

List of acronyms:

API - Application Programming Interface
BC – Betweenness Centrality
CHNOPS – Bio relevant elements
DBSCAN – Density Based Clustering
FA2 – ForceAtlas 2
fs- Femtosecond
Ga – Giga Annum
GD-MS – Glow-Discharge Mass Spectrometry
HDBSCAN - Hierarchical Density-Based Spatial Clustering
HR-LIMS-GT (LIMS-GT) – High-Resolution Laser Ionization Mass Spectrometer Gran Turismo
HV – High Voltage
IR – Infra Red
KDE – Kernel Density Estimate
LA-ICP-MS – Laser Ablation Inductively Coupled Plasma Mass Spectrometry
LDMS - Laser Desorption Mass Spectrometry
LE – Laplacian Eigenmaps
LIBS – Laser Ionization Breakdown Spectroscopy
LIMS – Laser Ionization Mass Spectrometry
LPF – Low pass filter
MC – Monte Carlo Integration
ML – Machine Learning
MSI – Mass Spectrometric Imaging
ns – Nanosecond
PC – Principal Component
PCA – Principal Components Analysis
ppm – part per million
Pyr-GC-MS – Pyrolysis-Gas Chromatography Mass Spectrometry
REE – Rare Earth Elements
RSC – Relative Sensitivity Coefficient
SGF – Savitzky–Golay filter
SIMS – Secondary Ion Mass Spectrometry
SNR – Signal to Noise Ration
SVD – Singular Value Decomposition
TDA -Topological Data Analysis
TOF – Time Of Flight
t-SNE – T-Distributed Stochastic Neighbors Embedding
UMAP – Uniform Manifold Approximation and Projection
UV – Ultraviolet
WMCA – Weighed Mass Correlation Analysis
k-NN – k Nearest Neighbor

Curriculum Vitae

PERSONAL INFORMATION:

Rustam A. Lukmanov / 08.09.1992 / Citizenship – Russian

CONTACT INFORMATION:

rustamlukmanov@gmail.com / rustamlukmanov@unibe.ch

EMPLOYMENT HISTORY:

- 2017 - now, PhD student in Physics, University of Bern
- 2015 –2017, Geophysicist, Russian Geological Research Institute (VSEGEI)

EDUCATION:

- 2017 - now, PhD student in Physics, University of Bern
- 2017 - Exchange semester - University of Tokyo
- 2017 - MSc in Geology and Geophyics - Saint Petersburg State University
- 2015 - BSc in Geology - Saint Petersburg State University

SOFTWARE AND PROGRAMMING SKILLS:

- Python/ Matlab / Mathematica
- Photoshop/ InDesign/ Illustrator/ CorelDraw
- Office/ Origin

LANGUAGE SKILLS:

- Russian - Mother tongue
- English - Fluent

INTERESTS:

- Running/ Hiking/ Swimming
- Illustration/ Reading

References

Akhmanov S. A., Vysloukh V. A., and Chirkin A. S. (1988) Optics of femtosecond laser pulses. *Moscow Izdatel Nauka*.

Arevalo Jr R., Ni Z., and Danell R. M. (2020) Mass spectrometry and planetary exploration: A brief review and future projection. *Journal of Mass Spectrometry*, 55: e4454.

Arevalo Jr R., Selliez L., Briois C., Carrasco N., Thirkell L., Cherville B., Colin F., Gaubicher B., Farcy B., Li X. and others. (2018) An Orbitrap-based laser desorption/ablation mass spectrometer designed for spaceflight. *Rapid Communications in Mass Spectrometry*, 32: 1875-1886.

Arvidson R. E., Squyres S. W., Bell J. F., Catalano J. G., Clark B. C., Crumpler L. S., De Souza P. A., Fairén A. G., Farrand W. H., Fox V. K. and others. (2014) Ancient aqueous environments at endeavour crater, mars. *Science*, 343: 1248097-1248097.

Barghoorn E. S., and Tyler S. A. (1965) Microorganisms from the gunflint chert. *Science*, 147: 563-575.

Bastian M., Heymann S., and Jacomy M. (2009) Gephi: an open source software for exploring and manipulating networks.Proceedings of the International AAAI Conference on Web and Social Media.

Becht E., McInnes L., Healy J., Dutertre C.-A., Kwok I. W. H., Ng L. G., Ginhoux F., and Newell E. W. (2019) Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, 37: 38-44.

Bednar J., Lead O. S. T., and Senior Solutions Architect A. (2017) Big Data Visualization With Datashader. *Anaconda, Inc.: Austin, TX, USA*: 12.

Behrmann J., Etmann C., Boskamp T., Casadonte R., Kriegsmann J., and Maaβ P. (2018) Deep learning for tumor classification in imaging mass spectrometry. *Bioinformatics*, 34: 1215-1223.

Belkin M., and Niyogi P. (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15: 1373-1396.

Biasotti S., Giorgi D., Spagnuolo M., and Falcidieno B. (2008) Reeb graphs for shape analysis and applications. *Theoretical computer science*, 392: 5-22.

Blondel V. D., Guillaume J.-L., Lambiotte R., and Lefebvre E. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008: P10008.

Boesl U. (2017) Time-of-flight mass spectrometry: introduction to the basics. *Mass spectrometry reviews*, 36: 86-109.

Böhm J. N., Berens P., and Kobak D. (2020) A Unifying Perspective on Neighbor Embeddings along the Attraction-Repulsion Spectrum. *arXiv preprint arXiv:2007.08902*.

Bonacich P. (2007) Some unique properties of eigenvector centrality. *Social networks*, 29: 555-564.

Borgatti S. P., Mehra A., Brass D. J., and Labianca G. (2009) Network analysis in the social sciences. *science*, 323: 892-895.

Brinckerhoff W. B., Managadze G. G., McEntire R. W., Cheng A. F., and Green W. J. (2000) Laser time-of-flight mass spectrometry for space. *Review of Scientific Instruments*, 71: 536-545.

Carlsson G. (2009) Topology and data. *Bulletin of the American Mathematical Society*, 46: 255-308.

Cloud P. E. (1965) Significance of the Gunflint (Precambrian) Microflora. *Science*, 148: 563-575.

Coifman R. R., and Lafon S. (2006) Diffusion maps. *Applied and Computational Harmonic Analysis*, 21: 5-30.

Cui Y., Moore J. F., Milasinovic S., Liu Y., Gordon R. J., and Hanley L. (2012) Depth profiling and imaging capabilities of an ultrashort pulse laser ablation time of flight mass spectrometer. *Review of Scientific Instruments*, 83: 093702.

Dahl D. A. (2000) SIMION for the personal computer in reflection. *International Journal of Mass Spectrometry*, 200: 3-25.

Deza M. M., and Deza E. (2009) Encyclopedia of distances. In: *Encyclopedia of distances*, Springer, pp 1-583.

Diaz-Papkovich A., Anderson-Trocmé L., and Gravel S. (2020) A review of UMAP in population genetics. *Journal of Human Genetics*: 1-7.

Dong W., Moses C., and Li K. (2011) Efficient k-nearest neighbor graph construction for generic similarity measures.Proceedings of the 20th international conference on World wide web.

Dührkop K., Nothias L.-F., Fleischauer M., Reher R., Ludwig M., Hoffmann M. A., Petras D., Gerwick W. H., Rousu J., and Dorrestein P. C. (2021) Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nature Biotechnology*, 39: 462-471.

Farahani F. V., Karwowski W., and Lighthall N. R. (2019) Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review. *frontiers in Neuroscience*, 13: 585.

Fassett C. I., and Head J. W. (2011) Sequence and timing of conditions on early Mars. *Icarus*, 211: 1204-1214.

Fenner N., and Daly N. (1966) Laser used for mass analysis. *Review of Scientific Instruments*, 37: 1068-1070.

Fruchterman T. M., and Reingold E. M. (1991) Graph drawing by force-directed placement. *Software: Practice and experience*, 21: 1129-1164.

Gemp I., McWilliams B., Vernade C., and Graepel T. (2020) Eigengame: Pca as a nash equilibrium. *arXiv preprint arXiv:2010.00554*.

Gil A., and Marco R. (2007) Systems Bioinformatics: An Engineering Case-Based Approach. Artech.

Goesmann F., Brinckerhoff W. B., Raulin F., Goetz W., Danell R. M., Getty S. A., Siljeström S., Mißbach H., Steininger H., Arevalo R. D. and others. (2017) The Mars Organic Molecule Analyzer (MOMA) Instrument: Characterization of Organic Material in Martian Sediments. *Astrobiology*, 17: 655-685.

Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., and Bengio Y. (2014) Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.

Han J., Kamber M., and Pei J. (2012) 2 - Getting to Know Your Data. In: *Data Mining (Third Edition)*. edited by J Han, M Kamber and J Peis, Morgan Kaufmann, Boston, pp 39-82.

Hillenkamp F., UNSöLD E., KAUFMANN R., and NITSCHE R. (1975) Laser microprobe mass analysis of organic materials. *Nature*, 256: 119-120.

Honig R., and Woolston J. (1963) Laser-induced emission of electrons, ions, and neutral atoms from solid surfaces. *Applied Physics Letters*, 2: 138-139.

Horvath S. (2011) Weighted network analysis: applications in genomics and systems biology. Springer Science & Business Media.

Hu Y. (2005) Efficient, high-quality force-directed graph drawing. *Mathematica journal*, 10: 37-71.

Jacomy M., Venturini T., Heymann S., and Bastian M. (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9: e98679-e98679.

Jolliffe I. T., and Cadima J. (2016) Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374: 20150202.

Kamada T., and Kawai S. (1989) An algorithm for drawing general undirected graphs. *Information processing letters*, 31: 7-15.

Kobak D., and Linderman G. C. (2019) UMAP does not preserve global structure any better than t-SNE when using the same initialization. *bioRxiv*: 2019.12.19.877522.

Kumar Kaliyar R. (2015) Graph databases: A survey.International Conference on Computing, Communication & Automation. IEEE.

Lepot K., Addad A., Knoll A. H., Wang J., Troadec D., Béché A., and Javaux E. J. (2017) Iron minerals within specific microfossil morphospecies of the 1.88 Ga Gunflint Formation. *Nature communications*.

Li X., Danell R. M., Pinnick V. T., Grubisic A., van Amerom F., Arevalo R. D., Getty S. A., Brinckerhoff W. B., Southard A. E., Gonnsen Z. D. and others. (2017) Mars Organic Molecule Analyzer (MOMA) laser desorption/ionization source design and performance characterization. *International Journal of Mass Spectrometry*, 422: 177-187.

Liang Z., Zhang S., Li X., Wang T., Huang Y., Hang W., Yang Z., Li J., and Tian Z. (2017) Tip-enhanced ablation and ionization mass spectrometry for nanoscale chemical analysis. *Science Advances*, 3: 1059-1059.

Ligterink N. F. W., Grimaudo V., Moreno-García P., Lukmanov R., Tulej M., Leya I., Lindner R., Wurz P., Cockell C. S., Ehrenfreund P. and others. (2020) ORIGIN: a novel and compact Laser Desorption – Mass Spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces. *Scientific Reports*, 10: 9641.

Lindenbaum O., Stanley III J. S., Wolf G., and Krishnaswamy S. (2018) Geometry-based data generation. *arXiv preprint arXiv:1802.04927*.

Makhzani A., Shlens J., Jaitly N., Goodfellow I., and Frey B. (2015) Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*.

Managadze G., Wurz P., Sagdeev R., Chumikov A., Tuley M., Yakovleva M., Managadze N., and Bondarenko A. (2010) Study of the main geochemical characteristics of Phobos' regolith using laser time-of-flight mass spectrometry. *Solar System Research*, 44: 376-384.

Manrique-Martinez J. A., Lopez-Reyes G., Alvarez-Perez A., Bozic T., Veneranda M., Sanz-Arranz A., Saiz J., Medina-Garcia J., and Rull-Perez F. (2020) Evaluation of multivariate analyses and data fusion between Raman and laser-induced breakdown spectroscopy in binary mixtures and its potential for solar system exploration. *Journal of Raman Spectroscopy*, 51: 1702-1717.

McInnes L., Healy J., and Astels S. (2017) hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2: 205.

McInnes L., Healy J., and Melville J. (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Meyer S., Riedo A., Neuland M. B., Tulej M., and Wurz P. (2017) Fully automatic and precise data analysis developed for time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, 52: 580-590.

Mohamed W., and Hamza A. B. (2012) Reeb graph path dissimilarity for 3D object matching and retrieval. *the visual computer*, 28: 305-318.

Navarro-González R., Navarro K. F., de la Rosa J., Iñiguez E., Molina P., Miranda L. D., Morales P., Cienfuegos E., Coll P., Raulin F. and others. (2006) The limitations on organic detection in Mars-like soils by thermal volatilization-gas chromatography-MS and their implications for the Viking results. *Proceedings of the National Academy of Sciences of the United States of America*, 103: 16089-16094.

Newman M. E., and Girvan M. (2004) Finding and evaluating community structure in networks. *Physical review E*, 69: 026113.

Nitsche R., Kaufmann R., Hillenkamp F., Unsöld E., Vogt H., and Wechsung R. (1978) Mass spectrometric analysis of laser induced microplasmas from organic samples. *Israel Journal of Chemistry*, 17: 181-184.

Nolet C. J., Lafargue V., Raff E., Nanditale T., Oates T., Zedlewski J., and Patterson J. (2020) Bringing UMAP Closer to the Speed of Light with GPU Acceleration. *arXiv preprint arXiv:2008.00325*.

Onstott T. C., Ehlmann B. L., Sapers H., Coleman M., Ivarsson M., Marlow J. J., Neubeck A., and Niles P. (2019) Paleo-Rock-Hosted Life on Earth and the Search on Mars: A Review and Strategy for Exploration. *Astrobiology*, 19: 1230-1262.

Orosei R., Lauro S. E., Pettinelli E., Cicchetti A., Coradini M., Cosciotti B., Di Paolo F., Flamini E., Mattei E., Pajola M. and others. (2018) Radar evidence of subglacial liquid water on Mars. *Science*, 361: 490-493.

Pearson K. (1901) LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2: 559-572.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., and Dubourg V. (2011) Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12: 2825-2830.

Press W. H., and Teukolsky S. A. (1990) Savitzky-Golay smoothing filters. *Computers in Physics*, 4: 669-672.

