ABSTRACT

Title of Dissertation:	CHARATERIZING CHEMCIAL SIGNATURES OF LIFE VIA MASS SPECTROMETRY
	Ziqin (Grace) Ni, Doctor of Philosophy, 2023
Dissertation directed by:	Professor Ricardo Arevalo Jr., Department of Geology

The search for signs of life beyond Earth has been fueled by a natural curiosity about whether we are alone in the universe. Organic molecules, as the primary chemical components of terrestrial living organisms, are major targets in life detection missions. However, organic compounds have also been found in abundance in meteorites. They can be synthesized via abiotic processes such as lightning strikes, and naturally degraded with time. Searching for chemical signatures of life requires analog experiments to constrain chances of false positive detection and advancement of instrumentation to reduce possibilities of false negative measurements.

The capacity to synthesize organics via abiotic mechanisms is influenced strongly by the redox condition. In this dissertation, the redox state of early Earth is estimated using trace element systematics in zircon grains. The Earth's surface is found to have reached a habitable redox condition as early as 4 billion years ago, coinciding with a time when Earth was still routinely bombarded by meteorites. Simulations of such high-speed impacts using high-power laser beams demonstrate the feasibility of synthesizing simple organic compounds from carbonates and nitrogen salts. Laboratory experiments

and numerical modeling suggest that crater-forming impacts could have synthesized a considerable concentration of organics on the surface of Earth, Mars, and Ceres.

The detection and characterization of organic molecules requires sophisticated analytical instrumentation. Laser desorption mass spectrometry (LDMS) and Orbitrap^{TM} mass analysis are examples of next-generation techniques under development for conducting comprehensive chemical investigations in space. Although a combination of these two techniques enables the determination of the atomic composition of organic molecules, even complex polymers such as peptides, such an approach fails to recognize the 3D structure and sequencing of polymers. To facilitate the ionization and sequencing capability of peptides via LDMS, silicon nanoparticles are incorporated in the substrate as an alternative to conventional organic matrices but with reduced risk of forward contamination. The simultaneous measurement of mass to charge ratio and collision cross-section via Orbitrap mass analysis allows for rapid separation of organic molecules by their class, structure, and composition in sample mixtures. The techniques developed here are valuable for astrobiological exploration beyond Earth.

CHARATERIZING CHEMCIAL SIGNATURES OF LIFE VIA MASS SPECTROMETRY

by

Ziqin Ni

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Associate professor Ricardo Arevalo Jr., Chair
Associate professor Nicholas C. Schmerr
Associate professor Alan Jay Kaufman
Associate professor Peter Nemes
Dr. Stephanie Getty
Professor Amy S. Mullin (Dean's Representative)

Dedication

To myself, a courageous young lady.

Acknowledgements

This dissertation, and the amount of work encompassed by this graduate research, could not have been completed without the support of the community around me. This is not merely completion of a Ph.D. but rather the start of a new life journey. I sincerely thank everyone who has accompanied me on this journey with gratitude from the depths of my heart. Ricardo Arevalo Jr. deserves special recognition for his role as my life mentor.

- Rick influenced me to run and bike. In the past few years, I have completed multiple half and full marathons, century bike rides, and one half-ironman race. Because of him, I now lead an active lifestyle and strive to be an athlete as a side job.
- Rick encouraged me to be a leader and supported me firmly when things went wrong. It has become clear to me that I want a leadership position. I have developed values that I advocate for, and I will become a leader of my kind at my own pace.
- Rick is not a perfect advisor, and that's why he is a nearly perfect mentor. The truth is, no human is perfect. This is why I can challenge authority for the things I believe in, even as a student, because I know they cannot always be right. This is also why I can be less afraid to step into any position just because I am new to the role. I am very sure that I will make mistakes, so I will learn along the way.

My gratitude to Rick goes on. I cherished the time when I was his protégé, regardless of whether it was good or bad, ups or downs. I will move on with the wisdom I have learned.

I am grateful for the guidance and friendship that I have received from Phil Piccoli. In particular, he taught me the importance of being patient. When I first entered the graduate program, I was deeply influenced by the competitive culture around me and thought that the speed of getting things done was more important than anything else, including my personal life and the quality of my work. In 2019, when I was preparing my first manuscript on how to select unaltered zircon grains in a faster way, Phil questioned my approach. I now realize that time is a friend, not an enemy. Sometimes, efficiency is important. But at other times, producing good work, developing original ideas, and even building relationships requires taking things slow. There is no need to rush.

I am also thankful for the encouragement and care that I have received from William McDonald. Bill is wise in many ways, but two things that have influenced me the most are his approach to making approximations and his dedication to science. Bill is skilled at thinking about numbers at a scale that matters to the scientific question. He simplifies cumbersome details to speed up the thinking process, while also focusing on critical details to ensure that the scientific prediction is accurate. Bill never stops working, reading, doing calculations, and asking hard questions. It has become clear to me that thinking smart requires lots of experience and dedication to train your mind.

I want to thank my mom and dad in China. They have been the ultimate cradle for me, no matter how tired, defeated, or sorrowful I was during the past 10 years of my academic journey in the United States. This has not been easy for them, not only because they really love me and miss me, but also because our lives, values, and future prospects have grown increasingly different due to the distance and life experiences between us. Their fears of missing out amplify with aging. I feel grateful and selfish for the selfless love they have given me, and I am growing to take responsibility for taking care of them too.

I want to thank my beloved Ph.D. fellows from the Geology department, in particular Laura Sammon, Ben Farcy, Joe Galena, Zach Zagas, James Bader, Rumya Ravi, Soumya Ray, Sourabh Shubham, Jiayang Sun, and so on. I have fond memories of graduate school, and for that, I thank my fellow graduate students as well as the UMD Geology faculty and staff, in particular Michelle Montero, Suzanne Martin, and Todd Karwoski throughout graduate school. I also want to thank important people outside of the Geology department. I cherish my 10 years of intimate friendship with Chengcheng Zhai since undergraduates. She has become a family of my choice. Garrett and Maggie have taught me the pleasure of suffering and the joy of overcoming physical and psychological pain to become stronger. Esther Xu and her perceptive thoughts on feminism have never stopped challenging my narrow view and inspired me to become more courageous at fighting bias. The friend circle I formed with Yupeng Zhao, Xuetian Xie, and Xinzi He has given me comfort and calm during stressful times. I love running, biking, and eating with them and cherish the invaluable memories. My

gratitude continues to many others who have shown up and enriched my life. Your presence, long or short, has made me who I am today.

Finally, I would like to thank Ricardo Arevalo Jr., Nicholas Schmerr, Peter Nemes, Stephanie Getty, Amy S. Mullin for serving on my Ph.D. committee as well as reading my dissertation and providing feedback. Ricardo Arevalo Jr., Nicholas Schmerr, Peter Nemes, Stephanie Getty are additionally thanked for their support throughout the proposal process and for their guidance over the course of this PhD. I am grateful for the funding provided by the University of Maryland and NASA for this dissertation.

Table of Contents

List of Tables

List of Figures

Figure 2.4. A sensitivity comparison between all three inclusion indicators: La concentration; $(Sm/La)_N$; and, LREE_I as a function of mixing ratio (%). The shaded areas represent regions that will be selected (or accepted) by the filter scheme.

Figure 2.5. The distribution of Th/U ratio in the compiled dataset (n=2173) from this study and zircon from Kirkland et al. (2015).

Figure 2.6. Hf concentration as a function of Eu/Eu* of zircon compiled in this study, categorized by host rock type. The inset shows the count distribution of Eu/Eu* of zircon by rock type.

Figure 2.7. Density plot of Ti-in-zircon temperatures determined from the zircon data set compiled here, categorized by host rock type. The area under each curve has a cumulative probability of 1.

Figure 2.8. An example of calculating the score of zircon that satisfy all criteria except that its $U/Nb > 20$, La concentration > 0.1 ppmw, and $(Lu/Gd)_N < 10$.

Figure 2.9. Application of the proposed filter scheme and scoring system to the zircon dataset compiled here (n = 2173). In general, the large spread in observed Ce/Ce* may be attributed to distinct sources of source melts, magmatic differentiation prior to zircon saturation, and/or mineral inclusions in trace element measurements. The black dotted and solid lines represent $\Delta FMQ \pm 2$.

Figure 2.10. Distribution of the compiled zircon data set within the pre-established magmatic and hydrothermal fields proposed by Hoskin (2005).

Figure 2.11. (a) Boxplots of zircon populations for each score. Based on the median and variance of inferred melt redox states, these populations may be divided into three groups: Group I (score 14 to 15), Group II (score 8 to 13), and Group III (score 1 to 7). (b) Selectivity and distribution of different rock types within each group (the percentage represents the proportion of all zircon that fall within each group).

Figure 2.12. Plot of Ce/Ce* and temperature in zircon as a function of morphology and textures. The texture of each individual zircon was deduced from BSE and CL imaging from source references. The black dotted and solid lines represent $\Delta FMQ \pm 2$.

Figure 2.13. Application of filter scheme to Jack Hill zircons (n=53). (a). The black dotted and solid lines represent mantle equilibrium state and has $\delta^{18}O$ 5.3 ± 0.3 ‰ (sd). This filter scheme has selected 7 mantle-equilibrated zircons and 6 of them are supported by mantle-equilibrated oxygen isotopes. The average Hadean magma oxygen fugacity is determined to be ΔFMQ 1.5 ± 1.3 (sd). (b). The black dotted and solid lines represent $\Delta FMQ \pm 2$. The Hadean mantle has oxidized to FMQ region since 4400 Ma.

Figure 2.14. Application of the filter scheme proposed here to a case study of TTG zircon from three independent samples (524906, 524907, 524994) from Greenland (see Apendix A for location). (a) Inferred melt redox state as a function of temperature. (b) Distribution of zircon by group.

Figure 3.1. A. Simulation of a crater-forming impact on planetary surface using a beam of fast-moving $^{13}C_{60}^{+}$ projectiles that accelerated towards a substrate rich in nitrogen

salts. **B.** simulation of vapor plume generated by a planetary-scale impact via high-power laser irradiation on a substrate rich in nitrogen salts and carbonates.

Figure 3.2. Synthesis of $^{13}CN^-$ and $^{13}CNO^-$ using a beam of fast-moving $^{13}C_{60}^+$ projectiles bombarding towards a solid substrate of ammonium nitrate (NH_4NO_3) at velocities >70 km/s. The uppermost panel shows mass spectra from experiments using unlabeled NH_4NO_3, while the middle panel and bottom panel are mass spectra from experiments using $NH_4{}^{15}NO_3$ and $^{15}NH_4NO_3$, respectively. The intensity on the y-axis is normalized to the total number of counts and reported as the average of the total number of measurements.

Figure 3.3. Mass spectra of ions generated in the vapor plume created via high power pulsed laser ablation on a solid target consisting of (A) $CaCO_3$ and $NaNO_3$ and (B) $CaCO_3$ and NH_4Cl powder mixture under near vacuum conditions (1.0×10^{-5} Pa). The intensity on the y-axis is the summation of 20 mass spectra recorded at a specific ion energy, between 1 to 20 eV; above 20 eV, the number of ions decayed exponentially. The 20 spectra were summed into a single spectrum to increase the signal to noise ratio and integrate ions of different energies, and then normalized to the total number of counts in the experiments. The background of each spectrum has been subtracted. The predictable peak shifts starting at m/z 26 (CN^-) and 42 (CNO^-) highlighted in both experiments demonstrate the recombination of C and N from solid. The raw data are accessible in the Apendix B.

Figure 3.4. Calibration curves between the number of detected CN^- ions and the weight percentage of N salts as starting materials in the laser simulation experiments. All extracted peak intensities, reported as counts per second, were first subtracted from the noise floor and then normalized by the 20 ms data acquisition time at each mass station. The intercepts of the linear regression curves were set to be zero. The errors of measurements are estimated using 1 relative standard deviation (1 RSD) of the noise floor. The shaded areas represent the predicted number of CN^- based on linear regression results at 1 SD. Outliers are identified in gray; the outliers in $NaNH_3$ and NH_4Cl experiments was likely due to incomplete dissociation of nitrate salt and contamination of other carbon- and nitrogen-rich residuals, respectively. The raw datasets from laser simulation experiments are available in the Appendix B.

Figure 4.1. Sample preparation and instrument setup at NASA Goddard Spaceflight center.

Figure 4.2. Single shot LDMS analysis of neat (a) SHD and (b) GDAE on flat Al plate shows detection of molecular ions at S/N < 100 but lack of diagnostic fragments for peptide *de novo* sequencing. Spectral intensities are normalized to the maximum peak intensity observed in each scan. The characteristics of all identified peaks are listed in Table C1. The raw transient data will be accessible in the Appendix C.

Figure 4.3. A schematic diagram that shows the nomenclature of fragments in a 3-mer peptide to enable *de novo* sequencing. The alpha carbon ($-C_n$) and the residues ($-R_n$) are highlighted in the peptide structure. A list of general types of molecular adducts and diagnostic fragments are also illustrated to demonstrate the identification and sequencing of a peptide in a single mass spectrum.

Figure 4.4. Single shot LDMS analysis of neat (a) SHD and (b) GDAE on a flat Al alloy plate with silicon nanoparticles demonstrates detection of various forms of molecular ion moieties and diverse diagnostic fragments that support peptide *de novo* sequencing. Spectral intensities are normalized to the maximum peak intensity observed in each scan. The characteristics of all identified peaks are listed in Table C1. The raw transient data will be accessible in the Appendix C.

Figure 4.5. Single shot LDMS analysis of peptide mixture AAG and SKSR obtained via LDMS (a) without nanoparticle and (b) with silicon nanoparticle shows detection of various forms of molecular ion adducts and diverse diagnostic fragments that support peptide *de novo* sequencing. Spectral intensities are normalized to the maximum peak intensity observed in each scan. Spectrum obtained without silicon nanoparticles was identified with a wider spread and more diverse forms of sequence ions of SKSR compared to analysis with silicon nanoparticles. The characteristics of all identified peaks are listed in Table C1. The raw transient data will be accessible in the Appendix C.

Figure 4.6. The pie charts to the right of each panel summarize the percent of total scans with at least one molecular ion adduct to identify SHD and GDAE without (top) and with (below) nanoparticles. The corresponding histograms show the frequency distribution of identified molecular ion adducts among *n* number of consecutive scans.

Figure 5.1. Schematics to demonstrate relationship of ion motion and ion distribution in phase domain to frequency domain.

Figure 5.2. The decay constant of identified ionic species in a single analysis of 200 μM oligopeptide KDA (top) and GDAE (bottom) in 0.1% formic acid solution. (**a, e**) A typical mass spectrum of KDA and GDAE, respectively. (**b-d, f-h**) The decay profile of each identified ionic species was extracted via StFFT from the single mass spectrum in panel (a) and (e), respectively. Each decay profile was fitted using an exponential function to derive decay constant c; the goodness of the fit was evaluated by R^2.

Figure 5.3. The decay constant of identified ionic species in a single analysis of a mixture of amino acids in water solution. (a) A typical mass spectrum of the solution mixture. (b-m) The decay profile of each identified species was extracted via StFFT from the single mass spectrum in panel (a). Each decay profile was fitted using an exponential function to derive decay constant c; the goodness of the fit was evaluated by R^2.

Figure 5.4. The decay constant of identified ionic species in a representative single scan of solution mixture (MRFA+GDAE+Ultramark1621). (a) A typical mass spectrum of the solution. (b-n) The decay profile of each identified species was extracted via StFFT from the single mass spectrum in panel (a). Each decay profile was fitted using an exponential function to derive decay constant c; the goodness of the fit was evaluated by R^2.