Riedo A., Lukmanov R., Grimaudo V., de Koning C., Ligterink N. F., Tulej M., and Wurz P. (2021) Improved plasma stoichiometry recorded by LIMS by using a double-pulse femtosecond laser ablation ion source. *Rapid Communications in Mass Spectrometry*: e9094.

Riedo A., Meyer S., Heredia B., Neuland M. B., Bieler A., Tulej M., Leya I., Iakovleva M., Mezger K., and Wurz P. (2013a) Highly accurate isotope composition measurements by a miniature laser ablation mass spectrometer designed for in situ investigations on planetary surfaces. *Planetary and Space Science*, 87: 1-13.

Riedo A., Neuland M., Meyer S., Tulej M., and Wurz P. (2013b) Coupling of LMS with a fs-laser ablation ion source: Elemental and isotope composition measurements. *Journal of Analytical Atomic Spectrometry*, 28: 1256-1269.

Riedo A., Tulej M., Rohner U., and Wurz P. (2017) High-speed microstrip multi-anode multichannel plate detector system. *Review of Scientific Instruments*, 88: 045114-045114.

Rohner U., Whitby J. A., and Wurz P. (2003) A miniature laser ablation time-of-flight mass spectrometer for in situ planetary exploration. *Measurement Science and Technology*, 14: 2159.

Sainburg T., McInnes L., and Gentner T. Q. (2020) Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning. *arXiv preprint arXiv:2009.12981*.

Sauer S., Freiwald A., Maier T., Kube M., Reinhardt R., Kostrzewa M., and Geider K. (2008) Classification and identification of bacteria by mass spectrometry and computational analysis. *PloS one*, 3: e2843.

Sauer S., and Kliem M. (2010) Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews Microbiology*, 8: 74-82.

Schaller R. R. (1997) Moore's law: past, present and future. *IEEE spectrum*, 34: 52-59.

Singh G., Mémoli F., and Carlsson G. E. (2007) Topological methods for the analysis of high dimensional data sets and 3d object recognition. *SPBG*, 91: 100.

Traag V. A., Waltman L., and van Eck N. J. (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9: 5233.

Tulej M., Neubeck A., Riedo A., Lukmanov R., Grimaudo V., Ligterink N. F. W., Ivarsson M., Bach W., de Koning C., and Wurz P. (2020) Isotope abundance ratio measurements using femtosecond laser ablation ionization mass spectrometry. *Journal of Mass Spectrometry*, 55: e4660.

Tulej M., Riedo A., Neuland M. B., Meyer S., Wurz P., Thomas N., Grimaudo V., Moreno-García P., Broekmann P., Neubeck A. and others. (2014) CAMAM: A miniature laser ablation ionisation mass spectrometer and microscope-camera system for in situ investigation of the composition and morphology of extraterrestrial materials. *Geostandards and Geoanalytical Research*, 38: 441-466.

Tulej M., Wiesendanger R., Riedo A., Knopp G., and Wurz P. (2018) Mass spectrometric analysis of the Mg plasma produced by double-pulse femtosecond laser irradiation. *Journal of Analytical Atomic Spectrometry*, 33: 1292-1303.

Tusi B. K., Wolock S. L., Weinreb C., Hwang Y., Hidalgo D., Zilionis R., Waisman A., Huh J. R., Klein A. M., and Socolovsky M. (2018) Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*, 555: 54-60.

Vago J. L., Coates A. J., Jaumann R., Korablev O., Rie Ciarletti V., Mitrofanov I., Josset J.-L., De Sanctis M. C., Bibring J.-P., Rull F. and others. (2017) Habitability on Early Mars and the Search for Biosignatures with the ExoMars Rover. *Astrobiology*, 17: 471-510.

Van der Maaten L., and Hinton G. (2008) Visualizing data using t-SNE. *Journal of machine learning research*, 9.

Wacey D., McLoughlin N., Kilburn M. R., Saunders M., Cliff J. B., Kong C., Barley M. E., and Brasier M. D. (2013) Nanoscale analysis of pyritized microfossils reveals differential heterotrophic consumption in the ~1.9-Ga Gunflint chert. *PNAS*, 110: 1053-1060.

Wagner D. E., Weinreb C., Collins Z. M., Briggs J. A., Megason S. G., and Klein A. M. (2018) Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, 360: 981-987.

Wattenberg M., Viégas F., and Johnson I. (2016) How to use t-SNE effectively. *Distill*, 1: e2.

Weinreb C., Wolock S., and Klein A. M. (2018) SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*, 34: 1246-1248.

Westall F., Foucher F., Bost N., Bertrand M., Loizeau D., Vago J. L., Kminek G., Gaboyer F., Campbell K. A., Bréhéret J.-G. and others. (2015) Biosignatures on Mars: What, Where, and How? Implications for the Search for Martian Life. *Astrobiology*, 15: 998-1029.

Wiesendanger R., Grimaudo V., Tulej M., Riedo A., Lukmanov R., Ligterink N., Fausch R., Shea H., and Wurz P. (2019) The LMS-GT instrument–a new perspective for quantification with the LIMS-TOF measurement technique. *Journal of analytical atomic spectrometry*, 34: 2061-2073.

Wiesendanger R., Tulej M., Riedo A., Frey S., Shea H., and Wurz P. (2017) Improved detection sensitivity for heavy trace elements using a miniature laser ablation ionisation mass spectrometer. *Journal of Analytical Atomic Spectrometry*, 32: 2182-2188.

Wiesendanger R., Wacey D., Tulej M., Neubeck A., Ivarsson M., Grimaudo V., Moreno-García P., Cedeño-López A., Riedo A., Wurz P. and others. (2018) Chemical and Optical Identification of Micrometer-Sized 1.9 Billion-Year-Old Fossils by Combining a Miniature Laser Ablation Ionization Mass Spectrometry System with an Optical Microscope. *Astrobiology*, 18: 1071-1080.

Wolff M., and Stephens W. (1953) A pulsed mass spectrometer with time dispersion. *Review of Scientific Instruments*, 24: 616-617.

Wu C.-H. J., and Irwin J. D. (2016) Introduction to computer networks and cybersecurity. CRC Press.

Wurz P., Abplanalp D., Tulej M., Iakovleva M., Fernandes V. A., Chumikov A., and Managadze G. G. (2012) Mass spectrometric analysis in planetary science: Investigation of the surface and the atmosphere. *Solar System Research*, 46: 408-422.

Wurz P., Riedo A., Tulej M., Grimaudo V., and Thomas N. (2020) Investigation of the Surface Composition by Laser Ablation/Ionisation Mass Spectrometry. *LPI Contributions*, 2241: 5061.

Wurz P., Whitby J., and Managadze G. (2009) Laser mass spectrometry in planetary science.AIP Conference Proceedings. American Institute of Physics.

Zavalin A., Yang J., Hayden K., Vestal M., and Caprioli R. M. (2015) Tissue protein imaging at 1 μm laser spot diameter for high spatial resolution and high imaging speed using transmission geometry MALDI TOF MS. *Analytical and Bioanalytical Chemistry*, 407: 2337-2342.

Appendix

11. 1 Supplementary Information for
Chemical identification of microfossils from the 1.88 Ga Gunflint chert.
Towards empirical biosignatures using LIMS - Chemometrics

Rustam A. Lukmanov[1*], Marek Tulej[1], Niels F. W. Ligterink[1], Coenraad De Koning[1], Andreas Riedo[1], Valentine Grimaudo[1], Anna Neubeck[2], David Wacey[3], Peter Wurz[1]

[1]Space Research and Planetary Sciences, University of Bern, Bern, Switzerland.

[2]Department of Earth Sciences, Uppsala University, Uppsala, Sweden.

[3]Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, Perth, Australia.

*Rustam A. Lukmanov
Email: rustam.lukmanov@space.unibe.ch

Scope of the study

The current state of in-situ space exploration requires new measurement techniques and novel analytical approaches to identify extinct or extant life on Mars. In this contribution, we demonstrate that laser (fs UV) ablation ionization time-of-flight mass spectrometry (UVLA-TOF-MS or LIMS) provides a highly sensitive (ppm level), high spatial resolution (4-10 μm), and high-speed analysis (10^{-3} s/ single laser shot spectrum) analysis of micrometer-sized objects, thus providing the necessary tool for in-situ exploration of the Red Planet and beyond. The LIMS measurements performed on the 1.88 Ga Gunflint sample reveals the capacity to identify the multielement composition of individual microfossils and to perform mass-spectrometric imaging, providing μm-spatially-resolved chemical analysis. Large-scale spectral sampling (>60k spectra) from the Gunflint microfossils and host mineral identified empirical biosignatures, defined for Gunflint-like microfossils embedded in silicified matrices. We also demonstrate that it is possible to create a machine learning solution that might aid in-situ analysis of samples on planetary surfaces. Additionally, we implemented a weighted mass correlation network analysis into the LIMS data processing routine to aid the identification of micro- to nano-scale mineralization present within the bodies of microfossils.
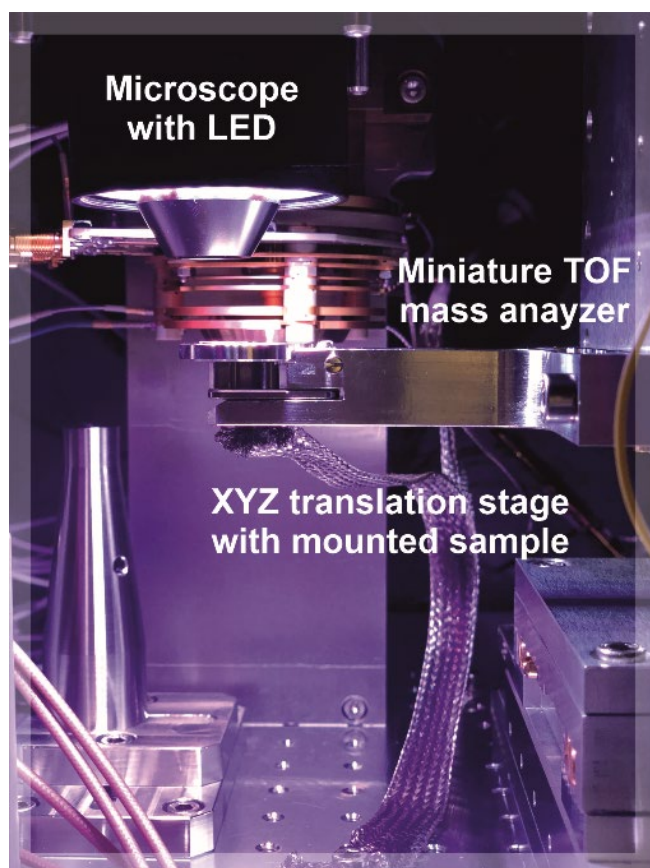
LIMS microprobe

A miniature time-of-flight mass spectrometer developed at the University of Bern has small dimensions: Ø 60 mm × 160 mm, which makes it suitable for space exploration programs as part of the lander or rover (1-3). In the current laboratory setup, a mass spectrometer is accompanied by an integrated microscopy system (4) and a femtosecond UV-258 nm ionization source. The instrument is designed to have spatial molecular (5, 6), element, and isotope (7, 8) mapping capabilities of solid samples. A picture of the instrument located inside of the vacuum chamber is shown in fig. S1. Multicolor LED vacuum compatible illumination system could be seen above the XYZ translation stage. On the backside of the picture, a miniature LIMS mass analyzer could be noted.

Fig.S1. Integrated LIMS system with microscope and XYZ translation stage inside of the vacuum chamber.
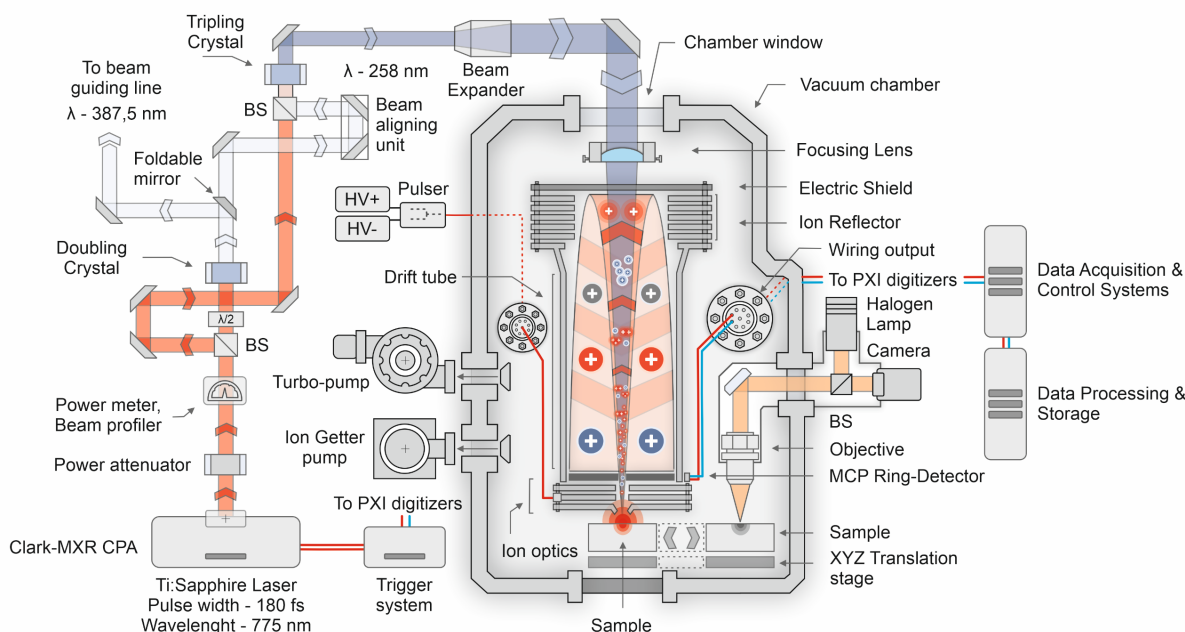
Tripling Crystal
To beam guiding line λ - 387,5 nm
λ - 258 nm
Beam Expander
Chamber window
Vacuum chamber
BS
Beam aligning unit
Foldable mirror
Focusing Lens
HV+ Pulser
HV-
Electric Shield
Ion Reflector
Wiring output
Doubling Crystal
To PXI digitizers
Data Acquisition & Control Systems
Drift tube
Halogen Lamp
Camera
λ/2
BS
Turbo-pump
Power meter, Beam profiler
Ion Getter pump
BS
Objective
Data Processing & Storage
Power attenuator
MCP Ring-Detector
To PXI digitizers
Sample
Ion optics
XYZ Translation stage
Clark-MXR CPA
Ti:Sapphire Laser Pulse width - 180 fs Wavelenght - 775 nm
Trigger system
Sample

Fig. S2. Schematics of femtosecond laser light transmission and subsequent mass spectrometric analysis of ablated positive ions with LIMS. UV-258 nm ablation occurs by guiding the appropriate wavelength to the beam guiding line and periscope system. (*Astrobiology, 2020, Lukmanov et al. – in review*).

Schematics of light generation and transmission to the sample surface illustrated in figure S2. The sample positioning and control of the laser system is achieved by using in-house made software. The vacuum inside of the chamber is maintained around 10^{-7} mBar, allowing a direct transfer of generated ions to the mass analyzer. A combination of the molecular turbo-pump and an ion getter pump is used to maintain the nominal pressure. The fs UV-258 nm beam is focused by a doublet lens, guided through the mass spectrometer, and projected on the surface of the sample. The entrance of the mass spectrometer is located in close proximity to the sample ($100 \sim 200$ μm), which allows direct coupling of ions with the mass spectrometer. The focused laser beam forms an ablation spot on the surface of the sample (with a diameter of about 4–10 μm for UV-258 nm, see fig. S3). Ions from the ablated plume are confined, focused, accelerated, and transferred to the field-free area of the mass spectrometer (drift tube), where ions experience mass per charge separation. After passing through an ion mirror, ions hit the chevron type ring-MCP detector. From the initial ion cloud, only positively charged species enter the mass analyzer, as it is shown in fig. S2. The time-of-flight spectrum is obtained by collecting current from the atomic and molecular ion signals, amplified by the microchannel plate detector (MCP), and recorded with a digitizer as a function of signal arrival time. The output current is measured as electrons*ns-1. The time-of-flight spectrum is converted to a mass spectrum with a simple fit equation $m=a(t+b)^2$, where m is the mass unit, t the arrival time, and a and b instrument dependent fit constants.

In the current setup, our system utilizes a multi-anode MCP detector system coupled with a fast acquisition system (9). Such modification of the detector allows collecting twice as much data with a reduced probability of detector saturation, thus providing better detection limits and record more stoichiometric signal (current from different parts of the detector digitized separately). Data acquisition occurs by the simultaneous recording of the signal from two anode rings on the MCP detector. PXI3 and PXI5 high-resolution digitizers from Agilent technologies provide 3.2 GS*s^{-1} sampling rate resulting in a ~ 0.3 ns sampling resolution. Control of ion-optical voltages, as well as translation stage positioning, are achieved with in-house made software. All collected data from the control PC are saved on the storage PC for post-processing and data analysis. All internal network connections are made with optic fiber cables to increase data transmission bandwidth.

LIMS data acquisition – MSI and depth profiling

The details about our system and data acquisition routine are published in a number of previous publications - (1, 3, 10-13). Only a brief description will be given here. A broad overview of the LIMS as an analytical method recently has been published and could be found here - (14). Our instrument represents a time-of-flight variation of LIMS.

Our LIMS instrument provides a point-analysis by recording a time-of-flight spectrum of positively charged species present in the ablation plume. By combining multiple points from the surface of the sample we could derive an image of the given mass by extracting peak intensities from specific time-of-flight windows.
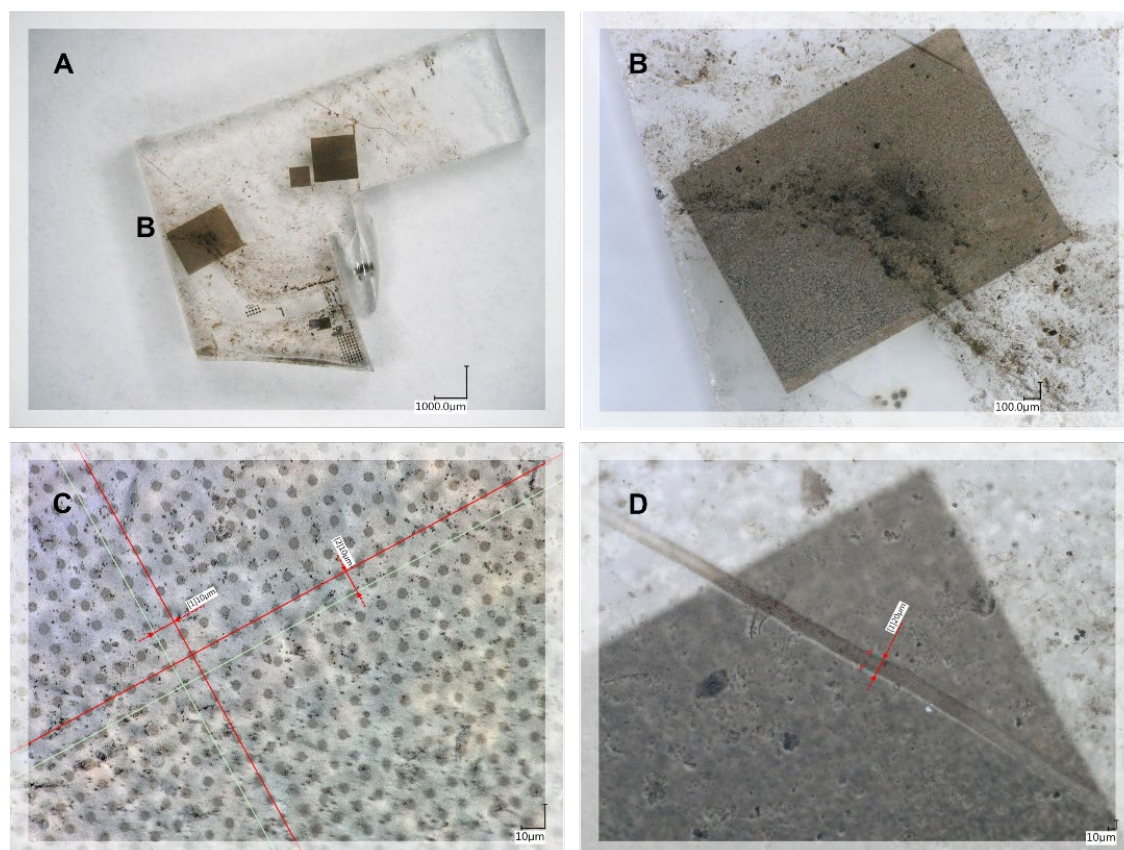
Fig S3. Optical images of the Gunflint chert thin section after imaging campaign. A) Panoramic picture of the Gunflint thin section sample with a designated location of the area chosen for the imaging. B) Image of the area after the imaging campaign. C) Close-up image of the single craters acquired during the mapping. D) Close up image of the corner of MSI area.

Fig. S3-A shows the Gunflint sample after the imaging campaign. The diameters of the craters are shown in figure S3-C. Most of the craters, formed with a fs UV laser, have a diameter of 4-5 micrometers. As it could be seen from figures S3-C and S4, the chosen gap (10 µm) between craters could be doubled without any analytical interference (overlapping of the craters). With our current setup, it is possible to improve the gap between craters down to 1 micrometer by programmatically forcing the stage to move to the absolute positions, so it is possible to reach a LIMS resolution that comes close to the resolution of our internal microscope (4) (1 micrometer). However, high-resolution imaging will increase the output of the data by the same factor. An example of the craters acquired during the MSI campaign could be seen in figs. S4 and S5. As it could be seen from figs. S4 and S5, the diameter of individual craters are comparable with sizes of individual microfossils. The diameter of individual craters is roughly equated to 4 micrometers (see fig. S4), which is well suited for the characterization of individual microfossils. In fig S4, one could note the

presence of spherical microfossil, which is highlighted with a black arrow. In figure S5, another type of lenticular microfossils is highlighted with the black arrow in comparison to the size of the individual craters. One could also note that most microfossils are embedded within the silicified matrix, as pictured in figure 1 in the main text.
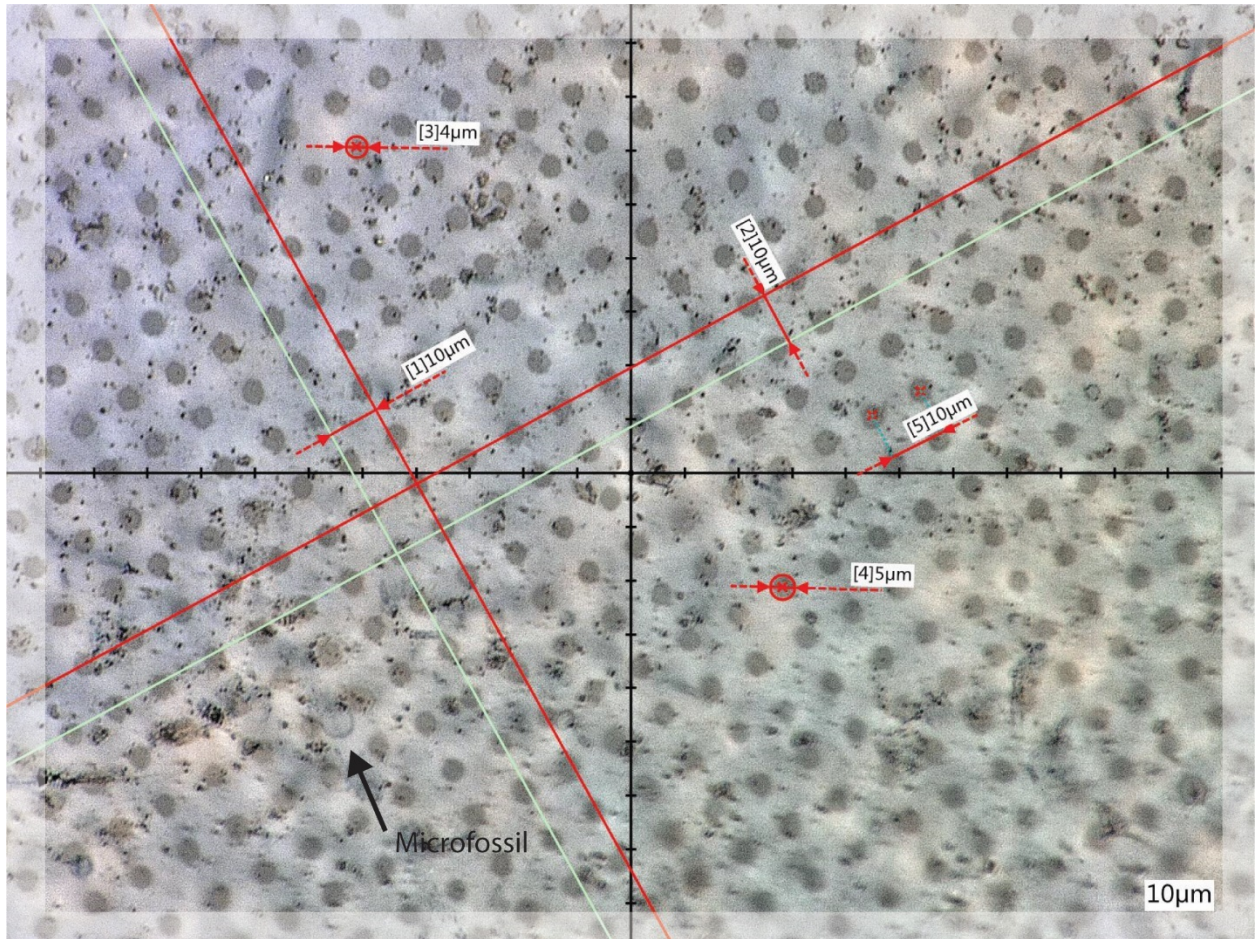
Fig S4. Close-up image of the single craters acquired from the Gunflint sample after the mapping campaign. The black arrow indicates location of the single microfossil embedded inside the host mineral. Red arrows indicate the diameter of ablation spots (4-5 μm). Scale bar 10 μm.
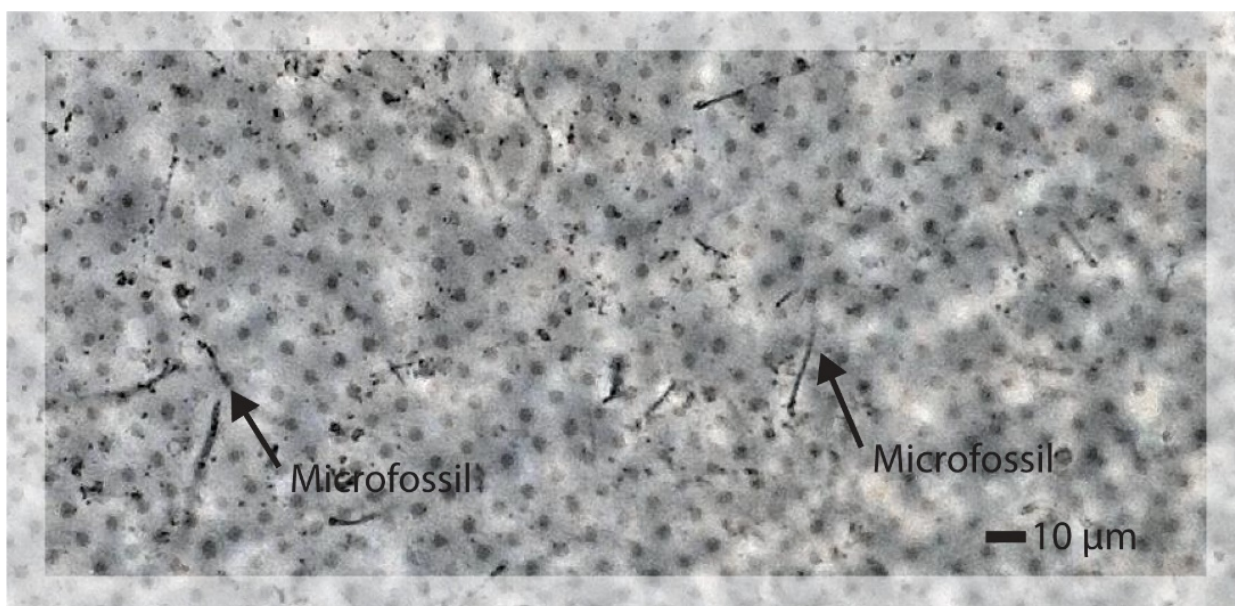
Fig S5. Close-up image of the single fs UV-258 nm craters and individual microfossils from the Gunflint chert sample. Note comparable sizes of LIMS ablation craters and lenticular microfossils highlighted with black arrows. Scale bar – 10 micrometers.
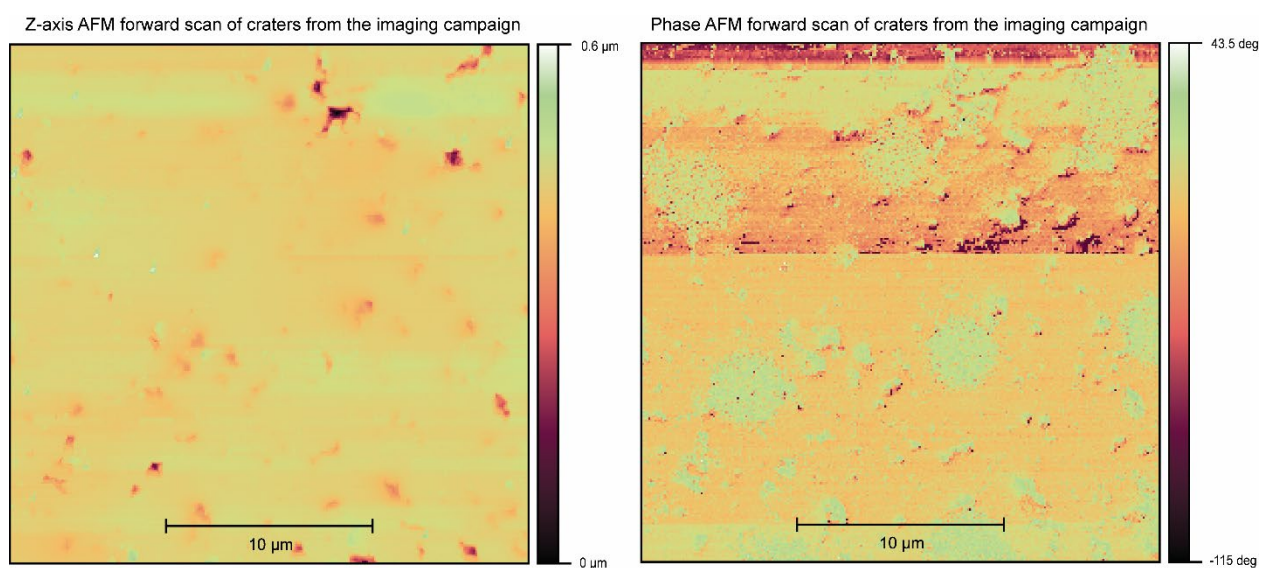
Fig S6. AFM images of the craters acquired from the imaging campaign. Left - AFM measurements of the z-axis. Right – Phase modification scan over the same area as in the picture on the left. See fig.S4 to compare with images of the same craters acquired with an optical microscope.