Figure 5.5. This figure shows the distribution of relative standard deviation (RSD) of the decay rates of each ionic species as a function of: (a) the RSD of CCS values; (b) the R^2 of their decay profiles; (c) the RSD of R^2, and (d) their average S/N in the mass

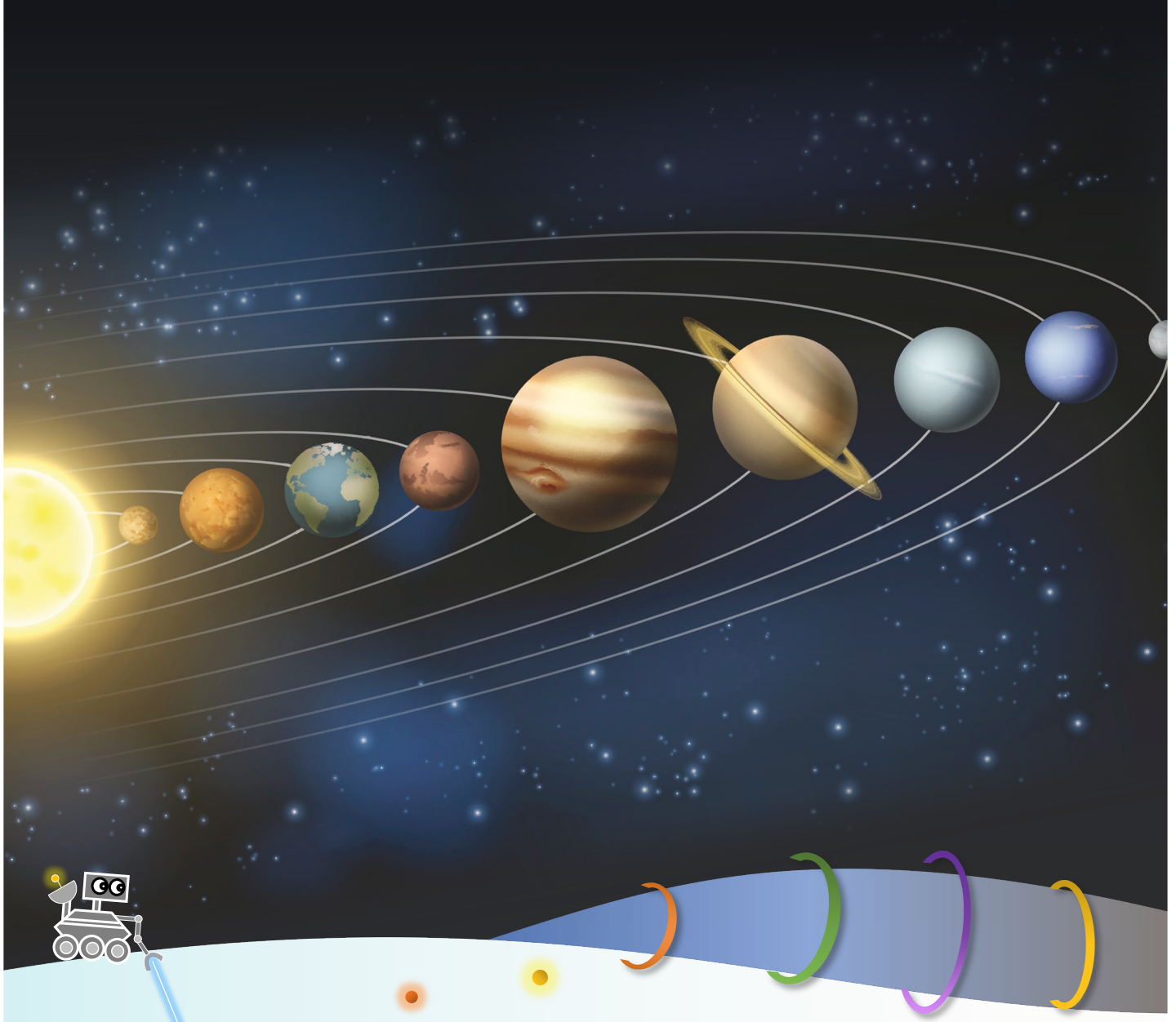
spectra. Threshold conditions are labeled to achieve CCS measurements with >99.5% reproducibility.

Figure 5.6. (a). A scatter plot and linear regression analyses show the CCS versus m/z distribution of ionic species identified in 14 consecutive scans of a solution mixture containing GDAE, MRFA, and Ultramark 1621. Ionic species from the same compound were grouped and characterized via a linear regression model. (b). Box plots show the distribution of slope of each compound obtained by individual scans (n = 14 total). Average values and standard deviation (SD) of the slopes are labeled.

Figure 5.7. (**a**) A typical mass spectrum of $[MRFA+H]^{++}$, $[GDAE+H]^{+}$, and $[MRFA+H]^{+}$ from the multiponent solution mixture. (**b**) The normalized decay profiles of three ionic species extracted from the mass spectrum reveal their modulated patterns in time domain. (**c-e**) The normalized peak intensities (to the maximum value of each ionic species) were plotted as a function of frequency shift at each segment of transient signals via StFFT. The frequency shift is calculated as the difference in frequency from current transient segment to the initial transient segment. (**f-h**) Mass spectra generated from each segment of transient signals via StFFT show the peak shapes and relative intensities of three ionic species through time. The color represents the sequential orders of each segment.

List of Abbreviations

AAG	Ala-Ala-Gly
BSE	Back scattered electron detector
CART	Classification and regression tree
CCS	Collision cross section
CE	Center electrode
CI	Carbonanceous chondrite
CID	collision-induced dissociation
CL	Cathodoluminescence
CORALS	Characterization of Ocean Residues and Life Signatures
Cp34H	Colwellia psychrerythraea strain 34H
DE	Detection electrode
ETD or ECD	electron transfer/capture dissociation
FFT	fast fourier transform
FMQ	Fayalite magnetite quartz mineral buffer
FT	Fourier Transform
FTICR	Fourier-transform ion cyclotron resonance
FWHM	Full Width Half Maximum
GDAE	Gly-Asp-Ala-Glu
HREE	Heavy rare earth element
LA-ICP-MS	laser ablation inductively coupled plasma mass spectrometer
LDI	laser desorption ionization
LDMS	laser desorption mass spectrometry
LREE	Light rare earth element
m/z	mass to charge ratio
MALDI	matrix assisted laser desorption mass spectrometry
MATLAB	Programming software
MFT	Mesa Falls Tuff
MOMA	Mars Organic Molecule Analyzer
MORB	Mid ocean ridge basalt
NASA	The National Aeronautics and Space Administration
REE	Rare earth element
RSD	Relative standard deviation
S/N	signal-to-noise ratio
SHD	Ser-His-Asp
SKSR	Ser-Lys-Ser-Arg
StFFT	short time fast fourier transform
TTG	Tonalite–trondhjemite–granodiorite
UMD	University of Maryland
UV	ultraviolet
UVPD	ultraviolet photodissociation



CHARACTERIZING

Chemical Signatures of Life via Mass Spectrometry

m/z

By Ziqin (Grace) Ni

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy, 2023

Chapter 1: Introduction

The question of whether life exists beyond Earth has driven the search for biosignatures within and beyond our Solar System. Chemical biosignatures relate to the characteristics, components, and structure of elements and compounds that provide evidence of past or present life. As knowledge of biological systems advances and deep space exploration continues, new chemical biosignatures will be proposed. Developing technologies and analytical strategies to detect these emerging biosignatures is a high priority in astrobiology. In this introduction, we will explore examples of chemical biosignatures, discuss evolving strategies for targeting these organic compounds, and review instrumentation capable of detecting and characterizing such materials during space missions.

1.1.Background

Essential building blocks of life, such as amino acids, lipids, and nucleic acids, are classical examples of chemical biosignatures. However, they have been found in carbonaceous meteorites and the interstellar medium (Callahan et al., 2011; Ehrenfreund & Charnley, 2000; Furukawa et al., 2019), suggesting that select organic molecules can be produced abiotically. Therefore, the detection of organic matter alone does not necessarily imply the presence of life.

Simple organic molecules can polymerize to produce macromolecules with more diverse constituents and complex structures. Of particular astrobiological interests, informational polymers (such as DNA and RNA) carry essential genetic codes that allow self-replication of molecules with increasing biological functionality under natural selection (e.g., Dorn et al., 2011; Summons et al., 2008). Organic compounds produced by living organisms on Earth are marked with specific structures and characteristics, including homochirality (Pace, 2001), even/odd carbon preferences (Dorn et al., 2011), and diagnostic isotopic signatures like $\delta^{13}C$ (Glavin et al., 2012). The abiotic to biotic transition, more specifically the evolution from pure chemical reaction of inanimate starting materials (i.e., abiotic synthesis) to autonomous synthesis and natural selection of informational polymers (i.e., biotic chemistry), sheds light on how and where life could have emerged on Earth and potentially on other habitable bodies. However, such a broad topic exceeds the scope of this singular dissertation.

1.2.Abiotic synthesis of organic compounds

Laboratory simulations have shown that simple organic compounds including amino acids can be produced abiotically with the assistance of sufficient energy sources, such as lightning (C. Chyba & Sagan, 1991), UV light (e.g., Botta et al., 2017) and/or impact events (e.g., (Ferus et al., 2019)). In particular, hypervelocity impacts may have played a critical role in delivering exogenous materials and enabling the synthesis of organic compounds in planetary environments (C. F. Chyba et al., 1990). The substantial energy released during a high-speed collision between two (or more) large planetary masses can be sufficient to reprocess C- and N-rich materials, including those stored in

the atmospheres and planetary surfaces, and generate new organic molecules in the post-impact plumes. Carbonates and nitrogen salts are two common inorganic matrices that have been observed on the surfaces of Mars (Niles et al., 2013) and Ceres (Carrozzo et al., 2018), and inferred to have existed on the early Earth (Morse & Mackenzie, 1998). Further research is necessary to investigate the potential for producing new molecules via crater-forming impacts using C- and N-rich geological matrices that could be widely accessible on planetary surfaces (***Chapter 3***).

Abiotically-produced organic compounds, even simple ones, can evolve chemically to produce larger, more complex macromolecules in planetary settings. For example, the refractory organic aerosols observed in Titan's atmosphere are likely polymerization products of simple HCN molecules originally produced via irradiation of a CH_4-N_2 dominated atmosphere with high energy photons sourced from the Sun and charged particles from the interstellar medium (Israël et al., 2005; Trainer et al., 2006). However, degradation and alteration of (macro)molecules are equally possible in harsh planetary environments under the exposure of ionizing radiation, such as that found on the surface of Europa (Blanco et al., 2018; Reisz et al., 2014). Therefore, the abundance and diversity of organic materials in planetary environments can be modulated due to abiotic synthesis and polymerization versus destructive physicochemical processes. As a result, a significant amount of research has been devoted to i) laboratory experiments and numerical simulations to investigate diverse planetary processes; ii) characterizing the organic inventory of meteorites to understand abiotic modes of molecular synthesis; and, iii) remote observations and in situ exploration of extraterrestrial environments to better understand boundary conditions of potentially habitable niches. A combination

of approaches helps to identify viable mechanisms to produce/destroy organic molecules and establish the "abiotic blank" of organic materials in planetary targets.

1.3.The redox condition of planetary environment

The redox state of planetary environments is crucial for abiotic synthesis of organic molecules. Laboratory experiments show that intense shock produced by meteors entering a thick atmosphere generated 10^8 times more organic material in a reduced atmosphere (i.e., 1×10^9 kg/yr) than in an oxidized atmosphere (3×10^1 kg/yr) (C. Chyba & Sagan, 1992). Additionally, local redox conditions regulate the valence state and bioavailability of essential nutrients (CHNOPS) and trace elements, which impacts the thermodynamic and kinetic behavior of multivalent elements in chemical reactions, influencing the possibilities for further molecular evolution.

The redox condition of magmatic environments can be probed via trace element systematics (e.g., Zn/Fe systematics in arc basalt (Lee et al., 2010)) and/or isotopic ratios (e.g., V isotope signatures in marine sediments (Wu et al., 2020)). Zircon, one of the most resilient mineral phases in Earth's history, can preserve unaltered chemical signatures related to paleoenvironments up to 4 billion years ago. By analyzing the abundances of multivalent elements such as Ce (3+, 4+) and Eu (2+, 3+) relative to other monovalent rare earth elements (REE, 3+ only), the local magmatic redox state can be inferred from trace element analysis of zircon grains, providing invaluable means to study and estimate planetary redox conditions on Earth and beyond (***Chapter 2***).

1.4.The evolving strategies to search for chemical signs of life

The active engagement of scientists from diverse fields has infused new knowledge into the frontier of astrobiology, expanding definitions and examples of chemical biosignatures. For instance, cold- and salt-loving psychrophiles collected from the Arctic Ocean and exposed to increasingly cold, saline, and nutrient-depleted conditions in the laboratory have shown an enrichment of a specific subset of oligopeptides composed of three or four amino acids (Mudge et al., 2021). These specific “3-mers” and “4-mers” could represent adaptive metabolic strategies evolved to enable these microbes to survive icy brines on Earth and serve as putative biosignatures of psychrophiles in habitable ocean worlds. ***Chapter 4*** focuses on the detectability of these putative biosignatures using a spaceflight prototype instrument, demonstrating the capability to search for targeted chemicals with specific functionalities in near-term missions to ocean worlds.

One strategy to improve the search for biosignatures is to prioritize measurements of chemical biosignatures with a lower likelihood of being produced abiotically or delivered externally, in other words, higher fidelity biosignatures. For example, in 2018, Neveu et al. proposed a prioritization scheme called the "ladder of life detection" that outlines the requirements of astrobiological payloads (Neveu et al., 2018). Another approach, as suggested by Chan et al. in 2019, emphasizes the importance of searching for chemical and physical biosignatures across a range of spatial scales, from the atom and molecule to morphological patterns (Chan et al., 2019). By considering

observations from multiple spatial scales, geological context can be provided for putative signs of biology, providing corroborative measures of biogenicity.

As life detection strategies continue to evolve, more advanced mission-enabling technologies and innovative analytical techniques will be required. In particular, highly sensitive (low limit of detection) instruments need to be developed to search for low abundance organic molecules in limited energy environments, such as the subsurface oceans observed on Enceladus, Europa, and Titan (to name a few). Reproducible (high precision) measurements, and corroborative observations of the structure, abundance, and isotopic ratios of organic molecules will be essential to molecular identification and to distinguish abiotic versus biotic sources. Reliable protocols to make *in situ* investigations and to report measurements of prospective chemical biosignatures need to be standardized and implemented in future space missions.

1.5.Mass spectrometry in space exploration

Mass spectrometry is an analytical technique that identifies isotopes, elements, or chemical compounds present in a sample via mass measurements. A mass spectrometer typically comprises three key subsystems: an ionization source, which generates ions from the analyte introduced into the instrument; one or more analyzers, which separate the ions based on their respective mass to charge ratios (m/z); and a detector, which measures the signal intensity of each measured m/z (**Figure 1.1**).