The imaging campaign was performed as a sequence of two separate scans over the same area (figs. S3-A and S3-B, S4, S5). The imaging results presented in the main text in fig.2 are calculated from the first scan, which consists of 200 individual fs laser shots applied to the single location on the surface of the sample. After the first scan, we applied additional 2200 single shots to gain even more statistics over the same craters. Due to the failure of the power attenuators controlling the laser power output, this scan was discarded from the analysis. The final craters shown in figs. S3, S4, and S5 are made with 2400 single laser shots (sum of two campaigns) and indicate a very shallow ablation of material from the quartz matrix of the sample. For example, fig. S6 (left) presents an AFM z-scan over the same area, where we could not measure any craters from the imaging campaign. Craters from the same area could be visualized only in the phase imaging mode (see fig. S6 - right), which indicates that 2400 applied laser shots only modified the surface roughness without actually creating any measurable craters. Laser pulses with an energy of 0.36 µJ per pulse (transmission corrected) were applied to the surface to ablate and ionize material. As it could be seen from figs.S6,7,8,9, and 12, applied laser pulse energies were enough to register an intense signal from a very little amount of analyte material. Our previous studies show that LIMS measurements could provide a femto to picogram level analysis in a single laser shot mode (12), which makes it very sensitive to the local enrichments and surface chemistry. An example of a spectrum acquired from the surface could be found in fig S12, and an example of spectra from the Gunflint subsurface provided in fig 4 in the main text.

High sensitivity, combined with the high-throughput acquisition, makes LIMS an attractive analytical method for the analysis of solids on planetary surfaces. The LIMS delivers a capability to probe microscopic inclusions and even to resolve some isobaric interferences by utilizing depth profiling. For example, hydrocarbons present at the surface spectra are partially eliminated by the acquisition of spectra recorded from the subsurface of the sample, where diameter of the analytical spot is rapidly decreasing (see fig.S11), thus providing even finer analysis of embedded inclusions. However, even in the case of the presence of isobaric input, it is possible to extract a significant amount of information. Figures S7, S8, and S9 illustrate the distribution of ^{11}B, ^{40}Ca, ^{39}K isotopes. Boron distribution pattern reflects the sedimentation pattern (from the upper-right corner to the lower-left corner) and shows a decrease of the boron supply after the appearance of the microfossils. Such dependencies might reflect the dynamics of the chemical system and represent a secondary biological phenomenon.
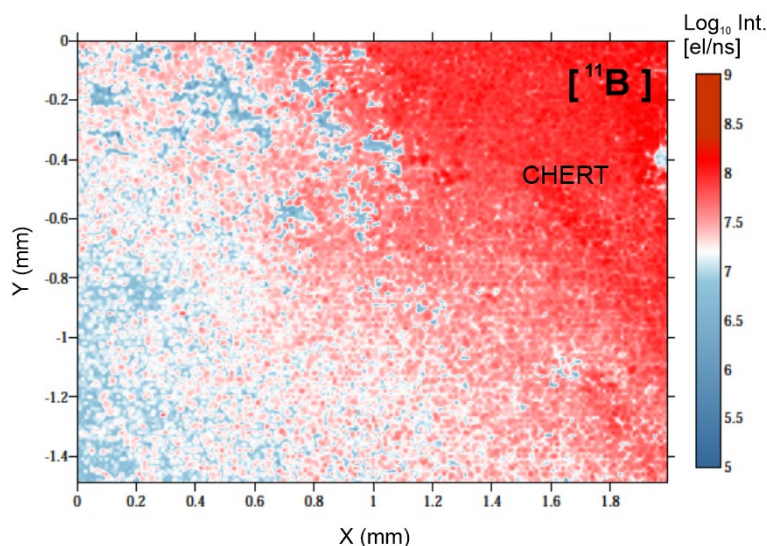

Fig S7. MSI of ^{11}B. The spatial distribution of B indicates a gradual decrease of B supply after the appearance of the microfossil community.

Individual crater from the depth profiling campaign

We studied the morphology and topography of the fs-UV-258 nm craters from the depth profiling campaign using atomic force microscopy (AFM). Results of the AFM measurements of the crater acquired from the host area shown in fig S8. As it could be seen from the figure, crater is relatively shallow (4 micrometers), and has a 10-micrometer wide diameter.
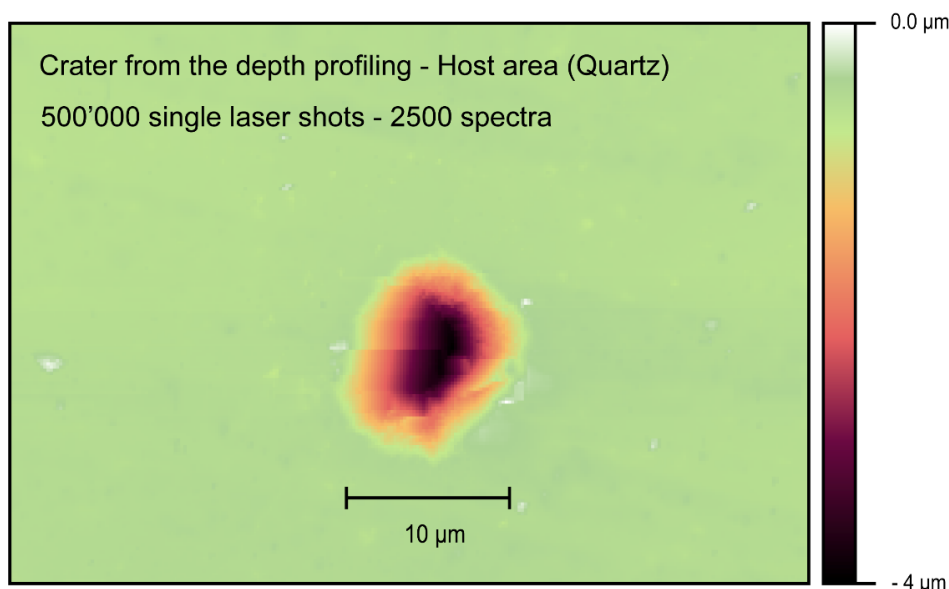

Fig S8. AFM z-scan image of the single crater acquired from the depth profiling campaign. To compare AFM measurements with an optical image of the same crater, see fig. 3A in the main text.
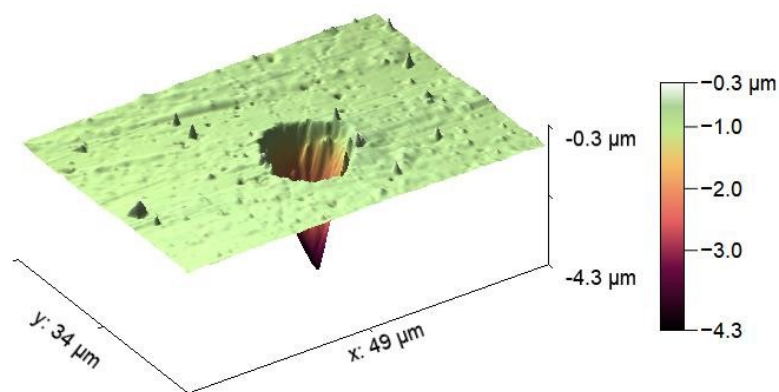
Fig S9. 3D AFM z-scan image of the single crater acquired from the depth profiling campaign. To compare AFM measurements with an optical image of the same crater, see fig. 3A in the main text and fig S8 (planar view).

In total, 500'000 single laser shots have been applied to the same location to aquire a crater presented in these figures. The crater measured with an AFM probe reveals the presence of conical crater with a 10 µm diameter. The depth of the craters most likly exceeds 4 µm and goes as deep as 20-28 µm. The estimated thickness of the thin-section is 30 µm, and according to the analysis of the LIMS depth profiles, the laser never went through the sample.

Identification of the true depths of the craters with an AFM probe was not possible since tip of the AFM probe was found to be too thick to resolve the true depth of the craters. Laser interferometry

measurements performed on the same craters also have not provided a satisfying estimate of the crater's depths since the Gunflint samples surface is not reflective enough and often produces artefactual measurements (interferometers laser passes through the sample and measure the topography of the bottom of the sample) for a green and red lasers provided with an interferometer.

Among the advantages of LIMS imaging is the possibility to image the distribution of virtually every element in the periodic table in a single run, since mass analyzer cover masses up to 1000 m/q and does not discriminate masses. However, as it was mentioned before, surface chemistry might affect MSI results. Fig. S10 depicts the spectra registered from the surface of the sample. Due to the isobaric input of CH chains, abundances of minor elements are masked. The spectrum depicted in fig.S10 reflects the chemical composition of the chert (quartz), microfossil (kerogen), metallic inclusion associated with the microfossil (intense Cr, Cu), and modern organic material present at the surface. From our observations, the composition of H/C to O/C ratios of the surface layers mostly corresponds to the lipidic profile. Thus, providing a piece of evidence that the contamination is registered from the fingerprint. However, other potential sources of contamination are possible, i.e., organic glue residue from the thin-section preparation process. Additionally, it is possible to improve the future imaging results by removing the uppermost layer of the surface by using the Ar+ ion gun implemented in our LIMS lab.
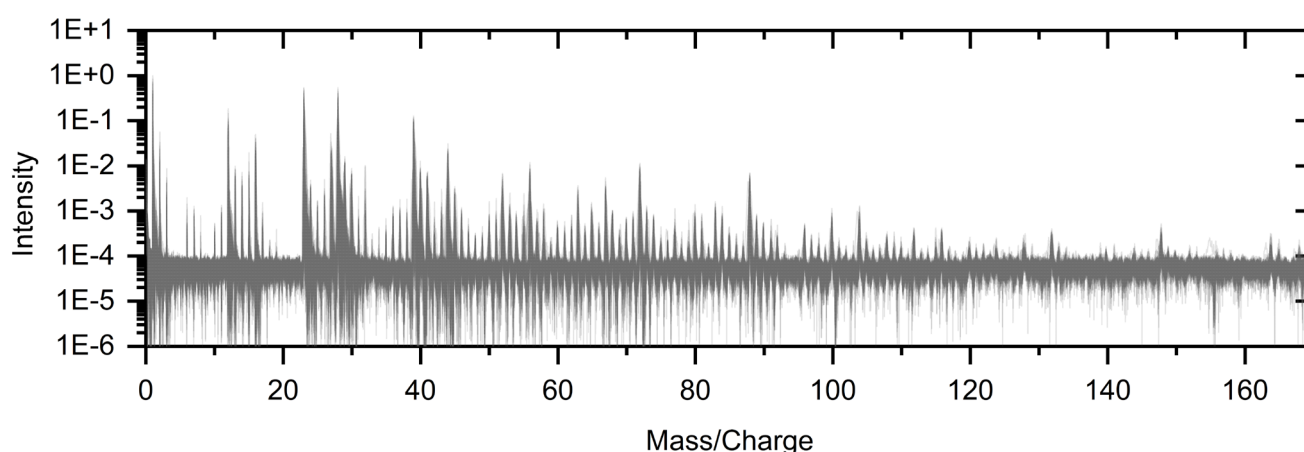
Fig S10. The mass spectrum registered from the surface of the sample. Hydrocarbons and Si oxides could be observed within the mass range from 36 to 160 amu.