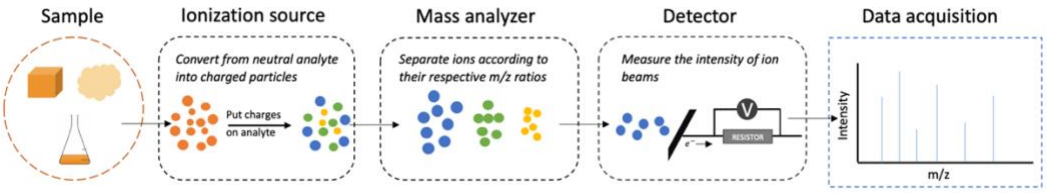


Figure 1.1. A mass spectrometer consists of three crucial subsystems: the ionization source, the mass analyzer, and the detector. Together, these subsystems perform essential functions that produce mass spectra, which are used to identify and quantify unknown samples. Adapted from Arevalo et al. (2020).

Designing a mass spectrometer for spaceflight applications requires more effort and time than designing a benchtop system due to additional considerations, such as miniaturization to meet limited reduced size, weight, and power (SWaP) resources. These considerations include the need to: i) ruggedize the instrument to withstand intense vibrations during launch and deployment; ii) ensure its qualification to endure thermal cycling and radiation levels expected in the targeted planetary environment; and, iii) confirm its ability to meet all functionality standards and performance requirements that align with the mission's scientific objectives.

Developing and testing a new technology for spaceflight mass spectrometers can take decades, so incremental modifications are often made to heritage designs to manage risk, budget, and time (**Figure 1.2**). However, whenever a new spaceflight technology is developed and matured, such as a new ionization source (**Session 1.5.1**) or mass analyzer (**Session 1.5.2**), the comprehensiveness of chemical investigation is significantly improved to answer more challenging scientific questions. Laser desorption mass spectrometry (LDMS) and Orbitrap MS are particularly promising techniques for characterizing chemical signatures of life in future missions.

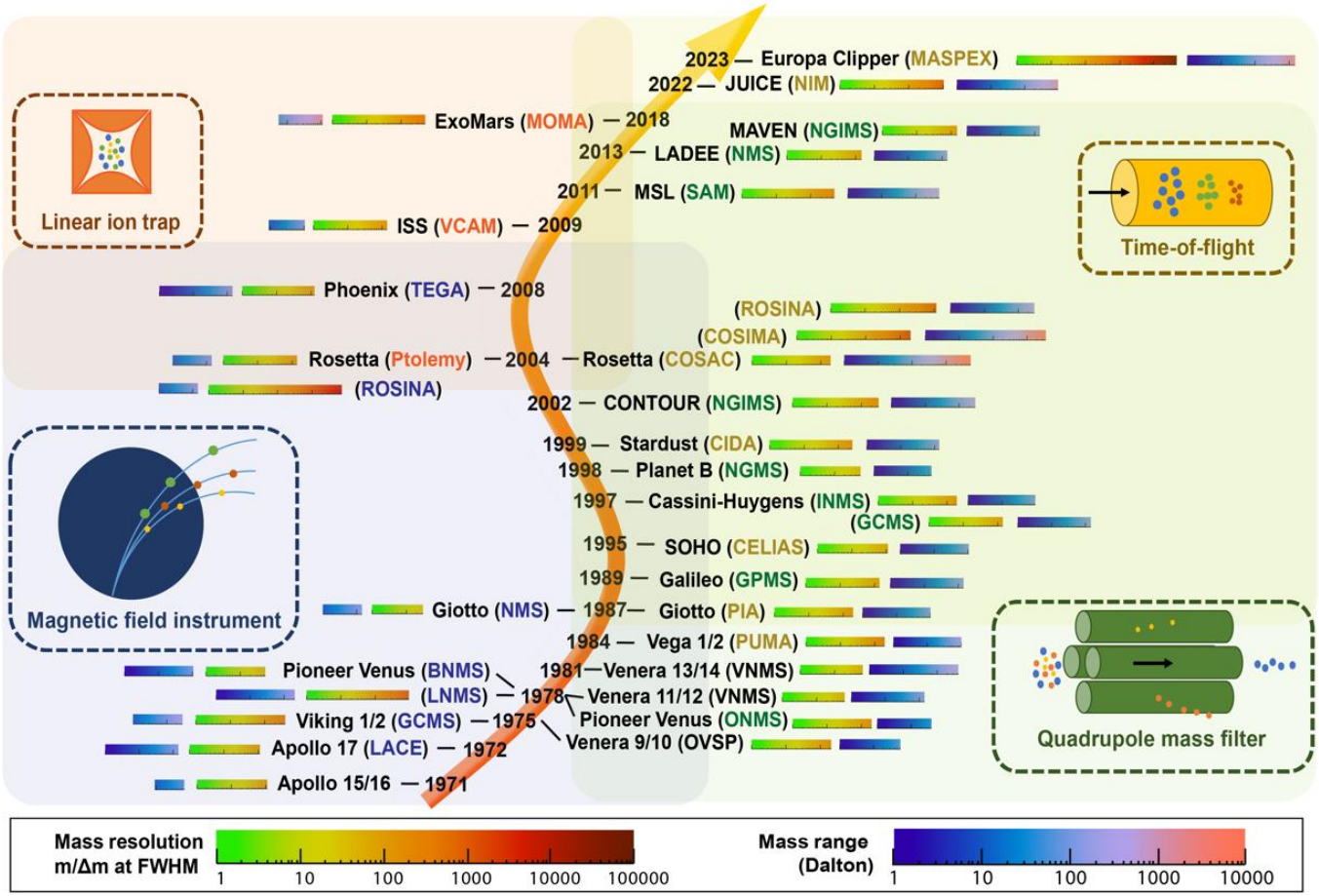


Figure 1.2. A flow chart that summarizes the evolution of mass spectrometers for planetary exploration since 1971. These heritage instruments are grouped by their respective mass analyzer type, as well as the transition from instruments with limited analytical capabilities to those with increased scientific output (e.g., wider mass ranges, greater mass resolving power, etc.). Adapted from Arevalo et al. (2020).

1.5.1. Laser desorption ionization mass spectrometry: advantages and challenges

Since the Viking space program in 1970s, pyrolysis/gas chromatography mass spectrometry (pyrolysis/GC-MS) has been one of the dominant means to characterize planetary surfaces. This method is capable of extracting volatile molecules via evaporation of bulk samples and decomposing refractory substances through pyrolysis (temperatures > 200 °C; Biemann, 1974). However, this approach is limited to the measurement of volatile organic materials. More refractory organics, such as progressively larger/more complex macromolecules, are susceptible to thermal degradation during pyr/GC-MS analysis.

Laser desorption mass spectrometry (LDMS) has gained increasing attention in space exploration. A high-power laser can directly sample and ionize refractory organic compounds (with limited fragmentation) (**Figure 1.3**). A focused laser beam allows for in situ measurements of sample composition with high spatial resolution, empowering chemical mapping and recognition of morphological patterns characterized by distinct isotopic/elemental/molecular spatial distributions.

The traditional approach to ionize nonvolatile organic molecules via LDMS is to admix samples with UV-absorbing, proton-rich organic matrices, a technique known as matrix-assisted laser desorption ionization (MALDI) mass spectrometry. Upon laser irradiation, UV-absorbing matrices and non-absorbing analytes are desorbed from the surface due to shock and rapid heating at the site of irradiation, generating a supersonic plume. The exact ionization mechanisms of analyte during LDMS remain controversial

(for example, Bae & Kim, 2015; Zenobi & Knochenmuss, 1998). One commonly accepted hypothesis state that protons could be directly transferred from the MALDI matrices to neutral analytes in the plume. Alternatively, free protons are generated due to thermal fragmentation of MALDI matrices and/or analytes, which enable ion formation. Alkali metals such as Na^+ and K^+ that released from the substrate could also combine with neutral analytes in the plume to generate metal adducts. A wide range of processes have been proposed to explain ion formation via LDMS.

The incorporation of organic MALDI matrices poses a risk of forward contamination to planetary environment. However, without an organic matrix, macromolecules such as 3-mer and 4-mer oligopeptides are more prone to degradation during laser irradiation, making their detection and identification more challenging. To address this issue, alternative substrates, such as inorganic nanoparticles, need to be explored as pathways to increase ionization efficiency, preserve molecular ion signals, and improve detection limits of low abundance complex molecules via LDMS techniques (***Chapter 4***).

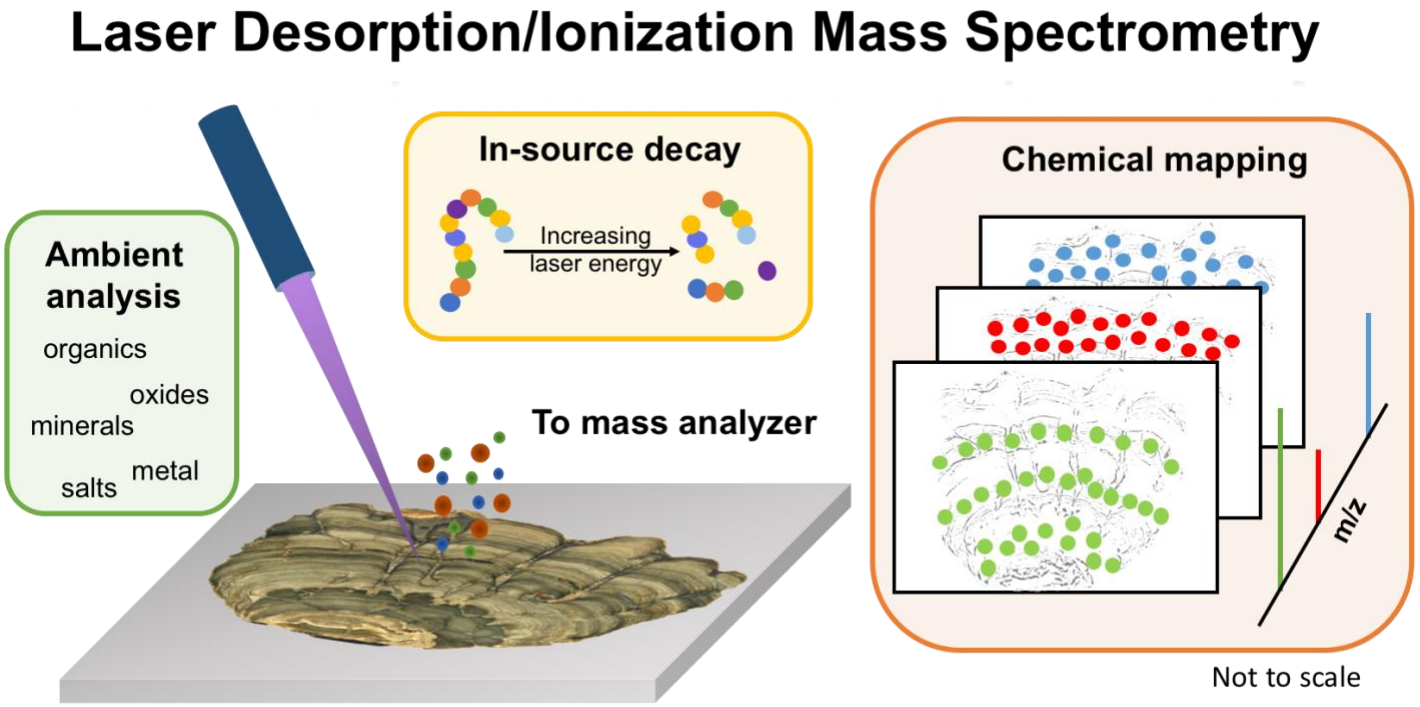


Figure 1.3. A schematic diagram that summarizes select analytical advantages of LDMS. LDMS techniques enable ambient chemical analysis of a broad range of materials, while fragmentation of molecules in the ionization source, or in-source decay, provides valuable insights into molecular structure. Additionally, chemical mapping can be used to characterize isotope/elemental/molecular patterns and recognize morphological features through spatially resolved analyses.

1.5.2. Orbitrap mass spectrometry: advantages and challenges

Mass spectrometers that have been integrated into deployed and/or planned spaceflight missions include magnetic sector instruments (e.g., Viking 1/2 GC-MS), quadrupole mass analyzers (e.g., Pioneer Venus), time-of-flight systems (e.g., Rosetta COSAC) and linear ion traps (e.g., ExoMars MOMA) (Arevalo et al., 2020). The evolution of these analyzers has resulted in progressive improvements in performance metrics, most notably mass resolving power, mass accuracy, mass range, and sensitivity (**Figure 1.4**). The next-generation Orbitrap mass analyzer originally developed for commercial applications (Makarov, 2000) was recently adapted for spaceflight applications (Arevalo et al., 2023). Compared to heritage spaceflight mass spectrometers, Orbitrap mass analyzers offer ~200× improvements in mass resolving power ($m/\Delta m \geq 10^5$, Full Width Half Maximum [FWHM]) and ~100× more accurate peak assignments relative to the MOMA instrument onboard the ExoMars Rosalind Franklin rover (Goesmann et al., 2017). Such high mass resolution and accuracy together enable the determination of molecular stoichiometry of organic molecules and geologic phases via exact mass measurements without the requirement for extensive sample purification/preparation in the laboratory.

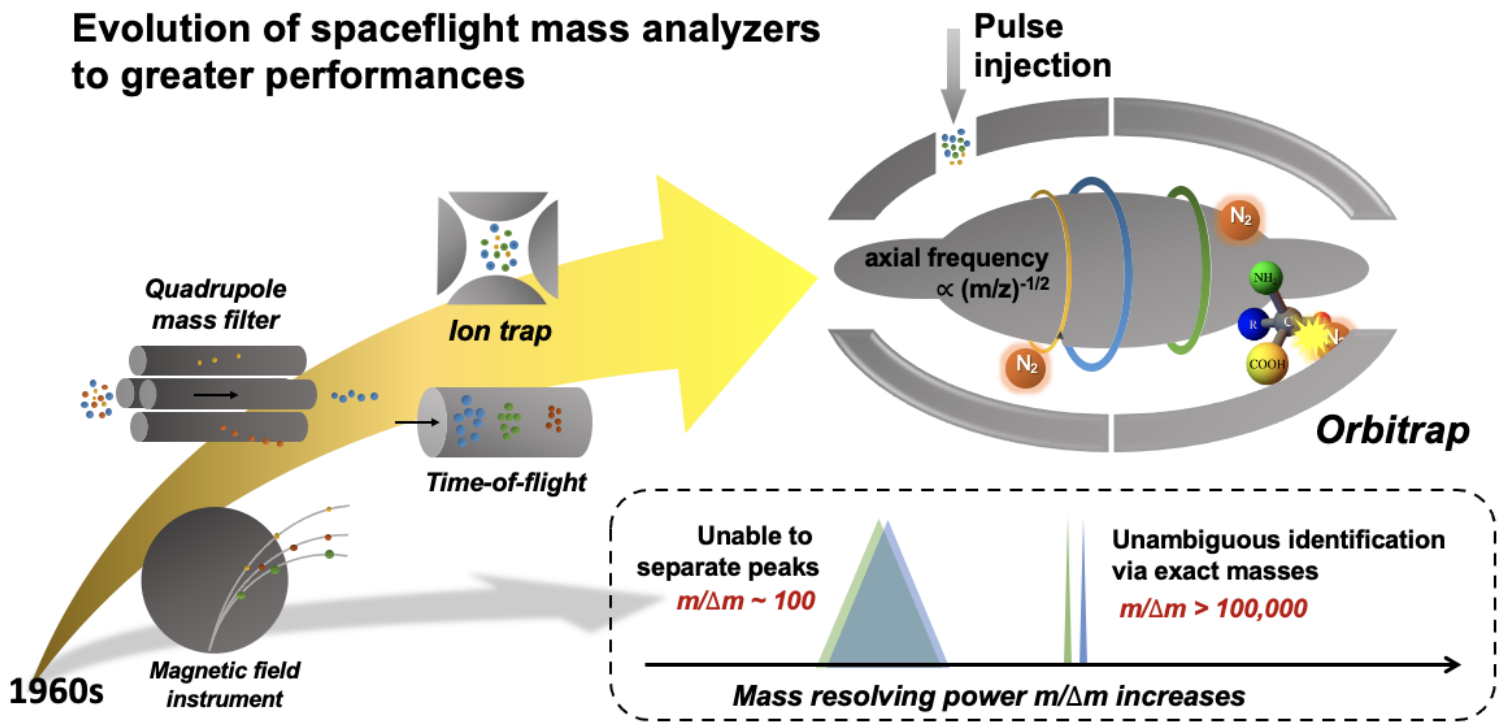


Figure 1.4. An overview of the evolution of mass analyzers in spaceflight mass spectrometers since the 1960s. The next generation Orbitrap analyzer demonstrates unparalleled mass resolution that allows unambiguous identification of the stoichiometry of analytes via exact mass determination.