Data processing and workflow

Fig. S11 depicts the data processing workflow used in this contribution. As it was mentioned in the main text, a depth profiling campaign was performed with the intention to remove layers contaminated with modern organic material and assess the original chemical composition of the microfossils. Due to the size of the microfossils (5-10 micrometer in length and nm scale thickness) (see the sketch in the main text – fig.1), microfossils embedded within the quartz matrix will be sparsely sampled. In an attempt to overcome this problem, we recorded more spots (nine depth profiles) from the microfossils-rich area and six spots (depth profiles) from the host area. As mentioned earlier, due to the specifications of our detector system, we can collect two spectra from the same laser shot on two high-speed digitizers - PXI3 and PXI5 (recording 64000 points per spectrum). The full amount of data recorded from the Gunflint sample resulted in 77'000 spectra from the depth profiling and 30'000 for the imaging campaign, making it 107'000 spectra in total.

Each of the presented peaks presented in fig. S10 has been integrated using corresponding time-of-flight windows. Details of integration procedure and spectra processing could be found here - (10); information about our laser and detector system - (9, 15) and integrated microscopy system - (4).

Using the in-house made software (10), we assigned mass scales to the acquired spectra and performed peak intensity extraction from 182 separate masses. Retrieved intensities from the same measurements

are summed and log-transformed. Each measured spectra assigned with an index of the depth location and class of the measurement, depending on which location depth profile was recorded (host/microfossils). Figure S11 presents a diagram of the data analysis pipeline. The first 500 spectra from the surface (data identified as surface contamination) were removed from the calculation of PCA scores, van Krevelen ratios, and classification models

Since it was identified from the LIMS imaging campaign that carbon is the main tracer of the microfossils, the whole spectral dataset collected from the microfossil-rich location was filtered on carbon intensities. To remove all spectra that belong to the quartz within the microfossil dataset and leave only spectra from the microfossils, we used the carbon threshold on 5.8 log10 el/ns level (see further details in the machine learning section). PCA values shown in fig. 7 (main text) were calculated from the z-scores of the retrieved intensities using correlation matrix and unstandardized scores. Variance showed in the PCA-2 (vertical axis) mainly reflects the carbon enrichment within the microfossils bodies, whereas PCA-1 variance mainly reflects the total intensity of the spectrum. As shown in Figure 9 (main text), the host mineral could be characterized with less variance and more uniform spectral features. The microfossils, on the other hand, are more diverse in PCA terms and spread within a larger area, which was interpreted as a sign of the chemical diversity of the microfossils. A similar pattern could be observed within the van Krevelen ratios (fig.7, main text), where host mineral measurements reveal a uniform distribution, whereas the ratios calculated from the microfossils are more diverse and reflects different mixing ratios with the quartz mineral and largely overlap in the middle area of the plot.

The $^{28}Si/^{16}O$ and $^{28}Si/^{1}H$ ratios were calculated from the host location dataset to subtract the isobaric interferences from the inorganic media present in the spectra measured from the microfossils. Statistical mode values within each of these distributions were chosen as representative ratios for the quartz mineral. Since the ^{28}Si mass peak is free of any interferences, peak intensities for ^{16}O and ^{1}H in microfossils have been recalculated using mode values of quartz mineral by subtracting mass peak intensities measured at the host from the intensities of the same isotopes, measured at the microfossils.
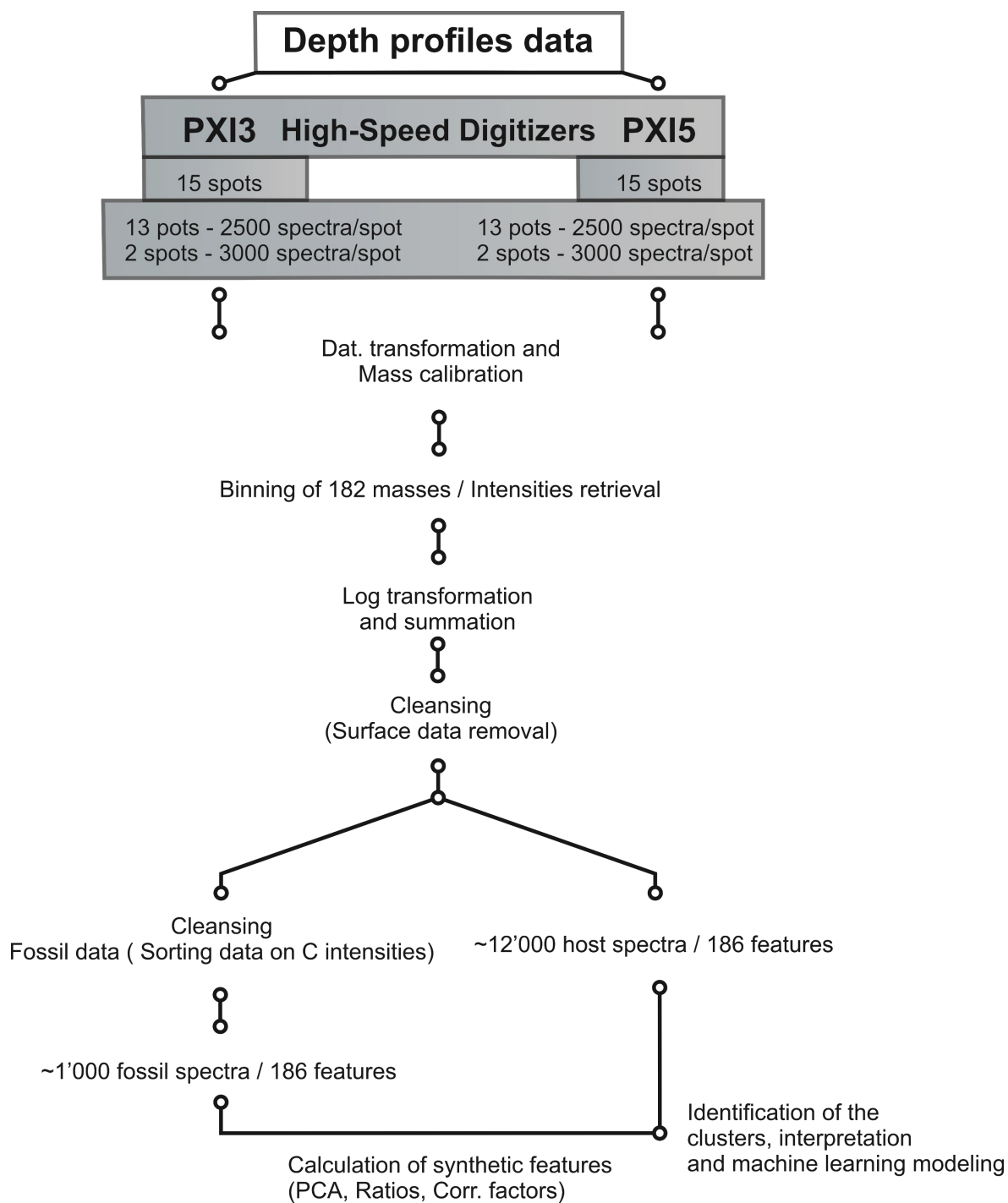
Depth profiles data

PXI3 High-Speed Digitizers PXI5

15 spots | 15 spots

13 pots - 2500 spectra/spot
2 spots - 3000 spectra/spot

13 pots - 2500 spectra/spot
2 spots - 3000 spectra/spot

Dat. transformation and
Mass calibration

Binning of 182 masses / Intensities retrieval

Log transformation
and summation

Cleansing
(Surface data removal)

Cleansing
Fossil data (Sorting data on C intensities)

~12'000 host spectra / 186 features

~1'000 fossil spectra / 186 features

Identification of the
clusters, interpretation
and machine learning modeling

Calculation of synthetic features
(PCA, Ratios, Corr. factors)

Fig S11. Data processing workflow.

Weighted mass correlation network analysis

Laser ionization mass spectrometry is a sensitive and quantitative method (3) that often produces spectra with complex chemical composition that might reflect the composition of the micrometer and even nanometer inclusions in a bulk of the sample. With an unknown sample, it might be challenging to assign specific masses to a given compound since it might represent a mixture of different minerals or a single complex mineralogical entity.

The ion yield from any given mineral registered with LIMS is a function of the chemical composition of the analyzed mineral with additional contribution from the ablation and ionization processes and local inhomogeneity of the sample. These instabilities form slightly different yields; however, on a larger scale, they form a linear pairwise dependency (see fig. S12-A) with a slope that determines their mixing ratio. If analyzed material within the diameter of the laser spot contains a few mineralogical inclusions, they should form a separate set of covariances. Thus, distinct mineralogical inclusions can be identified by scoring the pairwise linear correlation factors.

The relationship between measured LIMS mass/mass intensity values could be defined as the Pearson correlation coefficient:

$$(A, B) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{A_i - \mu_A}{\sigma_A} \right) * \left(\frac{B_i - \mu_B}{\sigma_B} \right)$$

Where each variable (A, B) has N scalar observations (number of spectra), where μ_A and σ_A are the mean and standard deviation of A (mass A) and μ_B and σ_B are the mean and standard deviation of B (mass B).

A dataset of pairwise correlation factors could be used to form the graph:

$$G\left[\rho > 0.4\right] = (V, E);$$

Where G is a graph, V are the vertexes in the graph, and E pairwise edges between vertexes weighted as $\rho(A, B)$, and ρ set to the specific threshold, to remove the clutter of insignificant correlation factors from the graph. In this study, we used the hard ρ threshold; however, soft (power) thresholds can be applied as well. The graph has a random initialization and should form mineral-specific topology as a $f(\rho)$ after implementation of the force-directed layout. Opensource graph drawing platforms Gephi and Cytoscape have been used to draw graphs.

There is a set of specific features that are needed to be checked before the correlation analysis. Firstly, we need to know that integrated masses correspond to the particular m/q assigned to the time-of-flight (TOF) window. For example, metastable molecules (molecules falling apart on the way to the detector) that form broad features might interfere with neighboring masses, creating artefactual detection. Secondly, a high mass resolution is required to integrate masses from strictly defined TOF windows. Furthermore, the signal registered from the background needs to be gaussian. Otherwise, all of the integrated masses will reveal relatively high covariance, determined by the oscillation modes of the baseline. The spectral resolution and ion yield stability, in our case, has been improved by using a fs UV-258 nm laser ablation/ionization and careful tunning of the ion-optical setup of the mass analyzer.
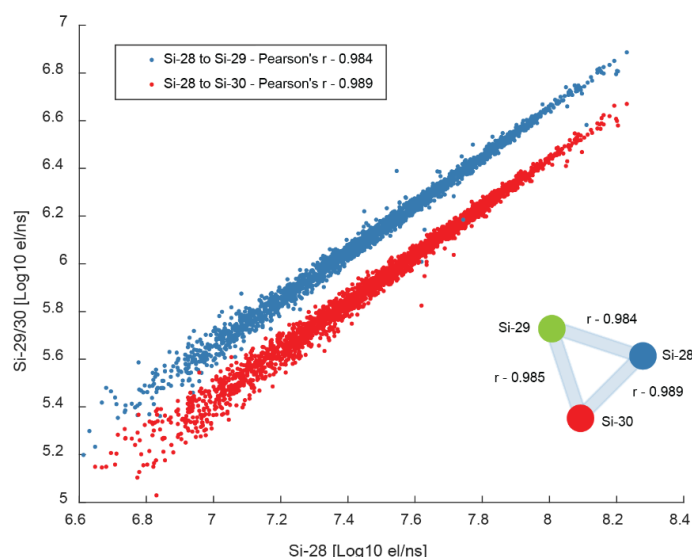
Fig S12-A. Schematic illustration of Si isotope correlation factors used as weights in weighted mass correlation network analysis.

The quality of the Si ion yield could be observed in fig. S12-A. A linear dependency registered from host location (quartz) for all Si isotopes, with enhanced scatter present in low-intensity spectra. These uncertainties could be mainly attributed to the contribution of noise. In principle, by histogramming more single-shot spectra, the quality of the ratios should grow proportionately. A second spectral feature that might affect the quality of the mass to mass ratios is an electronic ringing (unwanted oscillations of a voltage) caused by the impedance imperfections in the signal transmission line. The sketch provided in the lower part of fig. S12-A shows an embedding of Si isotope relationships, where nodes represent a single unit masses, and edges connecting each node are weighted according to the pairwise correlation factors.
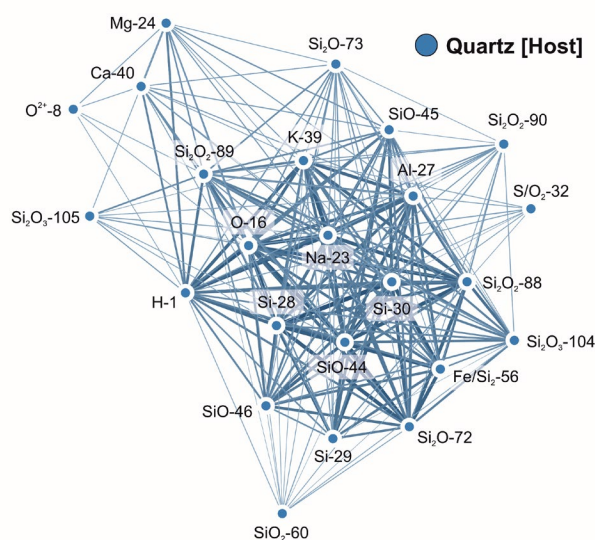
Fig S12-B. Correlation network of elements present at the host location. $\rho > 0.4$ used for the construction of the network. Thicker lines represent a higher correlation factor. Left-top nodes represent weakly connected impurities.

The useful feature of the Pearson correlation networks is that it is possible to use it for data mining purposes on large spectral datasets since Pearson ρ is sensitive to the outliers. Due to the gaussian character of the measured background intensities, covariance between two backgrounds (see the white area in fig. S15) reveal a correlation close to zero. However, if a few spectra among other thousands of measured spectra have an intense peak – the ρ value will reflect that change. The effect of that could be observed in fig. S13 (see the upper left nodes). Doubly charged O, Ca, and Mg could be observed within the host location. From the depth profile of this location, it is clear that these elements are weakly present in the depth profile (they appear only in a few spectra within specific depths) and represent an impurity within diagenetic quartz. In the case of simple averaging of the spectra, we would likely lose such fine information from the output spectra. To compare ρ values from the microfossil-containing spots, see fig. S14.

Modularity (Q) used in this contribution to separate clusters (communities) estimates the density of links inside communities compared to links between communities. Essentially it is just a measure of how many connections (edges) are between nodes within the group compared to other groups. In LIMS data, isotopes and molecules from the same minerals form a more tightly bounded group of nodes, representing a separate entity (Fig. S14). Micro- and nano- mineralization present within microfossil bodies embedded in a matrix form a complex time-of-flight spectrum, which is not always easy to decipher (i.e., figs. 4B, 4C in the main text). However, we know that they should form appropriate structure and density of links between nodes by using pairwise correlation factors. To divide the calculated networks into separate modules, we used a Louvain modularity score (16), which was developed for the analysis of large networks, though, it works well with small networks as well (>100 nodes).
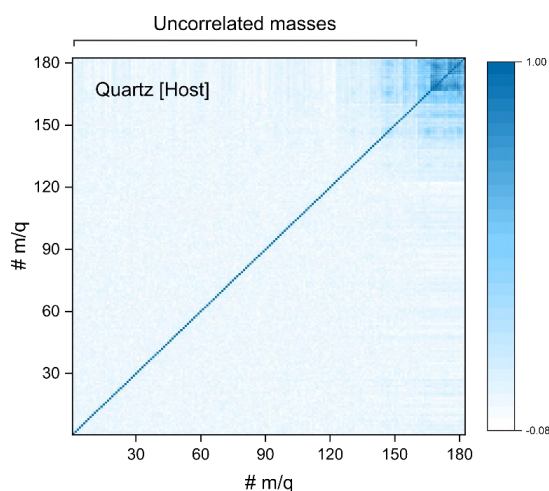
Fig S13. Visualization of the full feature space registered from the host location. Matrix represents hierarchically clustered values of mass-to-mass correlations (182 m/q in total) from the depth profile acquired from the host location. A single correlated entity could be observed, which corresponds to the diagenetic quartz.

The same information as in figs. 4-A and S12-B, could be visualized as a symmetric matrix and clustered using a hierarchical biclustering algorithm. Fig. S13 shows full feature space recorded from the depth profile acquired from the host location (2250 individual mass spectra) and corresponds to the spectrum in Fig. 4-A. Note that masses are indexed in an ascending manner (#m/q) and do not represent their actual m/q values. The data shown in fig S13 reveal a single correlated entity (quartz) with a gradual decrease of ρ values (from the heavy Si oxides due to their irregular formation pattern). The same results are acquired with the clustering of nodes using force-directed layouts (17) (see fig S12-B).
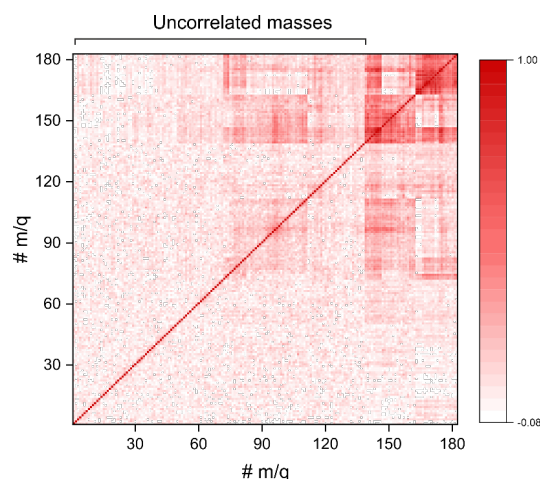
Fig S14. Visualization of the full feature space registered from the microfossil. Hierarchically clustered matrix of mass-to-mass correlations (182 m/q in total) from the depth profile acquired from the microfossil-rich location (Fig. 4 middle panel, location X, and Fig.5B). Two entities could be observed from the matrix, which are corresponds to the individual microfossil and quartz.

Machine learning results

In-situ space research is usually constrained in the amount of data that could be sent back to the Earth. However, modern high-throughput instruments are capable of creating relatively large datasets in a short amount of time. One of the possible approaches to deal with that is the initial data preprocessing and classification onsite. With the recent progress in machine learning (ML) and automated classification, it is possible to create deployable solutions that might perform classification tasks in real-time. For example, LIMS spectra acquired from the surface of the sample might be checked for the presence of biosignatures by comparing to the spectral datasets acquired from the "model" terrestrial microfossils. As "model" terrestrial microfossils, we can use samples with proven biogenicity, thus providing a targeted analysis of spectra of interest. Another application of ML models might be used in the searches of early life on Earth, where often, morphological features can be destroyed, but chemical composition might be preserved within small inclusions. The chemical fingerprints might still be identifiable with sensitive instruments, providing an additional line of evidence towards establishing the biogenicity of the sample.

We used MATLAB machine learning presets to score different classification models and identify the best-performing ones. Results of the scoring of ML models are illustrated in Table S1. The individual mass intensities retrieved from the mass spectra (from the depth profiling campaign) were combined with the ratios ($^1H/^{12}C$, $^{16}O/^{12}C$, $^{39}K/^{23}Na$, $^{40}Ca/^{24}Mg$, $^{27}Al/^{28}Si$, $SiO(44$ m/q$)/^{28}Si$, $^{52}Cr/^{55}Mn$, $^{32}S/^{54}Fe$, $^{19}F/^{12}C$, $^{39}K^{23}Na/^{40}Ca^{24}Mg$, $^1H^{12}C/^{16}O^{12}C$, $^{28}Si/^{12}C$, $^{28}Si/^{23}Na$) and two synthetic features – geometrical and arithmetic mean values of the light masses (from 1 to 74 m/q), and in total, it allowed us to form a dataset with 196 variables. Spectra collected from the microfossils rich region were filtered on the carbon intensities to remove the data corresponding to the matrix (quartz). Additionally, observations from both regions (microfossils and host area) were filtered to the depth region 500-2500, to remove the contribution from the surface. The primary working dataset, dataset #1 (left column in table S1), were filtered on 5.8 log10 el/ns intensity value (leaving only intense C signal from the spectra recorded from the microfossils region), which is slightly higher than the noise signal, and formed a spectral databank with 13454 spectra, where 1454 spectra are registered from the microfossils region and 12000 spectra are recorded from the host mineral. Dataset #2 was filtered on 6 log10 el/ns Carbon threshold (right column in table S1) and formed a spectral dataset with 12863 spectra, where 863 spectra are registered from the microfossils region (dataset #2 were used in PCA calculation (fig. 9), and ML scoring (table S1)). Further, all intensities and ratios have been log-transformed and z-score standardized.

As it could be seen from table S1, both datasets are performed very well (around 99%), and most of the misclassifications of microfossils are attributed to the low carbon intensity spectra, which are overlapping with the measurements from the host area. Visual assessment of the dataset and complexity of the classification task could be done by examining the PCA score shown in fig. 9 in the main text and Fig. S15 and S16. The ML modeling of the slightly overlapping datasets allowed learning true separation boundaries between the intensity values within different classes. To ensure that models do not overfit the data, we implemented five-fold cross-validation. Additionally, a Bayesian optimization (scoring the same model with different learning parameters) has been implemented to search through the hyperparameter space. After the optimization, adaptive boosting revealed the best score within both datasets, successfully classifying 98.8% and 99.7% of spectra. Overall, 19012 unique borderlines between different sets of mass and ratio pairs are used to create a multidimensional intensity-based classification model. To aid the interpretation of ML models, we calculated a volumetric binary PCA cube shown in Fig. S15 and S16.

Data transformations:	Z-scores of Log trans-n/ Depth (500-2500) and C (5.8-10) filtered	Z-scores of Log trans-n/ Depth (500-2500) and C (6-10) filtered
ML models	Accuracy %	Accuracy %
Fine Tree	97.9%	99.4%
Medium Tree	98.2%	99.4%
Coarse Tree	98.0%	99.6%
Linear Discriminant	96.8%	98.8%
Logistic Regression	97.9%	99.0%
Gaussian Naïve Bayes	91.0%	93.6%
Kernel Naive Bayes	94.0%	94.0%
Linear SVM	97.9%	99.4%
Quadratic SVM	98.2%	99.4%
Qubic SVM	98.2%	99.4%
Fine Gaussian SVM	89.2%	93.3%
Medium Gaussian SVM	97.9%	99.4%
Coarse Gaussian SVM	96.9%	99.1%
Fine KNN	94.3%	98.0%
Medium KNN	94.8%	98.2%
Coarse KNN	93.9%	97.0%
Cosine KNN	96.6%	98.7%
Weighted KNN	95.1%	98.1%
Boosted Trees	98.8%	98.4%
Bagged Trees	97.1%	98.4%
Subspace Discriminant	96.6%	98.9%
Subspace KNN	94.7%	97.9%
RUSBoosted Trees	98.4%	99.6%
AdaBoost (after Bayesian opt.)	98.8%	99.7%
Predictors	196	196
Observations (total/fossil)	13454/1454	12863/863
Cross-validation	5 fold	5 fold