The analyzer consists of a central spindle-like electrode, a pair of barrel-like outer detection electrodes, and an outer electrode next to the ion injection slit. The center and outer electrodes are shaped such that the quadro-logarithmic trapping electric potential is created inside the trap. During data acquisition, ion packets would be injected into the device via a slot located in one half of the outer electrode and deflected with assistance from a small deflection electrode situated nearby. During ramping of high voltage that is applied to the central electrode (e.g., 3.5 kV), ions are "squeezed" towards the center electrode, a process known as electrodynamic squeezing. and initiate ion oscillation at frequency proportional to $(m/z)^{-1/2}$ (Makarov, 2000). A pair of outer electrodes would record the signal intensity as ions moving closer or further away from each electrode, a process known as image current detection. This differential current will be amplified, digitized, and output as signal as a function of time, or transient signal (**Figure 1.5)**.

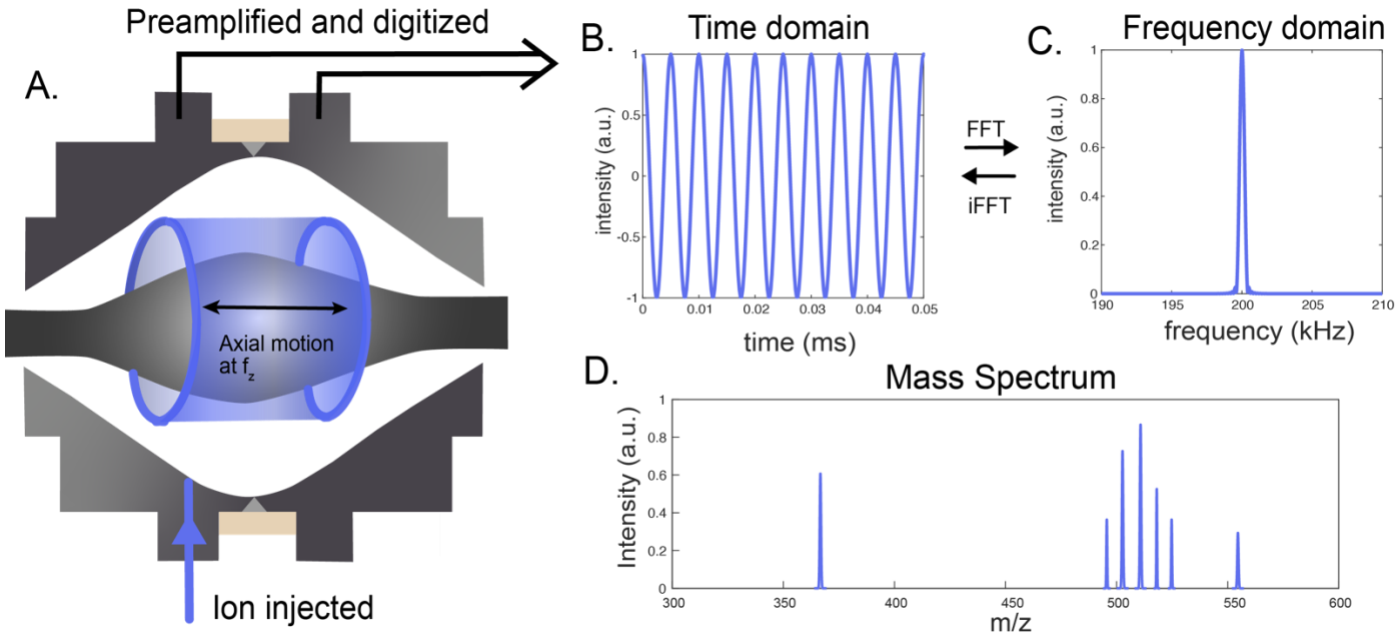


Figure 1.5. Schematics to demonstrate the relationship between ion motion and ion distribution in phase domain to frequency domain.

Following conventional high-speed data processing procedures, Fourier Transform (FT) is applied to convert the transient signal into a frequency spectrum, which is then calibrated using internal standards to produce a mass spectrum for molecular identification and quantification. This study applies fast fourier transform (FFT), a special variant of FT algorithm that occurs when the length of transient signal N is a power of 2. To further improve spectral accuracies, zerofillings and hanning windows are applied prior to FFT processing. A 10x zerofillings mean to add $10N$ number of zeros at the end of a transient signal. Additional y number of zeros is added to reach the nearest next power of 2 number to enable FFT. Reduction in spectral spacing allows for a better capture of peak apex and area, which in turn can be used to improve spectral accuracies. FFT processing of discrete transient signals also generates sidebands that extend many frequency bins away from the frequency peak apex. Sidebands can interfere with nearby low-intensity peaks, distort peak shapes, and confuse peak assignments. Hanning windows (an apodization function) are thus applied to suppress sidebands. All transient signals for this dissertation were processed in a custom-built MATLAB script.

In studies of complex sample mixtures, mass measurements remain insufficient to recognize and determine the molecular structure, configuration, and function of organic molecules. Recent studies show that transient signals obtained through Orbitrap MS decay proportionally to the collision cross section (CCS) of large ionic species (such as proteins), indicating the compound's conformation, related to its composition, size, and structure (James et al., 2022; Sanders et al., 2018a). However, it is unclear whether

CCS values of smaller molecules, such as amino acid monomers, can be determined using similar methods. Thus, further research is needed to demonstrate the simultaneous measurement of CCS and m/z of smaller molecules via Orbitrap MS (***Chapter 5***).

1.5.3. CORALS: a laser-orbitrap spaceflight prototype

The Characterization of Ocean Residues and Life Signatures (CORALS) investigation focuses on an advanced LDMS instrument specifically developed for the exploration of Europa and other ocean worlds (Arevalo et al., 2018, 2023; Briois et al., 2016; Willhite et al., 2021). The CORALS instrument is composed of a solid-state laser system that emits up to 450 uJ of 266 nm radiation (Fahey et al., 2020) and a high-performance Orbitrap analyzer that has been ruggedized to survive the mechanical and thermal stresses projected for an astrobiological mission to the outer planets. The precise and accurate measurement of m/z enabled by the Orbitrap analyzer can efficiently separate ions of similar m/z and allows the determination of chemical constituents of unknown molecules. The CORALS instrument has demonstrated the capability to detect and sequence 3-mer and 4-mer oligopeptides without introducing organic matrices (***Chapter 4***), thereby representing an emerging technology for planetary exploration and a pathfinder for advanced technique development for astrobiological objectives.

1.6.Thesis outline

This thesis contains four enumerated chapters with distinct but inextricably linked themes on the topics of planetary redox condition, abiotic synthesis of organic

compounds, and innovative life-detection measurements. The focus of these chapters is on the development of novel analytical methods using mass spectrometry to investigate planetary environments and characterize potential chemical signs of life. The mass spectrometry techniques developed here have broader implications for the field of mass spectrometry as a whole, as they offer alternative approaches to detecting and characterizing organic molecules using readily available technologies. Each chapter of the thesis is summarized below:

- **Chapter 2** is focused on the redox condition of planetary environments; a multidimensional filtering scheme was developed to identify unaltered "out of context" zircon grains up to 4 Ga and infer the local redox state using trace element systematics (*Published*).
- **Chapter 3** is focused on abiotic synthesis via hypervelocity impacts; impact-generated vapor plumes were simulated via a high-power laser irradiation and C_{60}^{+}physical bombardment of a solid substrates to quantify production rates of CN^{-} and CNO^{-} and inform on abiotic synthesis on early Earth, Mars, and Ceres (*Under review*).
- **Chapter 4** is focused on the development of novel methods for life detection objectives; an LDMS technique demonstrated the feasibility of detecting and sequencing 3-mer/4-mer peptides with and without silicon nanoparticles (an alternative to traditional organic matrices) using the CORALS prototype instrument (*Published*).
- **Chapter 5** is focused on the derivation of collision cross-section (CCS) via Orbitrap MS; a new data processing workflow enabled simultaneous

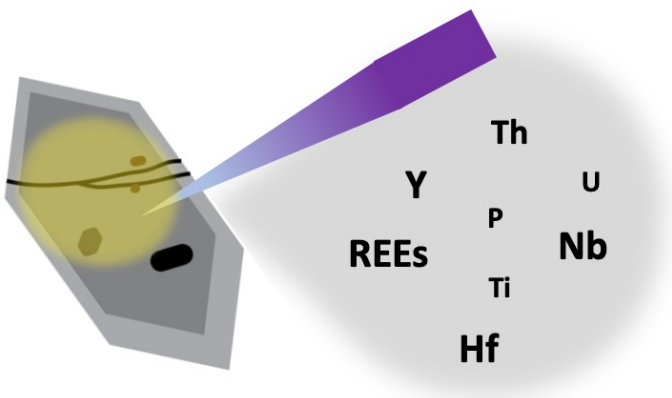
measurements of mass and CCS, delivering a corroborative measure of molecular identify and providing insights into ion motion inside the Orbitrap analyzer (*In preparation*).

Finally, the ***Conclusion*** chapter summarizes the key results of these studies, highlights the overarching impact of the findings reported, and proposes potential future directions for research in the field of life detection. ***Publications*** contains a list of publication records during the graduate studies. ***Appendix A-D*** contains supplementary materials to each chapter.

A novel approach to identifying mantle-equilibrated zircon by using trace element chemistry

Ziqin Ni[1], Ricardo Arevalo Jr.[1], Philip Piccoli[1], Barry Reno[2]

[1]University of Maryland, College Park, MD, 21401
[2]Northern Territory Geological Survey, Darwin, NT, 0800



Source melt petrogenesis

Zircon saturation context

Mineral inclusions

Chapter 2. A novel approach to identify mantle-equilibrated zircon by using trace element chemistry

Quick overview of the chapter:

Zircon is a robust accessory mineral commonly found in many rock types. The Ce/Ce and Ti content in zircon provide valuable insights about the redox states of local magmatic environment from which it crystallized. When it comes to study of mantle redox state via zircon Ce/Ce* oxybarometry and Ti-in-zircon thermometry, one of the requirements is to select mantle-equilibrated zircon (i.e., unaltered magmatic zircon without interaction to hydrosphere). Traditionally, such selection relies on comprehensive examinations of zircon and their host rock (i.e., whole rock geochemistry, zircon morphology, texture, and trace element and isotopic measurements) in order to obtain information for zircon saturation context, source melt petrogenesis, and post-magmatic history. This study, however, presents a simple filter scheme that facilitates the selections of mantle-equilibrated zircon and provides quick insights about zircon petrogenesis using trace element systematics alone (P, Ti, Y, Nb, REE, Hf, Th, U).*

In this chaper, I compiled and processed the dataset, interpreted the results, conceived the ideas, made the figures, and wrote the paper. This chapter has been published.
Ni, Z., Arevalo Jr, R., Piccoli, P., & Reno, B. L. (2020). A Novel Approach to Identifying Mantle-Equilibrated Zircon by Using Trace Element Chemistry. Geochemistry, Geophysics, Geosystems, 21(11), e2020GC009230.

2. Abstract

One of the requirements for inferring local mantle redox states via zircon Ce/Ce* oxybarometry and Ti-in-zircon thermometry is to select mantle-equilibrated zircon (i.e., unaltered magmatic zircon without interaction to hydrosphere). Traditional protocols for identifying mantle-equilibrated zircon require comprehensive examination of whole rock geochemistry in addition to zircon morphology, texture, and trace element and isotopic measurements to obtain information for zircon saturation context, source melt petrogenesis, and post-magmatic history. This study proposes a simple filter scheme for selecting mantle-equilibrated zircon using trace element systematics alone (P, Ti, Y, Nb, REE, Hf, Th, U). A total of 13 filtering criteria are synthesized from previous studies based on simulations, experiments, and compiled global datasets, all of which help to provide geological context and ultimately constrain the inferred melt redox state from three perspectives: (1) mineral inclusions; (2) source melt petrogenesis; and, (3) zircon saturation context. The filter scheme presented here, which is based on detailed classifications of zircon morphology/texture and host rock compositions of 2317 zircon analyses from 30 independent references, is shown to distinguish non-magmatic zircon (Group III), magmatic zircon with significant inclusions and/or sourced from highly enriched source melt (Group II), and mantle-equilibrated zircon (Group I). The filter scheme is validated with known, well-characterized mantle-equilibrated Jack Hill zircon (n=53); the selected mantle-equilibrated zircon are supported by BSE/CL imaging and oxygen isotopic signatures with an average ΔFMQ 1.5 ± 1.3 (sd; n=7), in agreement with previous study. Further, a case study of using this filter scheme to facilitate zircon research is demonstrated. Without contextual information about zircon

samples, application of filter scheme has selected mantle-equilibrated zircon (n=1), implying a mantle redox state ΔFMQ -0.5 since 2950 Ma. Future applications of this filter scheme include studies of out-of-context detrital and/or xenocrystic zircon.

2.1. Introduction

Magmatic fO_2 is an important control on elemental behavior, such as solubility (e.g., soluble Fe^{2+} versus insoluble Fe^{3+}), mobility (e.g., mobile UO_4^+ versus immobile U^{4+}), chemical affinity (e.g., siderophile S at low fO_2 versus lithophile S at high fO_2), and compatibility (e.g., compatible Eu^{2+} versus incompatible Eu^{3+} in plagioclase). Understanding the oxygen fugacity of the upper mantle in the early Earth is particularly valuable for quantitatively describing processes associated with planetary differentiation (e.g., Frost et al., 2008; Frost and McCammon, 2008; Yang et al., 2014) and the evolution of atmospheric composition (e.g., Kump et al., 2001, Trail et al., 2011; Lee et al., 2014).

Zircon ($ZrSiO_4$) is a resilient accessory mineral capable of surviving multiple igneous or metamorphic events, weathering and transport, hydrothermal alteration, etc. In the zircon crystal lattice, $^{VIII}Zr^{4+}$ may be exchanged for a range of incompatible trace elements, which provides valuable insights about the local magmatic environment from which it crystallized. For example, studies of zircon chemistry have been used to infer: melt crystallization temperatures based on Ti concentrations in zircon (e.g., Watson and Harrison, 2005; Ferry and Watson, 2007; Fu et al., 2008; Siégel et al., 2018); the timing of crystallization and/or metamorphism from U-Th-Pb radiometric dating (e.g., Zheng et al., 2004; Tichomirowa et al., 2013); the provenance of source material from

concentrations of LILEs and HFSEs (e.g., Hoskin and Ireland, 2000; Belousova et al., 2002; Grimes et al., 2007, 2015; Chapman et al., 2016); and, the degree of crustal assimilation in magmatic systems based on coupled O-Hf isotopes (e.g., Hawkesworth and Kemp, 2006; Trail et al., 2007; Page et al., 2007; Cavosie et al, 2009; Spencer et al., 2017).