Table S1. Performance of the classification models.
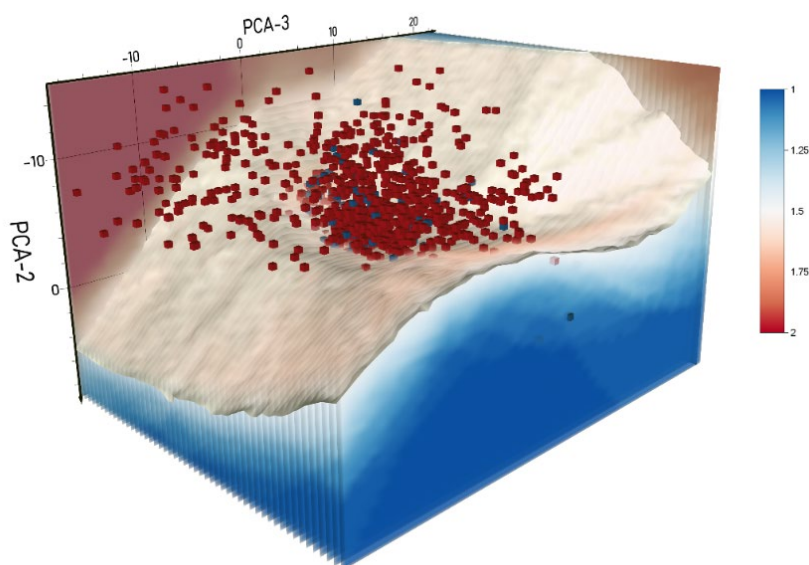
Fig S15. Visualization of the separation boundary between two classes – microfossils (red) and quartz matrix (blue) calculated on three principal components by using an inverse distance volumetric interpolation (view of the microfossils spectra). White areas under the curve represent the uncertainty of the class assignment. Blue area and blue nodes represent the PCA embedding of LIMS spectra acquired from the host area (quartz mineral). Red area and red nodes represent spectra measured from the microfossils.
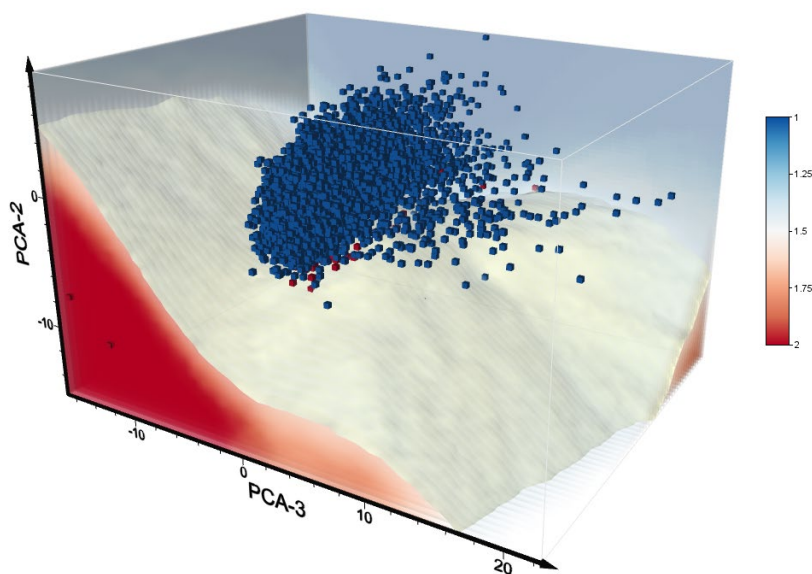
Fig S16. Visualization of the separation boundary between two classes – microfossils (red) and quartz matrix (blue) calculated on three principal components by using an inverse distance volumetric interpolation (view of the host spectra). White areas under the curve represent the uncertainty of the class assignment. Blue area and blue nodes represent the LIMS spectra acquired from the host area (quartz mineral). Red area and red nodes represent spectra measured from the microfossils.
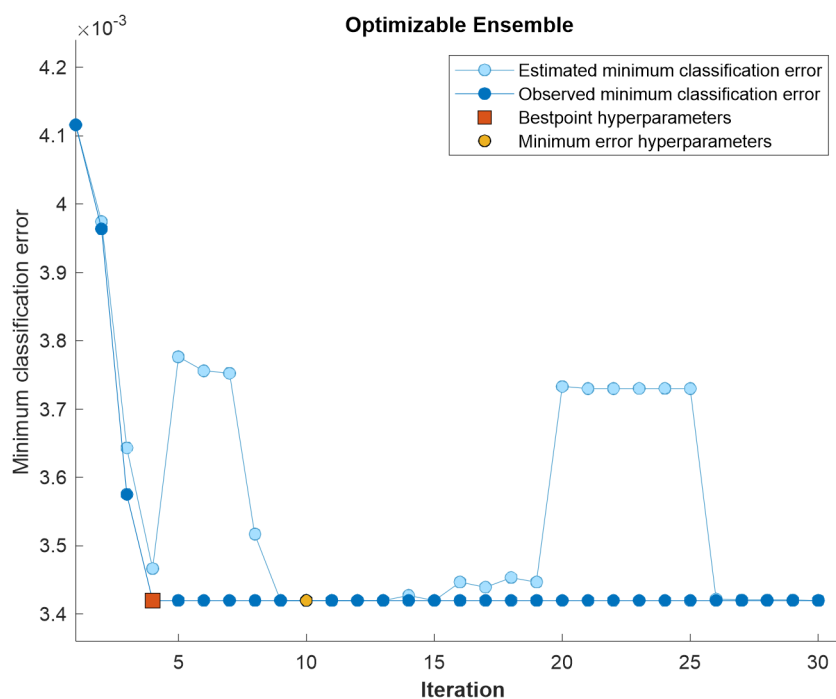
Fig S17. Minimum classification error change using the 30-step Bayesian optimization.
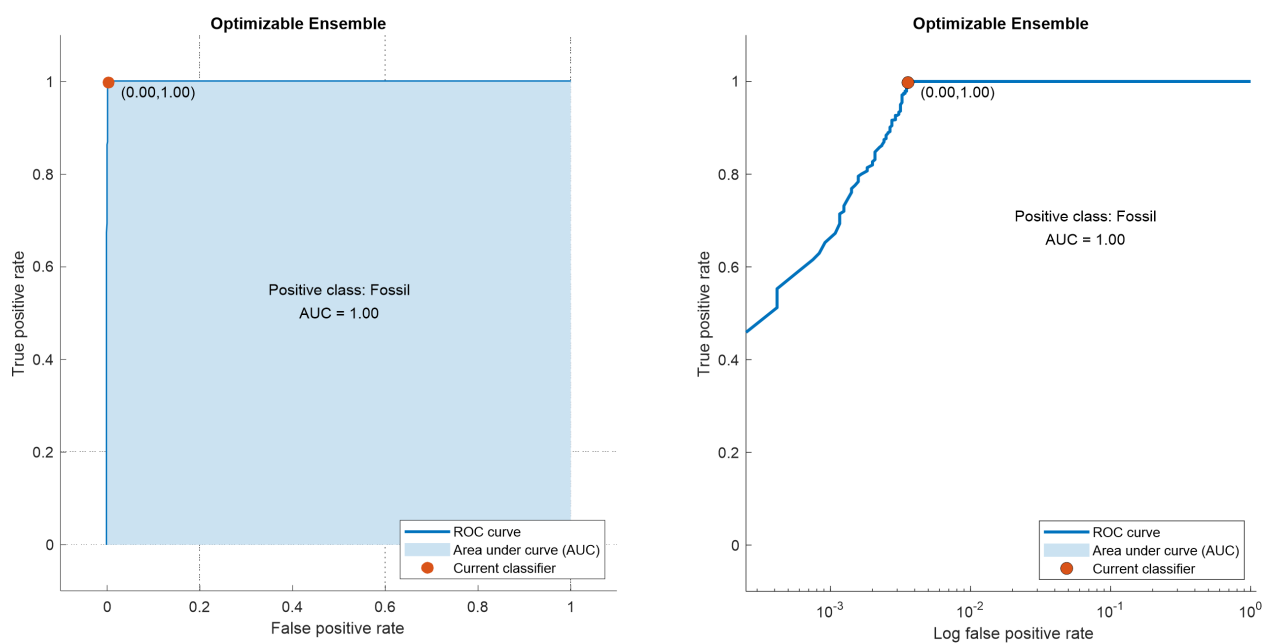
Fig S18. The classification quality of the best performing ensemble model. ROC curve and corresponding AUC value are very close to the perfect classifier. Log scale ROC curve presented on the left panel indicates the small discrepancy from the perfect model.
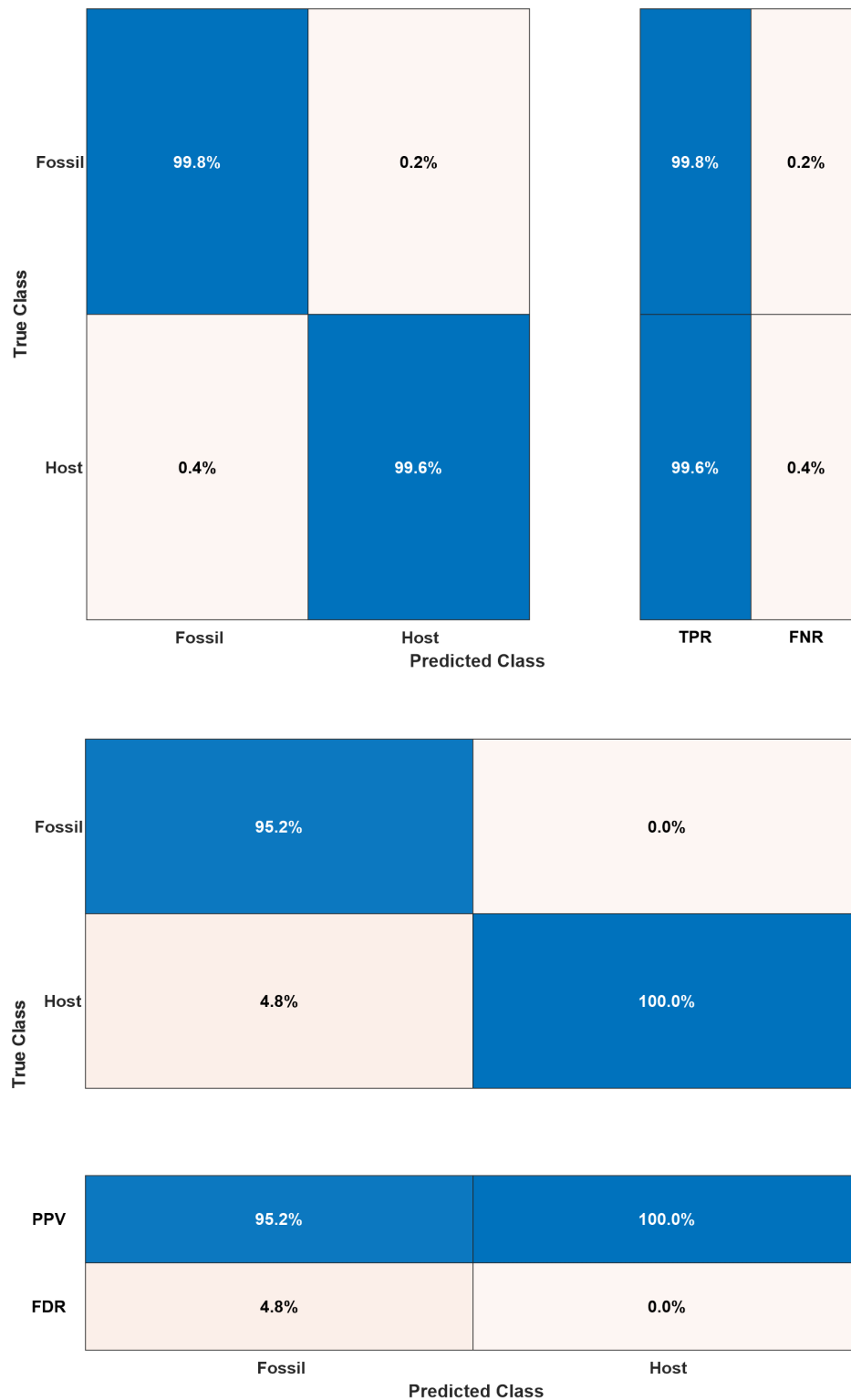
Fig S19. Classification quality of the best-performing model using different metrics. Upper panel – True positive rates (TPR) and false-negative rates (FNR). Lower panel – Positive predictive values (PPV) and False discovery rates (FDR).

Figure S17 shows the change of the classification error in different models using the Bayesian optimization procedure. The 30 step search was performed to find the optimal number of learners, predictors and number of splits. The hyperparameter search range included five ensemble methods – Bag, GentleBoost, LogitBoost, AdaBoost, and RusBoost. The search for the optimal number of learners was

conducted in the range from 10 to 500. The learning rate was probed in the range from 0,001 to 1. And lastly, the varying number of predictors was sampled from 1 to 196. The best-performing model revealed an average accuracy of 99.7% (see table S1). The search and optimization time on the personal laptop (i7 CPU) using five parallel workers resulted in 16 minutes. The lowest error values in fig. S17 indicates the best performing set of hyperparameters. The blue line shown on the plot indicates the minimum classification error that was observed during the search campaign. The light blue line indicates the calculated error of classification, the spread between two values (dark and light blue) indicates divergence of the error from the optimal model. One can see that local minimum found by the optimizer in the beginning, is persistent and present in 9 models out of 30 and likely represent a global minimum of the classification error. Thus, the ML model indicated with the red square represents the best-performing one.

The corresponding ROC curve and AUC values for the best-performing model are shown in fig. S18. The linear scale ROC curve indicates that overall performance is very close to the perfect classifier. However, the log scale ROC curve (right panel in fig. S18) suggests the presence of classification errors. Figure S19, the upper panel, shows accuracy of the model using true positive rate metric (TPR) - also referred as sensitivity or recall. This parameter is used to measure the percentage of actual positives, which are correctly identified. The overall performance of the model is over 99% for both classes – microfossils (99.8%) and host mineral (99.6%). The lower panel shows the PPV and FDR values. The FDR is the proportion of incorrectly classified observations per predicted class, whereas PPV is the proportion of correctly classified observations per predicted class. The results of the scoring indicate that false discovery rate for the fossils is larger than for the host mineral. The false discovery rate for fossils was found to be 4.8%, whereas FDR for the host mineral is 0. The discrepancies in the FDR values can be attributed to the unequal sampling rate of the classes. Thus, the classification model performance in FDR terms can be improved by increasing the sample rate of the microfossils.
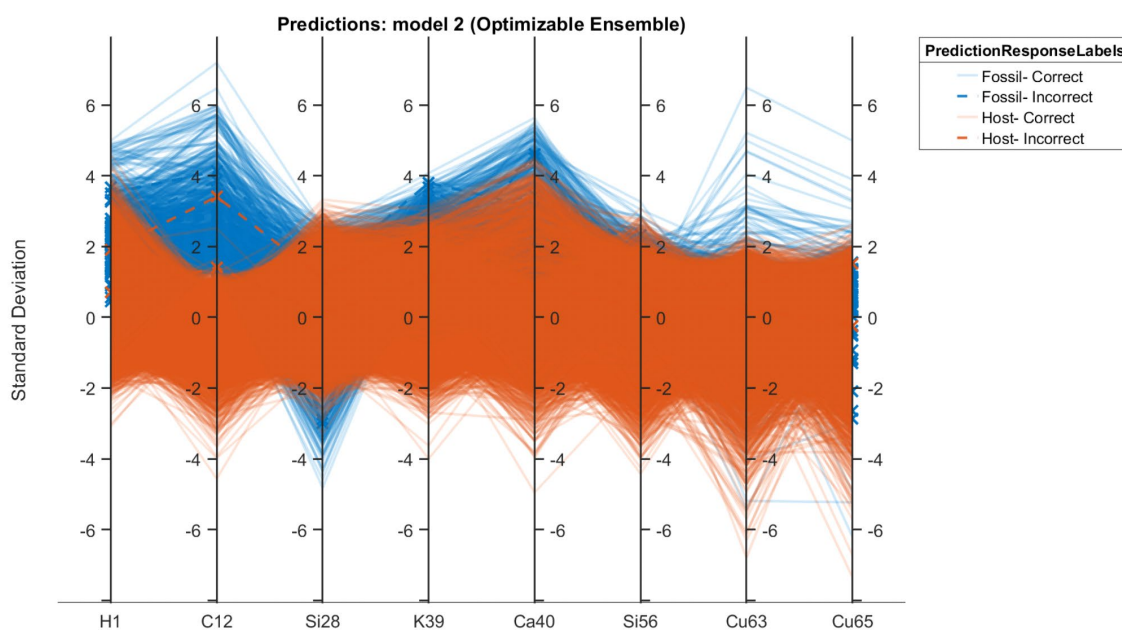

Fig S20. ML model predictions visualized using the parallel coordinates for selected isotope intensities.

Figure S20 shows the model predictions using the standardized isotope intensity distributions from two classes. The intensities from the microfossils are colored with blue lines. Intensity values from the host mineral are colored with the red line. The dashed line and crosses indicate the erroneous predictions of the model. As one can see, ^{12}C mass line intensities show the best separability, followed by ^{28}Si, ^{39}K, ^{40}Ca, and ^{63}Cu. The ^{1}H line shows significant overlap between classes. The 196 predictors in total were used to train the classifier.

Outlook

In this contribution, capability of the LIMS microprobe to detect and characterize two billion years old microfossils preserved in the silicified matrix were assessed. Further work will be required to collect an even higher statistic on the Gunflint microfossils to see if any pronounced chemically distinct types of microfossils are present within the Gunflint microbiome and with which particular morphology they are associated. Mass spectrometric imaging was found to be a handy tool in the visualization of the distribution of various microfossils on the surface of the sample. However, further work will be required to implement Ar+ sputtering or other methods of surface conditioning to ensure that modern organics will be removed from the surface and will not affect the imaging results. In general, the elemental composition of the microfossils from any silicified matrices should be located in the close neighborhood to the spectra acquired from the Gunflint microfossils. Nevertheless, the chemical composition of other Precambrian microfossils from silicified matrices needs to be assessed to understand how representative are the spectra acquired from the Gunflint microfossils in respect to other formations and which part of the chemical variance they cover. Future work will be required to assess the capability of LIMS in combination with the LD-MS (Laser Desorption Mass Spectrometry) in the identification and characterization of organic remains from complex matrices and their difference from the Fischer-Tropsch-Type (FTT) synthesis products.

SI References:

1. M. Tulej *et al.*, CAMAM: A miniature laser ablation ionisation mass spectrometer and microscope-camera system for in situ investigation of the composition and morphology of extraterrestrial materials. *Geostandards and Geoanalytical Research* **38**, 441-466 (2014).

2. P. Wurz *et al.*, Mass spectrometric analysis in planetary science: Investigation of the surface and the atmosphere. *Solar System Research* **46**, 408-422 (2012).

3. M. B. Neuland *et al.*, Quantitative measurement of the chemical composition of geological standards with a miniature laser ablation/ionization mass spectrometer designed for in situ application in space research. *Measurement Science and Technology* **27**, 035904-035904 (2016).

4. R. Wiesendanger *et al.*, Chemical and Optical Identification of Micrometer-Sized 1.9 Billion-Year-Old Fossils by Combining a Miniature Laser Ablation Ionization Mass Spectrometry System with an Optical Microscope. *Astrobiology* **18**, 1071-1080 (2018).

5. N. F. W. Ligterink *et al.*, ORIGIN: a novel and compact Laser Desorption – Mass Spectrometry system for sensitive in situ detection of amino acids on extraterrestrial surfaces. *Scientific Reports* **10**, 9641 (2020).

6. P. Moreno-García *et al.*, Towards matrix-free femtosecond-laser desorption mass spectrometry for in situ space research. *Rapid Communications in Mass Spectrometry* **30**, 1031-1036 (2016).

7. M. Tulej *et al.*, Chemical Composition of Micrometer-Sized Filaments in an Aragonite Host by a Miniature Laser Ablation/Ionization Mass Spectrometer. *Astrobiology* **15**, 669-682 (2015).

8. R. Wiesendanger *et al.*, Improved detection sensitivity for heavy trace elements using a miniature laser ablation ionisation mass spectrometer. *Journal of Analytical Atomic Spectrometry* **32**, 2182-2188 (2017).

9. A. Riedo, M. Tulej, U. Rohner, P. Wurz, High-speed microstrip multi-anode multichannel plate detector system. *Review of Scientific Instruments* **88**, 045114-045114 (2017).

10. S. Meyer, A. Riedo, M. B. Neuland, M. Tulej, P. Wurz, Fully automatic and precise data analysis developed for time-of-flight mass spectrometry. *Journal of Mass Spectrometry* **52**, 580-590 (2017).

11. A. Riedo, A. Bieler, M. Neuland, M. Tulej, P. Wurz, Performance evaluation of a miniature laser ablation time-of-flight mass spectrometer designed for in situ investigations in planetary space research. *Journal of Mass Spectrometry* **48**, 1-15 (2013).

12. A. Riedo *et al.*, Laser ablation/ionisation mass spectrometry: Sensitive and quantitative chemical depth profiling of solid materials. *Chimia* **70**, 268-273 (2016).

13. R. Wiesendanger *et al.*, A method for improvement of mass resolution and isotope accuracy for laser ablation time-of-flight mass spectrometers. *Journal of Chemometrics* **33**, 1-10 (2019).

14. V. A. Azov, L. Mueller, A. A. Makarov, Laser ionization mass spectrometry at 55: Quo Vadis? *Mass Spectrom Rev* 10.1002/mas.21669 (2020).

15. A. Riedo, M. Neuland, S. Meyer, M. Tulej, P. Wurz, Coupling of LMS with a fs-laser ablation ion source: Elemental and isotope composition measurements. *Journal of Analytical Atomic Spectrometry* **28**, 1256-1269 (2013).

16. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).

17. M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* **9**, e98679-e98679 (2014).

11. 2 Supplementary Information for:

"On topological analysis of fs-LIMS data. Implications for in situ planetary mass spectrometry" - (Frontiers in Artificial Intelligence)

Rustam A. Lukmanov[1*], Andreas Riedo[1], David Wacey[2], Niels F.W. Ligterink[1], Valentine Grimaudo[1], Marek Tulej[1], Coenraad de Koning[1], Anna Neubeck[3], Peter Wurz[1]

[1]Space Research and Planetary Sciences (WP), University of Bern, Bern, Switzerland
[2]Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, Perth, Australia
[3]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

*Correspondence:
Rustam A. Lukmanov: rustam.lukmanov@space.unibe.ch

In order to check that cluster assignments are not artifactual (see Fig.5 in the main text), we performed a clustering robustness analysis. The Rand Index (RI) metric was used to assess the clustering similarity between 10 random subsamples of the data registered from microfossils. In total, 75% of the data was used to generate random subsamples. The output UMAP subsamples were clustered using the Louvain community detection algorithm. The RI similarity matrix for Louvain clustering of random samples could be found in the supplementary information (see Table S1 and Fig. S3). Overall, 45 different clustering pairs revealed an average RI score of 92.5% with a standard deviation of 2%, which indicates that communities shown in Fig. 5 are not artefactual and that the cluster assignments are robust. Most of the clustering uncertainty can be attributed to the transition zone between two types of microfossils. The type-2 microfossils reveal more inhomogeneity (see Figs.S1 and S2) in comparison to the type-1 microfossils and represent more intermixed with the host mineral material.

Figure S1. The Mapper networks constructed from subsampled data, shaped
(1473, 6) and using lens shaped (1473, 3) - 75% of original data
Number of filter cubes = 20, percent of overlap = 0.5
UMAP projection dimensions - 0, 1, 2
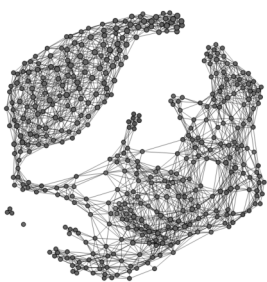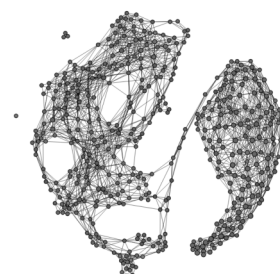Clusterer - DBSCAN (epsilon = 25, minimal number of samples = 8, metric = 'cosine')
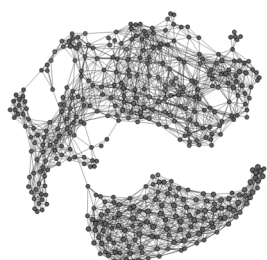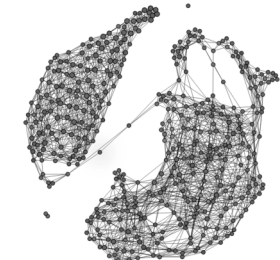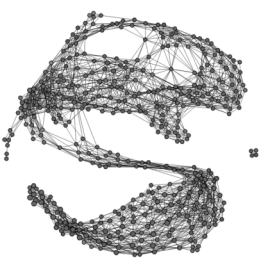
Subsample-1
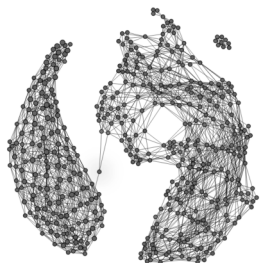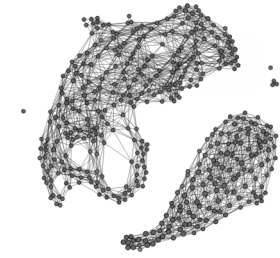
Subsample-6
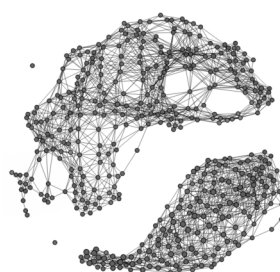
Subsample-2

Subsample-7

Subsample-3

Subsample-8

Subsample-4

Subsample-9
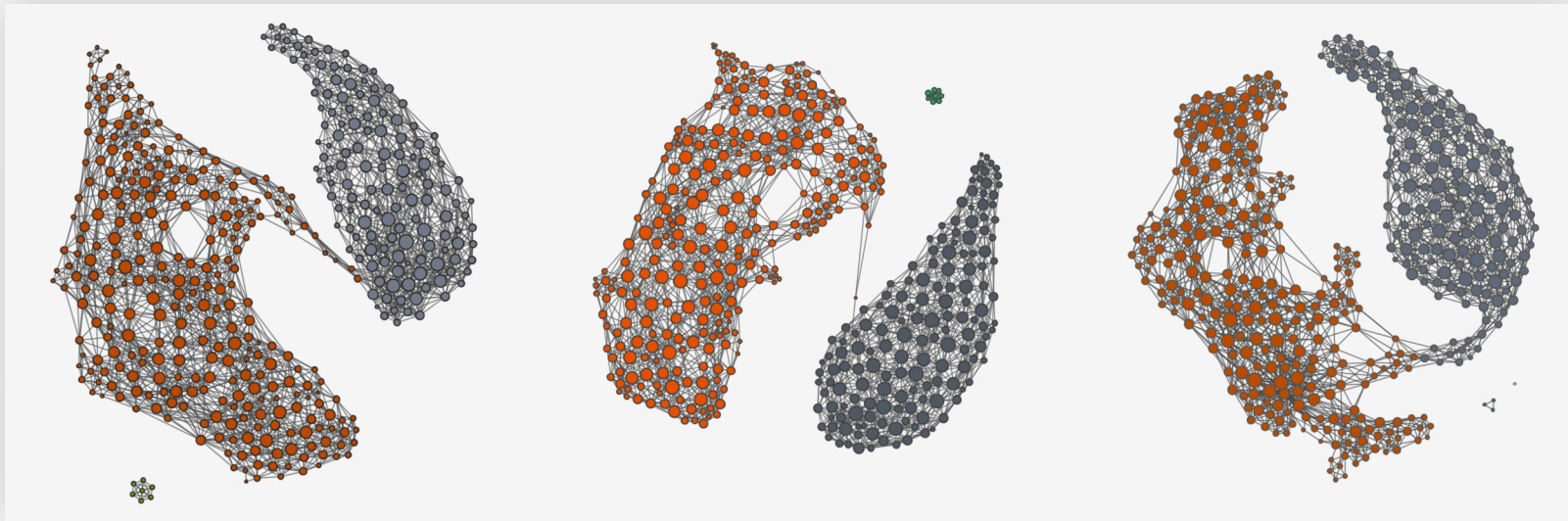
Subsample-5
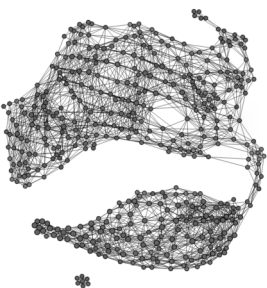
Subsample-10
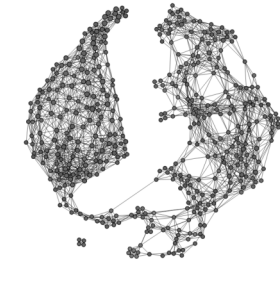
Figure S2. Exemplary Louvain clustering of random subsamples.
The type-1 microfossils (grey nodes) reveal stable clustering results. Whereas,
type-2 microfossils (red nodes) show an increased level of inhomogeneity.
Node sizes are scaled according to the nodal degree.

Rand Index

| Samples	Subsample1	Subsample2	Subsample3	Subsample4	Subsample5	Subsample6	Subsample7	Subsample8	Subsample9	Subsample10
Subsample1	1.000	0.941	0.920	0.957	0.909	0.956	0.934	0.942	0.938	0.955
Subsample2	0.941	1.000	0.902	0.934	0.918	0.943	0.933	0.906	0.920	0.919
Subsample3	0.920	0.902	1.000	0.926	0.871	0.923	0.892	0.929	0.913	0.926
Subsample4	0.957	0.934	0.926	1.000	0.901	0.932	0.937	0.959	0.935	0.972
Subsample5	0.909	0.918	0.871	0.901	1.000	0.913	0.910	0.875	0.907	0.907
Subsample6	0.956	0.943	0.923	0.932	0.913	1.000	0.920	0.912	0.939	0.925
Subsample7	0.934	0.933	0.892	0.937	0.910	0.920	1.000	0.924	0.920	0.926
Subsample8	0.942	0.906	0.929	0.959	0.875	0.912	0.924	1.000	0.913	0.966
Subsample9	0.938	0.920	0.913	0.935	0.907	0.939	0.920	0.913	1.000	0.926
Subsample10	0.955	0.919	0.926	0.972	0.907	0.925	0.926	0.966	0.926	1.000
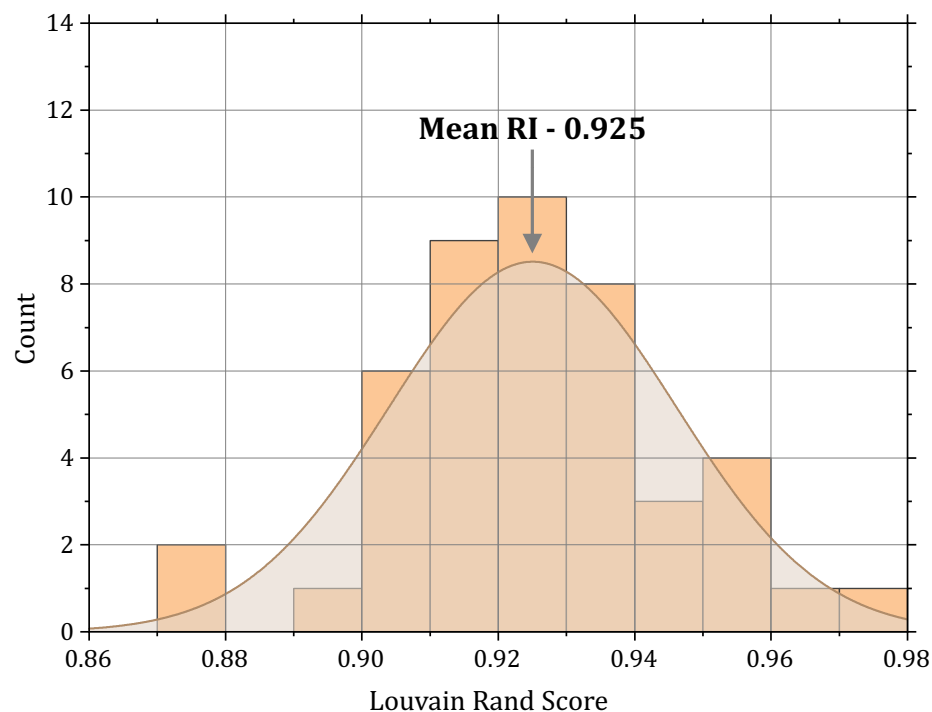
Table S1. Rand Index (RI) scores calculated for Louvain clustering of random subsamples (1473 mass spectra or 75% of original data). Minimal number of samples for DBSCAN clusterer were set to 1 to avoid unequal sampling. Majority of the RI values are above 0.9.

Figure S3. Distribution of the RI scores. Mean value equals to 0.925 (derived from 45 observations). Standard deviation - 0,02. Maximum value - 0.97; Minimum value - 0.87. Distribution indicates robust clustering results.

Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name:

Registration Number:

Study program:

Bachelor ☐   Master ☐   Dissertation ☐

Title of the thesis:

Supervisor:

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.

For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

2021

Place/Date

Rustam Lukmanov

Signature