Of particular importance, magmatic oxygen fugacity (or fO_2) has been inferred from the abundances of multivalent elements, namely Ce (3+, 4+) and Eu (2+, 3+), relative to other monovalent REE (3+ only). In zircon, REE^{3+} can replace Zr^{4+} in a coupled substitution with additional cations (e.g., P^{5+}) to balance charge. The compatibilities of trivalent REE in zircon increase systematically from La to Lu due to progressively smaller ionic radii with increasing atomic number (commonly termed the lanthanide contraction; **Figure 2.1**).

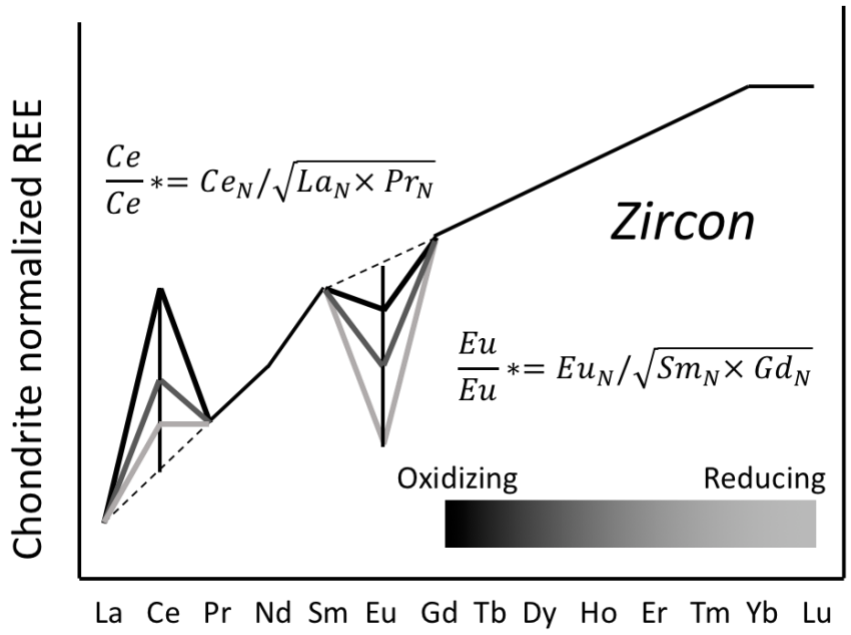


Figure 2.1. Schematic chondrite normalized pattern of REE in zircon. Ce/Ce* and Eu/Eu* are calculated as $Ce_N/\sqrt{La_N \times Pr_N}$ and $Eu_N/\sqrt{Sm_N \times Gd_N}$ respectively, where the subscript N indicates trace element values normalized to CI chondrites (McDonough and Sun, 1995). As the oxygen fugacity of the system increases (darker shade of gray), Ce/Ce* is expected to increase whereas Eu/Eu* is expected to decrease. The majority of terrestrial magmatic zircon will have both positive Ce/Ce* and negative Eu/Eu*.

However, under oxidizing conditions, a greater proportion of Ce is present as the higher Ce^{4+} valence state, resulting in preferential incorporation into the zircon structure (relative to Ce^{3+}) due to its smaller ionic size and equivalent charge to Zr^{4+}. This substitution bias results in a "positive spike" of Ce relative to nearby trivalent elements La and Pr in normalized composition diagrams, enabling the quantification of a Ce anomaly as: Ce/Ce* = $Ce_N/\sqrt{La_N \times Pr_N}$. In comparison, under reducing conditions, a greater proportion of Eu takes on the lower Eu^{2+} valence state. Because Eu^{2+} is too large to fit into the zircon structure without incurring significant lattice strain, Eu exhibits a "negative trough" relative to neighboring trivalent Sm and Gd, defining a Eu anomaly as: Eu/Eu*= $Eu_N/\sqrt{Sm_N \times Gd_N}$. As a compounding effect, fractionation of plagioclase also contributes to Eu/Eu* in zircon because Eu^{2+} is highly compatible to plagioclase structure. Consequently, Ce/Ce* (and to a lesser extent Eu/Eu*) in magmatic zircon can be used as a proxy for magmatic fO_2.

2.1.1. Quantitative zircon Ce/Ce* oxybarometry and Ti-in-zircon thermometry

The Ce/Ce* recorded in zircon has been experimentally calibrated to fO_2 of silicate melts at various alumina and water concentration, such as anhydrous peralkaline melts at high temperature (i.e., 1300 ℃) (Burnham and Berry, 2012; Trail et al., 2011 & 2012; Smythe and Brenan, 2015 & 2016; Loucks et al., 2020). However, statistically, the majority of natural zircon crystallizes in hydrous alumina-saturated melts (i.e., determined by the molar ratio of $Al_2O_3/(Na_2O + K_2O)$, or A/NK > 1) at relatively lower temperatures (e.g., ~ 800 ℃). Therefore, the empirical relationship that most

universally describes Ce partitioning behavior in natural zircon is from Trail et al. (2012),

$$\ln\left(\frac{Ce}{Ce^*}\right)_{CHUR} = (0.1156 \pm 0.0050) \times ln(fO_2) + \frac{13860 \pm 708}{T(K)} - (6.125 \pm 0.484)$$

Eqn (2.1),

where the experiments were conducted at 10 kbar and 800 – 1300 ℃ with 2 to 10 wt% H_2O (Trail et al., 2012). The zircon crystallization temperature T in Eqn (2.1) can be inferred from the Ti concentration in zircon following the quantitative relationship below (Ferry and Watson, 2007):

$$log\ (\ Ti\ in\ zircon\ ppm) = (5.711 \pm 0.072) - \frac{4800 \pm 86}{T(K)} - \log\ (a_{SiO_2}) + log(a_{TiO_2})$$

Eqn (2.2).

The activity of SiO_2 (a_{SiO_2}) and TiO_2 (a_{TiO_2}) are dimensionless quantity that describes the chemical potential of SiO_2 and TiO_2 of the system relative to quartz and rutile, respectively. Ce/Ce* oxybarometry and Ti-in-zircon thermometry are often applied together to provide a streamlined estimation of melt redox state, which are often reported as orders of magnitude deviations from the fO_2 of mineral assemblage fayalite-magnetite-quartz (FMQ) at a specific temperature, i.e., ΔFMQ (Chou, 1978).

2.1.2. Challenges at selecting mantle-equilibrated zircon

Inferred magmatic fO_2 from mantle-equilibrated zircon provides valuable insights into mantle oxygen fugacity. Mantle-equilibrated zircon refer to unaltered magmatic zircon that crystallized from source melt without interaction with the hydrosphere and/or assimilation of exogenous materials (e.g., sediments, crustal materials). However, such mantle-equilibrium state is often violated by many geological processes which summarized to be three main types: (1) assimilation of melt from multiple sources (mantle magmas, assimilated crustal rock, etc.); (2) dynamic growth conditions (disequilibrium crystallization, mineral decomposition during metamorphism, etc.); and/or (3) modification of zircon trace element concentration via post-magmatic processes (subsolidus crystallization, hydrothermal alteration, etc.). These processes diversify the composition of source melt, and change trace element signatures in zircon syn- and/or post- zircon saturation, resulting an unrealistic spread in inferred magmatic fO_2 that extends below conditions expected at the core mantle boundary (ΔFMQ < -4; McCammon, 2005) to those found in the modern atmosphere (ΔFMQ > +10; McCammon, 2005). Therefore, selection of mantle-equilibrated zircon is necessary to obtain meaningful mantle oxygen fugacity.

Evaluation of zircon mantle-equilibration state traditionally requires a number of analytical protocols including (but not limited to):

- determination of whole rock chemistry to estimate the source melt composition, calculate apparent zircon/melt partition coefficient for Ce, infer the

crystallization behavior (equilibrium versus disequilibrium) and growth environment of zircon;

- petrographic examination and/or chemical mapping (e.g., via back-scattered electron (BSE) or cathodoluminescence (CL) imaging) to characterize zircon morphology and texture, identify mineral or fluid inclusions, evaluate intragranular and intergranular chemical and isotopic variation based on zoning, and infer petrogenetic history (Corfu et al., 2003; reference therein);
- measurement of isotopes in individual zircon crystal. For example, U-Pb dating to understand the age population of zircon in host rock, and oxygen isotopes to screen zircon that crystallized from source melt with negligible interaction to hydrosphere, and within the mantle-equilibration zone ($\delta^{18}O$ 5.3 ± 0.3 ‰, sd) (Valley et al, 1998); and,
- chemical filtering to exclude zircon that carry trace element signatures indicative of mineral inclusions (e.g., apatite, etc.; Bell et al., 2019) and non-magmatic features (i.e., hydrothermal and metamorphic zircon; Hoskin & Black, 2002; Hoskin, 2005).

These requirements are highly recommended to obtain a thorough understanding of zircon saturation conditions, constrain statistical variances, and add confidence to inferred melt redox states. For instance, the application of these methods has enabled the demonstration that mantle-equilibrated zircon center around the FMQ buffer (e.g., Trail et al., 2011). However, such strict protocols also reduce the size of viable data sets and ultimately undermine the capability/advantage of the method by requiring

extensive analyses that may not be feasible given financial (cost), programmatic (time), and/or logistical (sample volume/size) limitations. For example, in situ measurements of oxygen isotope in zircon require Secondary Ion Mass Spectrometry (SIMS), which are not only expensive but also demand a critical size of zircon for in situ analyses (Valley, 2003).

2.1.3. Purpose of this study

The purpose of this study is to establish a multidimensional filter scheme that facilitates the selection of mantle-equilibrated zircon based only on a narrow range of trace element systematics. Instead of relying on extensive contextual observations (e.g., petrography, whole rock compositions, and oxygen isotope ratios), trace element abundances and dynamics are used to isolate zircon that have experienced diverse geological processes (e.g., inclusions, subsolidus crystallization, crystallized from melt with potential interaction to hydrosphere etc.), so that the combined filters can screen effectively for mantle-equilibrated zircon (**Figure 2.2**). These chemical criteria are first justified based on previous published review studies, experiments, and simulation, and then evaluated objectively by applying the methods developed here to robust, publicly available zircon databases. Zircon with the greatest compositional similarity to the canonical trace element chemistry of mantle-equilibrated zircon should theoretically have the highest probability of recording (and preserving the signature of) the local mantle redox state.

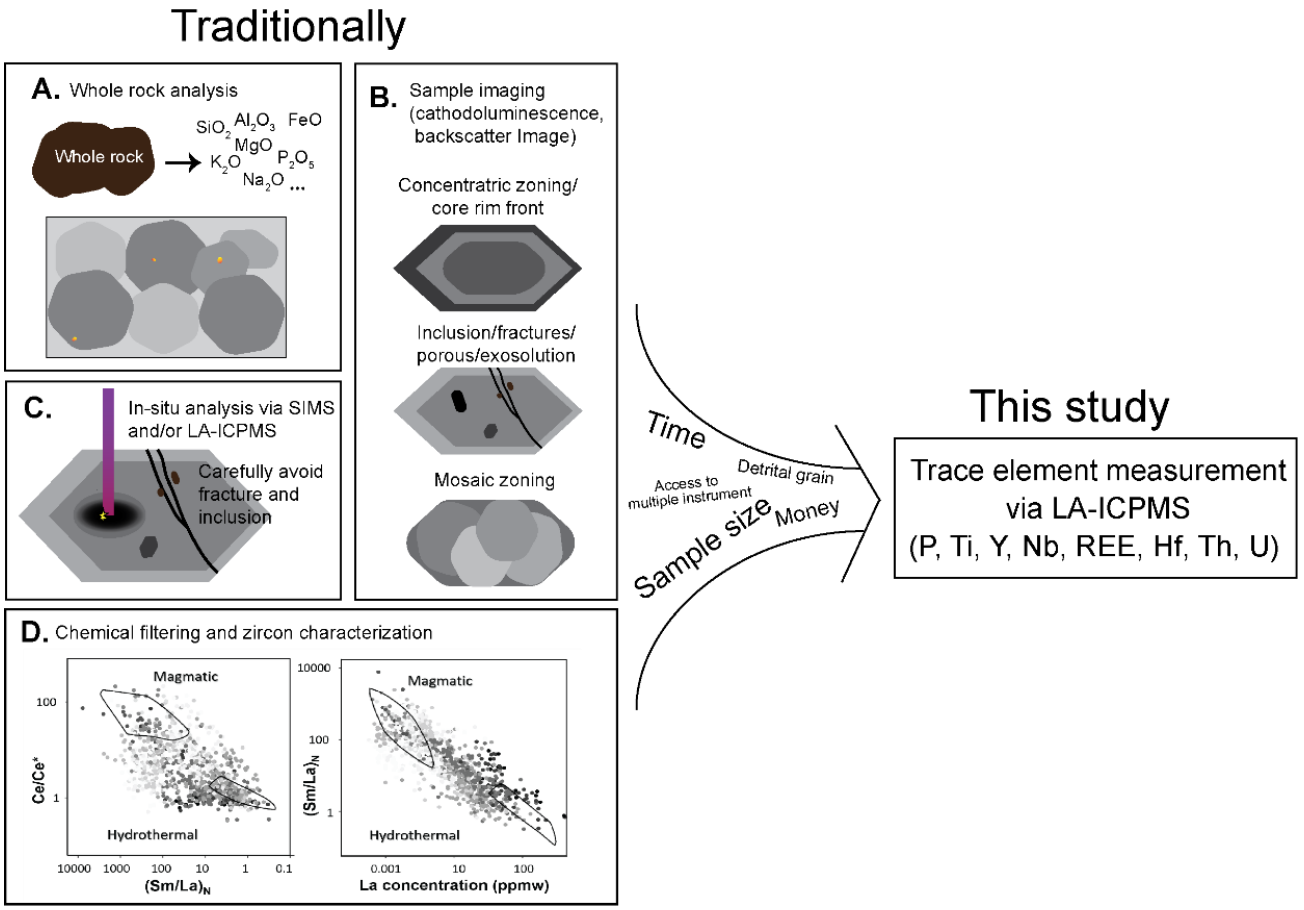


Figure 2.2. A comparison diagram between traditional approach and this study at screening mantle-equilibrated zircon.

2.2.Materials and Methods

In order to establish, validate, and demonstrate the filter scheme proposed here, this study compiled and prepared three separate zircon datasets, which correlate to sections 5.1-5.3 respectively.

2.2.1. Establishment of multidimensional filter scheme

The establishment of filter scheme has to rely on the selection and validation of individual trace element criterion at constraining zircon saturation conditions. Though each selected criterion is summarized from previously published review studies and/or justified with experiments and simulation, we also compiled a zircon dataset from 30 independent references with following trace element measurements: P, Ti, Y, Nb, REE, Hf, Th, U to confirm the selection statistically. Zircon entries without trace element concentrations listed, or with values below detection limits, are not used in this study. In the end, the dataset contains 2173 individual zircon analyses from a wide age distribution (i.e., Mesozoic to Archean zircon) and tectono-environments (e.g., continental/ island arc, post-collisional boundary, Ocean Island, etc.). Each zircon entry was labelled with type of their host rock (i.e., igneous, metamorphic, detrital, carbonatite, kimberlite, and mineral deposits) and zircon description from BSE and CL imaging (i.e., magmatic, magmatic with inclusion, magmatic with metamorphic rims, magmatic with hydrothermal rims, metamorphic, hydrothermal, metamict). The classification was left as blank if no host rock or contextual information was provided in the literature. Such detailed labelling facilitates the selection of individual trace

element criterion in section 2.3, and further helps to propose and validate the multidimensional filter scheme in section 2.4 and 2.5.

2.2.2. Validation of filter scheme with Jack Hills zircon

We validate the filter scheme with Jack Hills zircon. This zircon dataset has been characterized via BSE/CL imaging, oxygen isotopes, U-Pb ages, trace element measurements (except Nb) by Cavosie et al., 2006, and Ti contents of zircon are measured by a separate study Fu et al., 2008. After excluding unreliable zircon analyses that identified by Cavosie et al., 2006, this validation dataset contains 53 zircon in total, though some of the age and oxygen isotope measurements are missing. The same dataset has been used to select mantle-equilibrated zircon and derive oxygen fugacity of Hadean magma in Trail et al., 2011. Therefore, a direct comparison of our selected zircon and inferred melt redox state to previous study corroborate the performance of proposed methods.

2.2.3. Case study: Application of filter scheme to TTG zircon from Greenland

We further apply this filter scheme to an unpublished Tonalite-Trondhjemite-Granodiorite (TTG) zircon dataset (n=76) from three rock samples (524994, 524906, 524907) with trace element measurements only. This case study has a full set of trace element measurement required, and is meant to demonstrate the use of this filter scheme at facilitating zircon research. The trace element measurements, including U-Pb dating, were collected via LA-ICP-MS (see methods in Appendix A).

2.2.4. Estimates and uncertainty of Ti-in-zircon temperature and inferred magmatic fO2

The Ce/Ce*, Ti-in-zircon temperature, and melt redox state of all zircon entries in three studies were calculated using Eqn (2.1) and (2.2). The values of two unknown variables a_{SiO_2} and a_{TiO_2} are justified as following. In the literature, the value of a_{SiO_2} is commonly assumed to be unity as zircon most commonly crystallize from melt that is close to quartz saturation (Ferry and Watson, 2007). The value of a_{TiO_2} in silicic melts at appropriate magmatic temperature is typically between 0.6 and 1, and it can be estimated based on the crystallization of mineral phases at equilibrium with zircon, such as rutile (a_{TiO_2}=1), and ilmenite and/or titanite (a_{TiO_2}= 0.6) (Watson et al, 2006; Ferry and Watson, 2007). In applications to quartz-free crustal rocks and/or detrital zircon with limited or no geological context, the uncertainty in a_{SiO_2} and a_{TiO_2} will impose only limited effects on calculated temperatures. For example, assuming $a_{SiO_2} = 1$ and $a_{TiO_2} = 1$ yields an uncertainty up to 70 ℃ at 750 ℃ (Ferry and Watson, 2007). Therefore, for simplicity, this study calculates Ti-in-zircon temperature and derive inferred melt redox state assuming unity a_{SiO_2} and a_{TiO_2} respectively. However, when justification of a_{SiO_2} and a_{TiO_2} is available, accurate a_{SiO_2} and a_{TiO_2} values are highly recommended to reduce the uncertainty of inferred melt redox state. By comparing the melt redox state values determined from this study to previously published values (see 5.2), the analytical uncertainty of magmatic fO_2 calculation is estimated to be 1.3 log unit (sd).

2.3.Results

As mentioned in section 2.1.3, the variance of inferred melt redox state derived from Ce/Ce* in zircon can be attributed to three categories: (1) mineral inclusions; (2) source melt petrogenesis; and (3) saturation context of zircon. The trace element systematics of zircon associated with these processes have been explored independently in previous studies via simulations, experiments, and global datasets, which are summarized below as foundations for selecting filtering criteria.

2.3.1. Mineral inclusions in zircon

Mineral inclusions are common in igneous zircon. Some inclusions act as nucleation sites for crystallizing zircon crystals in melt (i.e., primary inclusion); others, however, are alteration or exsolution features reflecting post-magmatic processing (e.g., hydrothermal alteration, metamorphism, exsolution features) (i.e., secondary inclusion). The scale of such intragranular heterogeneity varies from visible micron-sized inclusions to nanoscale trace element enrichment zones (Corfu et al., 2003; Anderson et al., 2008; Hofmann et al., 2014; Bell et al., 2015). Therefore, sampling of inclusion phases and/or local enrichment zones cannot be completely avoided even with high-resolution petrographic imaging. Spatially resolved analytical techniques that characterize exceedingly small sample volumes (e.g., Secondary Ion Mass Spectrometry, SIMS) are particularly susceptible to disproportionate sampling of trace amounts of inclusions or nanoscale enrichments; the chemical signatures of localized features may be artificially amplified by individual spot analyses.

Based on this study, measurements of zircon plagued by primary and secondary inclusions can lead to up to 15 orders of magnitude variation in inferred melt redox states, as such phases may inflate estimated Ti-in-zircon temperatures (e.g., via excess Ti derived from titanite and magnetite) and/or attenuate the magnitude of observed Ce/Ce* anomalies (e.g., via overprinting REE signatures). Simple two-component mixing between a variety of phases commonly included in zircon (i.e., apatite, titanite, allanite, xenotime, and monazite) and the most primitive zircon grain from the case study reviewed in Section 2.5.2 (see below) demonstrates that as little as 0.0001 wt % of monazite and allanite inclusions will reduce Ce/Ce* by more than 80%. Interestingly, 0.5 wt % of inclusions enriched in light REE will overprint the Ce/Ce* signature completely (**Figure 2.3**). Therefore, due to the high sensitivity of Ce/Ce* oxybarometry to mineral inclusions and the potential risks of sampling them, filtering magmatic zircon for such based on diagnostic chemical trademarks is one of the most important steps for constraining variation in inferred melt redox states.

Historical criteria for filtering hydrothermal zircon include Ce/Ce*, La concentration, $(Sm/La)_N$, and light rare earth element index (LREE_I), calculated as the concentrations of Dy/Nd+Dy/Sm (Hoskins 2005; Bell et al., 2019). Based on petrographically predetermined magmatic and hydrothermal zircon, filtering schemes previously proposed in literature targeting the selection unaltered magmatic zircon are Ce/Ce* $\leq$ 100, La between 0.002 to 0.5 ppm, $(Sm/La)_N$ between 30 to 1000, and LREE_I > 60 . Because Ce/Ce* is one of the input parameters for calculating oxygen fugacity via Eqn (1), the range of Ce/Ce* is especially important for accurately determining melt redox state. The traditional upper boundary Ce/Ce* $\leq$ 100 is adopted

as a filter for magmatic zircon because Ce/Ce* >100 is approximately equivalent to an unrealistic magmatic environment as oxidized as modern atmosphere (ΔFMQ > +10; McCammon, 2005). The lower boundary of Ce/Ce*, however, cannot be directly inferred because it varies with the amount of REE in the melt, as well as the mixing ratio of inclusions in the sample. Therefore, instead of setting a lower boundary of Ce/Ce*, the other three parameters, La concentration, $(Sm/La)_N$, and LREE_I, are used to change the fraction of inclusions within zircon.

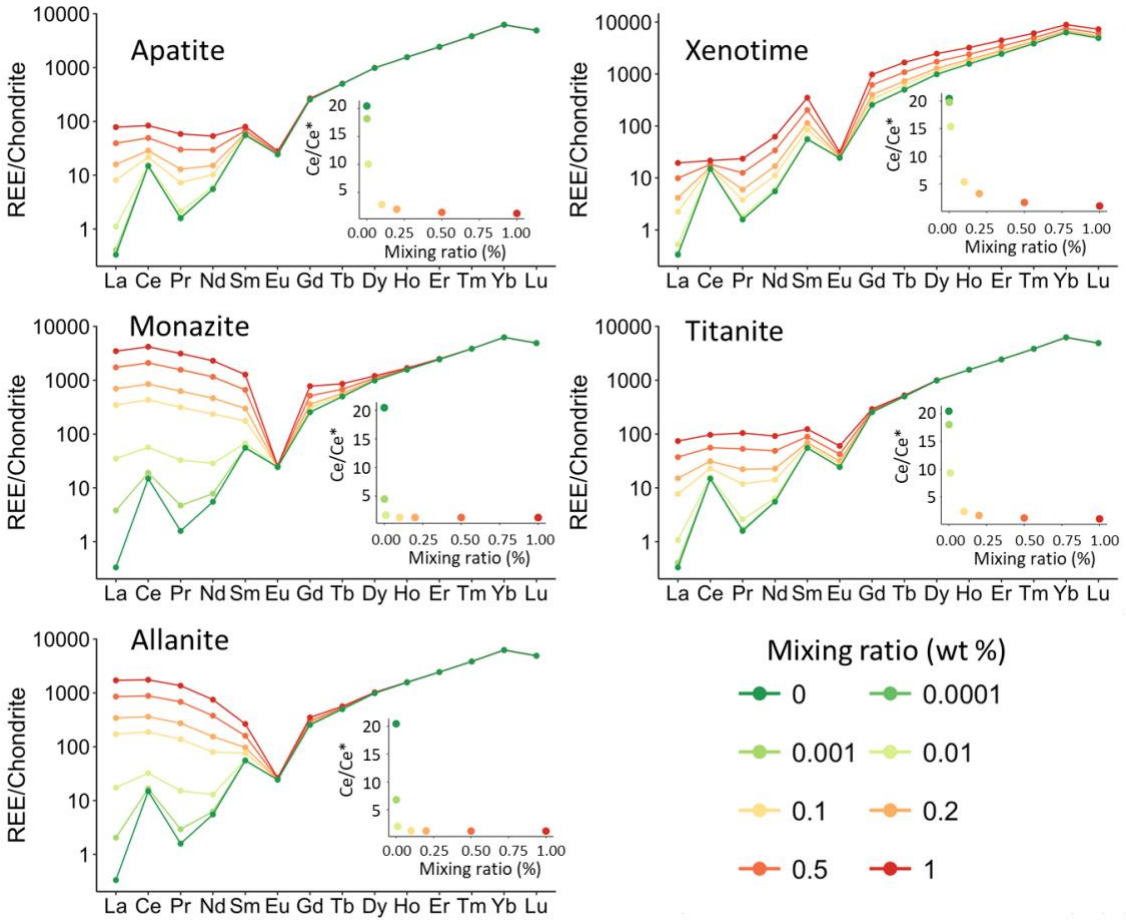


Figure 2.3. Mixing models between zircon and potential inclusions across a range of mixing proportions demonstrates the sensitivity of Ce/Ce* to chemical overprinting. The REE concentration of unadulterated zircon is modeled after sample 524994 Zircon 76, the most primitive zircon grain from the case study reviewed in Section 5.2. The average REE concentration of apatite is from Zhong et al. (2018a), and the compositions of the other phases (xenotime, monazite, titanite, allanite) are from Bea et al. (1996). The insets describe the relationship between Ce/Ce* and the mixing ratio of inclusions.

Figure 2.4 shows the sensitivity of La concentration, $(Sm/La)_N$, and LREE_I for filtering mineral inclusions of different mixing ratio to zircon. La < 0.1 ppm is the most sensitive parameter that is able to filter out zircon with mineral inclusions as low as 0.0001 wt%, but such sensitivity is reduced by up to three orders of magnitude if La < 0.5 ppm is used instead. Similarly, $(Sm/La)_N > 100$ is able to discern 0.0001 wt% monazite and allanite, but relaxing the requirement to $(Sm/La)_N > 30$ masks up to 0.2 wt% of xenotime, which can attenuate Ce/Ce* by 80 %. Although LREE_I is the least sensitive parameter among these three, it is important to use an additional LREE-independent criterion to monitor inclusions because the LREE measurements in zircon are analytically challenging due to their low count rates. Other LREE-independent criteria recommended to detect inclusions include using elevated Ca and Fe content in zircon for apatite and magnetite inclusion respectively (Bell et al., 2019). However, these proxies are less sensitive than La content and $(Sm/La)_N$ (e.g., La content < 0.1 ppm and $(Sm/La)_N > 100$ can select zircon with Fe content < 150 ppm based on dataset in Bell et al., 2019), thus they are not favored in these studies. Further, because trace element concentrations of each mineral inclusion vary over two orders of magnitudes in natural samples, multidimensional criteria can increase the filtering power and reduce the probability of detecting inclusions. Therefore, the boundaries of current filtering scheme are updated to identify samples with the highest probabilities of reflecting unaltered mantle-equilibrated magmatic zircon: La concentration between 0.1 and 0.002 ppm, $(Sm/La)_N$ between 100 and 1000, and LREE_I > 60. In addition, Ce/Ce* $\leq$ 100 is adopted to distinguish magmatic zircon.

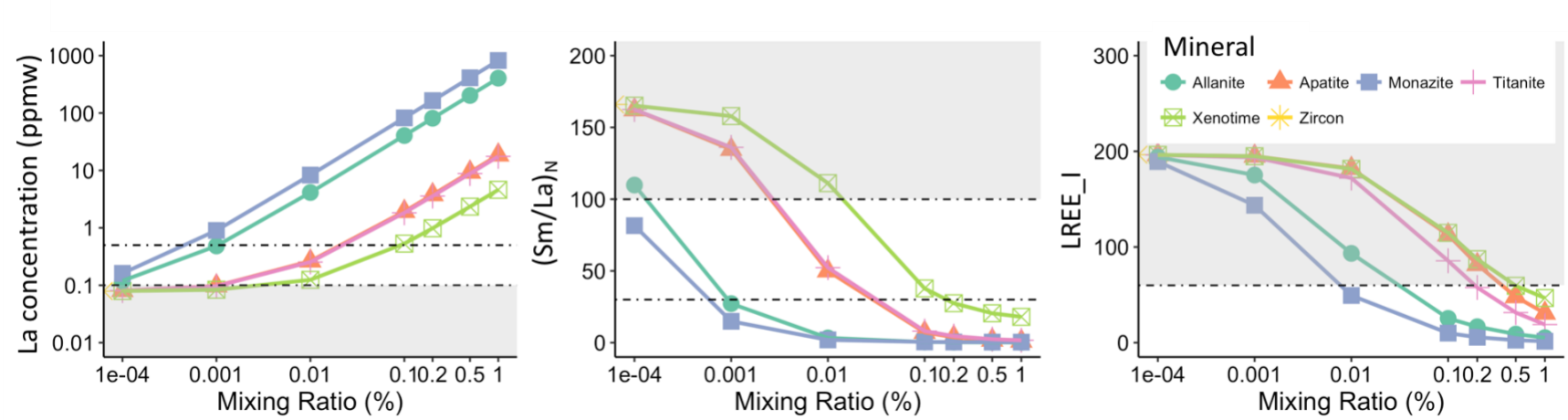


Figure 2.4. A sensitivity comparison between all three inclusion indicators: La concentration; $(Sm/La)_N$; and, LREE_I as a function of mixing ratio (%). The shaded areas represent regions that will be selected (or accepted) by the filter scheme.

2.3.2. Source melt petrogenesis

Zircon starts to crystallize when the Zr availability in the system (e.g., source melt) is greater than or equal to the zircon saturation content. Zircon saturation content, or the minimum Zr concentration required to precipitate zircon in the melt, tends to decrease with an increasing degrees of melt polymerization and/or decreasing magma temperature (Boehnke et al., 2013; Hanchar and Watson, 2003). For example, the zircon saturation content in basaltic magma is about 2000 to 30000 ppmw Zr (Shao et al., 2019) but reduces to ~100 ppmw in silicic magma (Watson, 1979). The availability of Zr in melt, however, is constrained and remains relatively constant; for example, the average Zr content in global mid ocean ridge basalt (e.g., ~100 ppmw; e.g., Arevalo and McDonough, 2010) is similar to the composition of the bulk continental crust (~130 ppmw Zr; e.g., Rudnick, 1995). Despite similar Zr availability, the zircon saturation contents of these two systems are dramatically different because melt polymerization serves as a dominant control, and prevents the crystallization of zircon in mantle-derived magma unless the source melt interacts with metasomatic fluids or assimilates felsic materials (e.g., marine sediments, slab-derived melt, or crusts). In fact, the source melts of most zircon in ultramafic and/or mafic host rocks in compiled dataset are often interpreted as melt products of metasomatized mantle-wedge, or contaminated mantle-derived melt that mixed with slab-derived fluids or crustal materials during ascent (e.g., Janoušek et al 2019; Robinson et al., 2015; Tichomirowa et al., 2012; Sun et al., 2018; Vilalva et al., 2019, etc.). Because fluids and surface materials have undergone various low temperature processes (chemical and mechanical

weather, erosion, etc.) and commonly possess enriched incompatible trace element abundances, assimilation of these materials can easily overprint the inherited fO_2 signatures from the original mantle source.

Historical studies of source melt petrogenesis focus on whole rock geochemistry and isotopic measurements of zircon crystals (e.g., Robinson et al., 2015; Liang et al, 2018; Janoušek et al., 2019). Specifically, whole rock geochemistry provides insights about the provenance and composition of the source melt, and oxygen and Hf isotopes of zircon track the interaction to metasomatic fluids and the contribution of felsic materials. In this study, instead of relying on whole rock or detailed isotopic analyses, previously developed trace element proxies are used to infer the formation and composition of source melt.

2.3.2.1. Source melts from igneous sources: P $\leq$ 750 ppm and (REE+Y)/P molar ratio > 1

Zircon is commonly found in granitic systems, but felsic melt can derive from both igneous sources (I-type granite, alumina-saturated metaluminous melt) and sedimentary sources (S-type granite, alumina-oversaturated peraluminous melt). Generation of melt from a sedimentary protolith implies little to nothing about mantle redox conditions, so zircon from S-type granitoids provide less insight into the deep Earth compared to mantle-equilibrated zircon.

Zircon P concentration provides a gauge to identify the type of granitic protolith. The average P concentration of global MORB is ~600 ppmw (Arevalo Jr. and McDonough, 2010). Due to its incompatibility to most major minerals, the P concentration of a melt

may increase as the magma differentiates until apatite precipitates. Apatite ($Ca_5(PO_4)_3(OH, Cl, F)$) is a common accessory phase that crystallizes prior to zircon saturation in metaluminous melt (Lee et al., 2014), but dissolves in peraluminous melt as magmatic differentiation continues (Mysen et al., 1997). Despite the insignificant difference of bulk rock P_2O_5 concentration between S-type and I-type granites, the saturation and dissolution of apatite control the availability of P in residual melts.

In general, S-type zircon inherit a higher P concentration, whereas only 10 % of I-type zircon have P concentration > 750 ppm (Burnham and Berry, 2017). Moreover, coupled substitution between P^{5+} and $(REE,Y)^{3+}$ for Zr^{4+} becomes dominant in melt with a higher P concentration. S-type zircon are observed to have a strong linear correlation between the molar ratios for REE+Y and P, whereas I-type zircon are observed to have (REE +Y)/P molar ratio > 1. The lack of linear correlation in I-type zircon implies preference of additional REE incorporation mechanisms in P-deficient metaluminous melt. In a case study by Burnham and Berry (2017), the authors suggest that the combined criteria $P \leq 750$ ppm and (REE+Y)/P molar ratio > 1 enables the isolation 96% of I-type zircon based on 144 zircon analyses from characterized S-type and I-type granites.

Similar source melt information has been inferred form Eu/Eu* (Wang et al., 2012, see section 2.4.3.4) and Al content in detrital zircon (Trail et al., 2017). Although they are not specifically addressed in these sections, they could be considered as alternative criteria for differentiating source of source melt.

2.3.2.2. Limited interaction with fluids and surface materials: U/Nb ≤ 20 and/or Th/Nb ≤ 10

Partial melting of basaltic systems, especially at arc settings, often involves fluid components and/or more surface materials (e.g., marine sediments and continental crust). The degree of such interaction can be inferred from U/Nb, Th/Nb ratios in zircon. Fluid immobile Nb fractionates from fluid mobile Th and U during low temperature surface weathering, erosion, and/or fluid interaction. Thus, normalization of Th and U to Nb can be used as a monitor for mixing with such exogenous materials.

Based on over 5300 zircon analyses compiled by Grimes et al. (2015), the zircon formed in arc magmas may be distinguished from those derived from mid-ocean ridges and intraplate ocean island by notably higher U/Nb >20 and Th/Nb >10. These chemical signatures can be explained by the Th and U enrichment in relative to Nb due to assimilation of Th and U enriched sediments, or due to fractionation of fluid mobile Th and U into resultant magma during hydrous melting. Therefore, zircon with U/Nb ≤ 20 and/or Th/Nb ≤ 10 best represent crystallization from mantle-equilibrated melt without sufficient interaction to surface materials and fluids.

Similar trace element systematics for assessing zircon with limited fluid interaction and surface material assimilation include U/Yb and Th/Yb (Grimes et al., 2007, Grimes et al., 2015). However, the specific values of these ratios vary if source melt originates from variably enriched mantle sources. For example, while low U/Yb (< ~ 0.1) appears to be diagnostic of zircon from differentiated melt originating from a depleted mantle source, zircon from relatively enriched geochemical reservoirs typically carry higher

U/Yb ratios (Grimes et al., 2007, Grimes et al., 2015, Carley et al. 2014). Normalization of U/Yb and Th/Yb to Nb/Yb helps to reduce such effects, making U/Nb and Th/Nb preferred proxies in this study. However, when Nb measurement is not available (such as in case study 5.2), U/Yb and Th/Yb can be used interchangeably. Because continental arc zircon are offset to Mid Ocean Ridge (MOR) and Ocean Island (OI) zircon by higher U/Yb and Th/Yb ratios, this study recommends U/Yb $\leq$ 0.5 and Th/Yb $\leq$ 0.2 to exclude arc zircon, where the cutoff values are derived from the first quantile of continental arc zircon and third quantile of MOR and OI zircon summarized in Grimes et al., 2015.

2.3.3. Zircon saturation context

As a magma progressively differentiates prior to zircon saturation, the precipitation of major mineral phases will continuously change the availability of major elements in the melt, including redox-sensitive ratios such as Fe^{2+}/Fe^{3+}, thereby manipulating the melt redox state. Saturation and/or dissolution of accessory phases, such as monazite or xenotime, can control trace element budgets in residual melts, thus changing the REE signatures inherited by zircon. In addition, the speed of magma differentiation and oversaturation of trace elements can lead to disequilibrium crystallization of zircon, resulting in observed partitioning behavior that deviate from calibrated relationships (e.g., Eqn (2.1)). Such magmatic processes play significant roles in changing melt redox states and influencing trace element concentrations in zircon. Beyond saturation from magmatic melt, zircon can also experience re-equilibration, dissolution, and recrystallization in Zr-rich melts sources from the decomposition of Zr-rich mineral phases, or Zr-rich metasomatic fluids. In such cases, saturation of non-magmatic zircon will not appropriately record the melt redox state via Eqn (2.1). We employ five chemical filters to monitor for: (1) fractionation of redox-sensitive mineral phases and REE-enriched accessory phases; (2) degree of magma differentiation; (3) equilibrium crystallization of zircon; and, (4) non-magmatic zircon signatures.

2.3.3.1.Absence of flat HREE pattern: (Lu/Gd)N between 10 and 74

The partitioning of heavy REE (HREE) zircon is suppressed due to fractionation of garnet prior to or contemporaneous with zircon crystallization, leading to a characteristic flat HREE pattern (or a “plateau”) in the zircon. More problematic, the

fractionation of garnet can alter the melt redox state by up to 6 orders of magnitude in arc settings (Tang et al., 2018), as Fe^{2+} is more compatible in garnet, leaving behind residual melt with a disproportionate concentration of Fe^{3+}. Because Fe serves as a major control on local fO_2, an increase in $Fe^{3+}/\Sigma Fe$ in the residual melt (as promoted by fractionation of garnet) could effectively oxidize the system, consequently changing the Ce^{4+}/Ce^{3+} ratio in the melt relative to the original mantle setting. In addition, flat HREE patterns may also be imparted by metamorphic processes. Under high-temperature and high-pressure environments where garnet is stable, partial melting or decomposition of Zr-rich mineral phases (e.g., titanite) can contribute Zr, leading to subsolidus crystallization of zircon. Similarly, zircon in mantle-derived host rocks devoid of garnet (i.e., kimberlite, carbonatite) can also have flat HREE pattern due to incomplete melting of garnet in mantle source region (Hoskin and Schaltegger, 2003; Rubatto, 2017). Saturation of zircon in mantle-derived melt is thermodynamically hindered unless it is aided by a large amount of metasomatic fluids or slab-derived felsic melt. Therefore, zircon with flat HREE pattern often associate fractionation of garnet in melt, metamorphic origin of zircon, or disequilibrium crystallization in mantle-derived host rock, so they imply little about the local magmatic environment.

Experimental results show that $D_{Lu}^{Zircon/Melt}/D_{Gd}^{Zircon/Melt}$ in hydrous granitic melt is approximately 9 at 850℃, the upper zircon saturation temperature identified in this study (see 2.3.3.5), and increases to a higher value as magma cools (Rubatto and Hermann, 2007). The presence of garnet also suppresses the partition of Lu and Gd into zircon; for example, $D_{Lu}^{Zircon/Garnet}/D_{Gd}^{Zircon/Garnet}$ is 6 at 850℃, and further reduces

to 1 as temperature increases to above 900℃ (Rubatto and Hermann, 2007; Taylor et al., 2015). This agrees with the empirical observations that mantle-affinity zircon (co-presence of garnet) have $(Lu/Gd)_N$ ranging from 1~10, metamorphic zircon with $(Lu/Gd)_N$ approaching 1 (Kelly and Harley, 2005; Whitehouse and Platt, 2003), whereas typical granitic zircon have $(Lu/Gd)_N$ ranging from 16-74 (Hoskin and Schaltegger, 2003). Therefore, magmatic zircon without coexistence of garnet tends to have $(Lu/Gd)_N$ between 10 and 74.

2.3.3.2.Avoiding fractionation of accessory phases: Th/U ratio between 0.2-4

Prior to or during zircon saturation, crystallization of Th-enriched phases (e.g., monazite and thorite) and/or U-enriched phases (e.g., coffinite, uraninite, or monazite) induces local heterogeneity of Th and U in the melt, enhancing observed Th/U variances in cogenetic magmatic zircon. Such heterogeneity can be further augmented via post-magmatic processes, such as hydrothermal alteration, metamorphism, and/or low temperature weathering. Hydrothermal fluids may mobilize, redistribute, or precipitate Th-rich and U-rich inclusions along cracks, or promote mass loss of U by oxidizing U^{4+} to form fluid mobile UO_4^{2-}. Metamorphism, decomposition, and crystallization of Th- and U-rich accessary phases can also incur spatial variations in Th and U concentration in melts, facilitating the dissolution and disequilibrium growth of zircon (Xiang et al., 2011). As a result, zircon with extreme values of Th/U are often associated with significant inclusion/fractionation of accessory phases, disequilibrium crystallization, and non-magmatic origins.

Based on 10,693 magmatic zircon analyses compiled in Kirkland et al. (2015), zircon Th/U ratios range from near 0 up to 19.8, with a median of 0.65 and a positive skewness of 6.7. Although metamorphic zircon exhibit a similar spread in Th/U, a greater percentage of metamorphic zircon have Th/U < 0.1 (Yakymchuk et al., 2018). Historically, Th/U < 0.1 was used as filtering tool for metamorphic zircon, but even zircon in ultrahigh temperature metamorphic assemblage (> 900 ℃) can exhibit Th/U > 0.1, reinforcing the complexity of Th/U systematics (Kelly and Harley, 2005).

The large variance in Th/U recorded by natural zircon may be attributed to reasons mentioned above. However, the effects of each of these controls on Th/U ratios have not been fully characterized. Based on variation of zircon Th/U observed in compiled datasets from this study and from Kirkland et al. (2015), the Th/U of magmatic zircon is set to be within 0.2 and 4 (**Figure 2.5**).

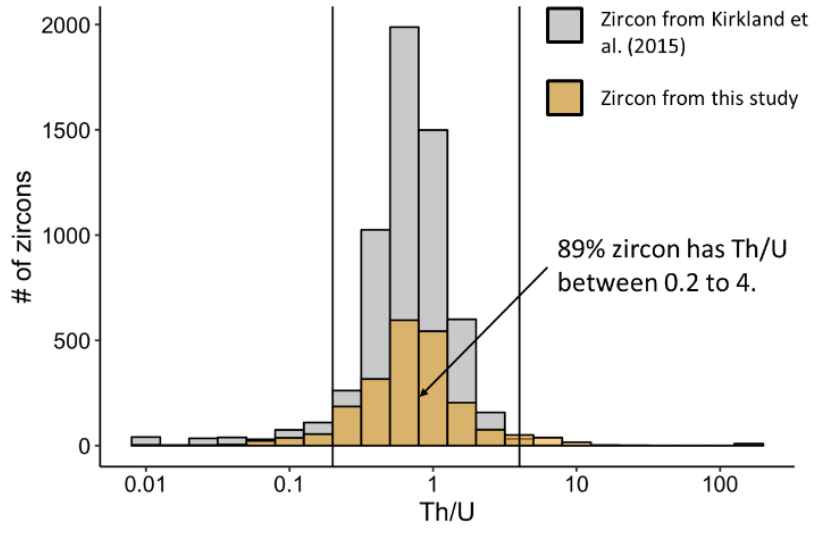


Figure 2.5. The distribution of Th/U ratio in the compiled dataset (n=2173) from this study and zircon from Kirkland et al. (2015).

2.3.3.3.Constraining differentiation and metasomatism: Hf concentration: 5,000-14,000 ppmw

As HFSE, Hf and Zr are both highly incompatible in most mineral phases, so the Hf and Zr concentration in melt increases significantly as magma differentiates. Zircon, as one of the primary carriers of Hf in magmatic systems, readily incorporates Hf into its crystal lattice due to the same valence states and similar ionic radii to Zr^{4+} (Wang et al., 2010). Therefore, the Zr/Hf ratio in zircon remains relatively constant in granitic magma and extreme Zr/Hf ratio can be used as markers for fractionation of Hf-enriched accessary phases (e.g., hafnon) and significant magma cooling.

According to 2201 zircon electron microprobe analyses complied in Wang et al. (2010), the Zr/Hf signatures recorded by granitic zircon range from 5 to 100, showing a nearly normal distribution with a mean, median, and mode of 39. This mean Zr/Hf ratio is reproducibly observed among different studies of granitic zircon from 35 to 38, and is notably higher than zircon from pegmatite or highly evolved rocks (i.e., Zr/Hf ratio ~ 20) (Linnen and Keppler, 2002; Pupin, 2000; Claiborne et al., 2006). Interactions with a fluid phase will reduce Zr/Hf ratio in zircon, resulting in enrichment and positive correlation between Hf and Rb (Linnen and Keppler, 2002).

By assuming an average Zr concentration in zircon based on stoichiometry, a Zr/Hf ratio between 35 (the lower bound of granitic zircon; Claiborne et al., 2006) and 100 (the max value observed in the same data set; Wang et al., 2010) equates to 5,000 to 14,000 ppmw Hf in magmatic zircon. This compositional range overlaps with the Hf

content observed in natural zircon (Claiborne et al., 2010) and includes approximately 88% of the data compiled in this study. Notably, zircon with Hf > 14,000 ppmw commonly exhibit higher La contents, corroborating a potential metasomatic influence and/or precipitation of LREE-enriched mineral phases (Claiborne et al., 2006).

2.3.3.4.Equilibrated crystallization from silicic melt: Eu/Eu*: 0.1 - 0.6

Besides Ce, Eu is another REE that has two valence states, Eu^{2+} and Eu^{3+}. Eu^{2+} is strongly incompatible in the zircon structure due to the large ionic size and required coupled substitution in zircon structure, but Eu^{2+} is strongly compatible to the Ca^{2+} site in plagioclase. Therefore, fractionation of plagioclase depletes Eu content in the melt. As a result, the Eu/Eu* ratios in zircon rarely exceed one, unless zircon saturates in disequilibrium conditions (Hoskin and Schaltegger, 2003; Trail et al., 2012).

The relationship between Eu/Eu* and Hf depicts a continuum of magma differentiation that connects zircon from mantle-derived melts (i.e., carbonatite and kimberlite) to silicic melt (i.e., igneous rocks; **Figure 2.6**). Majority of zircon in carbonatite and kimberlite are observed to have low Hf concentration and Eu/Eu* > 1, supporting hypothesis that saturation of zircon in mantle-derived melt is thermodynamically unfavorable. However, as magma sufficiently evolves, zircon inherits a higher concentration of Hf and lower Eu/Eu*; for example, the majority of igneous zircon from this study have Eu/Eu* $<$ 0.6 (Figure 6). The distribution of Eu/Eu* broadly correlates to types of parental magma. Zircon from I type granitoid has a distinctive lower En/Eu* $\leq$ 1 than zircon from S- and A-type granitoid (Wang et al., 2012; Sawaki

et al., 2017). Therefore, mantle-equilibrated zircon is expected to have Eu/Eu* between 0.1 to 0.6.

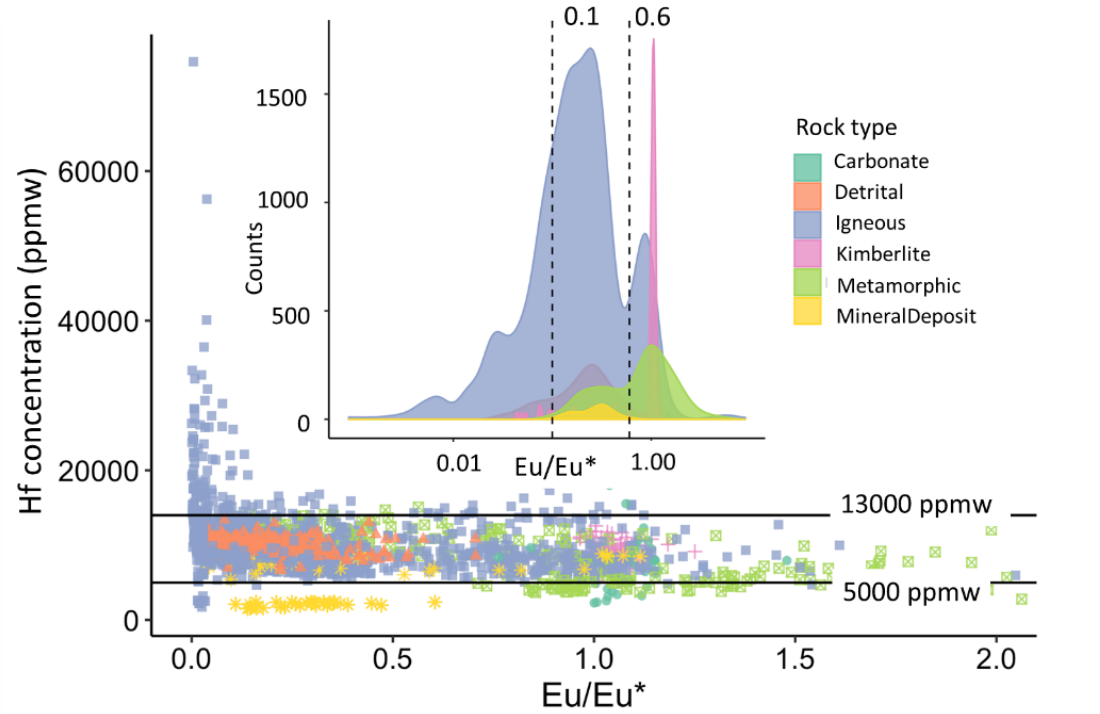


Figure 2.6. Hf concentration as a function of Eu/Eu* of zircon compiled in this study, categorized by host rock type. The inset shows the count distribution of Eu/Eu* of zircon by rock type.

2.3.3.5. Crystallization temperature for magmatic zircon: Ti concentration: 2 - 24 ppm

Titanium content in zircon is used to infer the zircon saturation temperature in magmatic system (see section 2.2.1). In a closed magmatic system, the onset of zircon saturation is controlled by the Zr availability versus the zircon saturation content (i.e., the minimum Zr content required to precipitate zircon from melt) in the melt as magma cools. By assuming a 100 ppmw initial Zr concentration in the system and a highly

incompatible behavior of Zr in the phases that first crystallize from a basaltic magma ($D_{Zr} \leq 0.01$), zircon is expected to saturate starting at 850 °C based on simulation of magma differentiation (Lee et al., 2014; Nandedkar et al., 2014; Borisov, 2019). This 850 °C represents an extreme upper bound, because the actual zircon saturation temperature is estimated to be ~800 to 600 °C, based on their Ti concentration and melt composition parameter M (i.e., the cation ratio of (Na + K + 2Ca) / (Al·Si) of the melt/whole rock) of 1,705 igneous zircon (Samperton et al., 2017; Siégel et al., 2018). Therefore, zircon saturating at temperature > 850 °C reflects an open magmatic system which inherits a disturbed trace element chemistry from mantle condition.

The majority of zircon data compiled in this study has Ti-in-zircon temperatures between 710 and 810 °C, which converges to the crystallization temperature range suggested by the simulations described above, 600 to 850 °C as well as in literatures (**Figure 2.7**; Harrison et al., 2007; Fu et al., 2008). By assuming unity a_{SiO_2} and a_{TiO_2} in Ti-in-zircon thermometry and considering the errors associated with the equation, we constrained the Ti concentration in magmatic zircon between 2 and 24 ppm, corresponding to zircon saturation temperature between 600 °C to 850 °C respectively. Zircon with Ti > 24 ppm is indicative of zircon saturation from an open magmatic system at > 850 °C, or an artificial enrichment of Ti contents due to sampling of contamination phases, such as magnetite. The later scenario can be effectively identified if zircon fails mineral inclusions proxies La content, $(Sm/La)_N$, and LREE_I mentioned in section 2.4.1, and can be further elaborated to magnetite inclusion if zircon shows elevated Fe content.

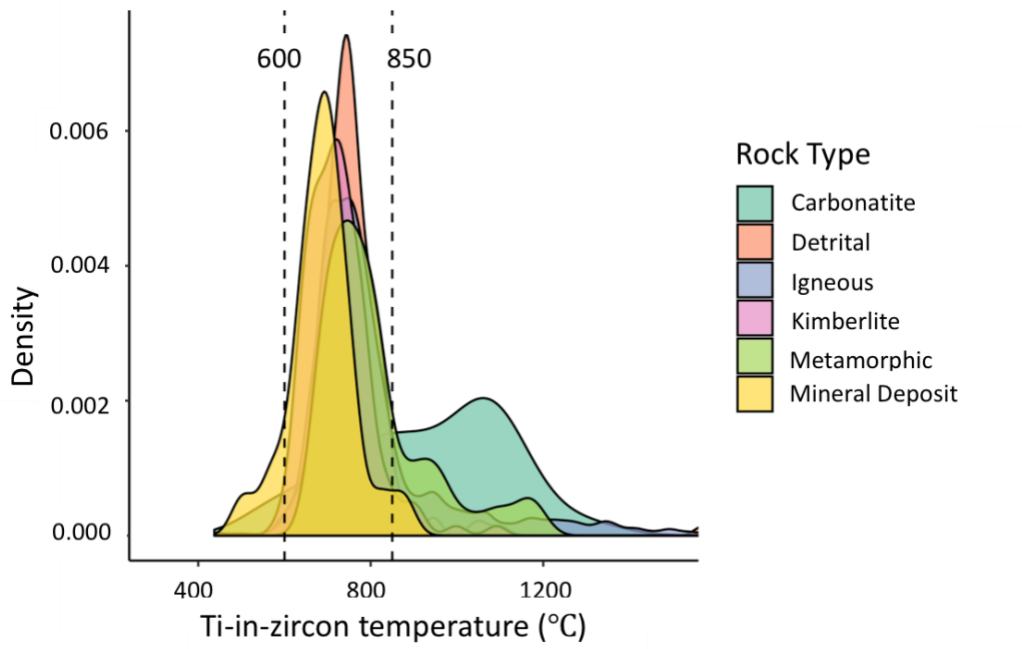


Figure 2.7. Density plot of Ti-in-zircon temperatures determined from the zircon data set compiled here, categorized by host rock type. The area under each curve has a cumulative probability of 1.

2.4.Discussion

In this section, a multidimensional filter scheme is proposed to select unaltered mantle-equilibrated zircon. Scoring system is established to guide the confidence of selection. Based on the discussion above, a total of 13 filtering criteria are summarized, all of which help to limit external interferences and artificial variances in inferred redox states from three perspectives:

(1) Mineral inclusions in zircon:

La concentration: 0.002 - 0.1 ppm

$(Sm/La)_N$ between 100 and 1000

LREE_I > 60

(2) Source melt petrogenesis:

U/Nb $\leq$ 20

Th/Nb $\leq$ 10

P $\leq$ 750 ppmw and (REE+Y)/P molar ratio > 1

(3) Zircon saturation context:

Ce/Ce* $\leq$ 100

$(Lu/Gd)_N$ between 10 and 76

Hf concentration between 5,000 and 14,000 ppm

Th/U ratio between 0.2 and 4

Eu/Eu* between 0.1 to 0.6

Ti concentration between 2 to 24 ppm

These criteria may be translated into a numeric scheme by employing a binary scoring system. A zircon will score 1 for each criterion it satisfies in Categories (2) and (3); if the zircon falls outside the predefined range, it receives a score of 0 for that proxy. However, because inclusion phases have extraordinary control on both Ce/Ce* and Ti-in-zircon temperatures, the criteria outlined in Category (1) are given a 2x weight factor; that is to say, a zircon will receive a score of 2 for each criterion it satisfies in Category (1). If the zircon falls outside the predefined range, it receives a score of 0 for that Category (1) proxy. For example, a zircon entry that satisfies all criteria except that its U/Nb $>$ 20, La concentration $>$ 0.1 ppmw, and $(Lu/Gd)_N$ $<$ 10 ppmw will have a total score of 11 (**Figure 2.8**).

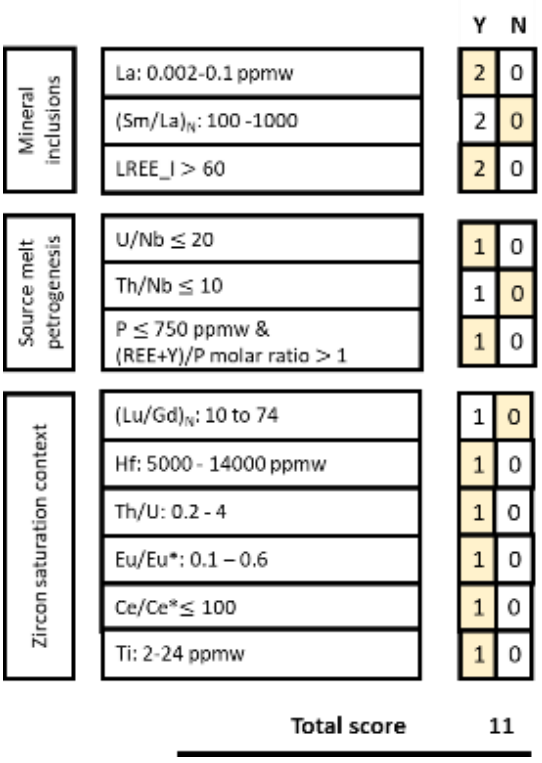


Figure 2.8. An example of calculating the score of zircon that satisfy all criteria except that its U/Nb $>$ 20, La concentration $>$ 0.1 ppmw, and $(Lu/Gd)_N < 10$.

Noted that because filter scheme relies on multiple categorical questions on trace element measurements, large analytical uncertainty at the boundary condition of each criterion could lead to a false negative selection. For example, Hf concentration of 14,100 ppm might be rejected by the criterion (Hf < 14,000 ppm) but it is in fact satisfactory considering the 5% (sd) analytical uncertainty of Hf measurement (i.e., 14,000 $\pm$ 700 ppm). Therefore, we strongly recommend to add analytical uncertainty at the boundary of each criterion depending on the quality of trace element measurements. In this study, a 5% analytical uncertainty is applied for each criterion at

their boundary conditions; for example, Hf concentration is between 5000×0.95 to 14000×1.05.

The total score estimates the similarity of a given zircon entry to a mantle-equilibrated zircon. Zircon with the high scores are most likely to represent mantle-equilibrated zircon, whereas zircon with low scores have a reduced probability of preserving the identities of their respective mantle sources. Samples with the highest confidence of reflecting mantle-equilibrated zircon will have a maximum score of 15 (i.e., satisfying all criteria).

2.4.1. Establishment, validation, and application of filter scheme to zircon samples

The scores for all 2173 zircon entries compiled here cover the entire scoring range from 1 to 15 (**Figure 2.9**). Application of Hoskin's differentiation diagrams suggest that most low score zircon are derived from hydrothermal sources, whereas high score zircon cluster within the magmatic domain (**Figure 2.10**). In general, the zircon population with the highest scores converge at melt redox states near the FMQ region. Based on the distributions of scores, the zircon may be divided into three groups, as described further below (**Figure 2.11a**).

Group III zircon have scores ≤ 7 with a median oxygen fugacity ΔFMQ -5.7 $\pm$ 5.0 ($\pm$ median absolute deviation; **Figure 2.11a**). Consequently, these samples are interpreted to represent non-magmatic zircon. The median absolute deviation is calculated as the

median of the absolute distances of each data value to the median of whole sample set; compared to standard deviations, this value is less affected by outliers and better represents statistical dispersion within dataset. Group III zircon generally yield Ce/Ce* approaching unity but surprisingly high Ti-in-zircon temperatures (>1000 ℃), highlighting unusual trace element characteristics (**Figure 2.9**). Group III zircon represent > 40% of zircon from metamorphic rocks and nearly all zircon from carbonatites (**Figure 2.11b**). According to the textures and morphologies of these samples provided in their source references, zircon with unity Ce/Ce* and high temperature often associates with metamict, hydrothermal, and metamorphic zircon, which validate that Group III zircon represent non-magmatic origin (**Figure 2.12**).

Group II zircon have scores from 8 to 13 with a median oxygen fugacity ΔFMQ -0.5 ± 4.9. They mostly represent zircon with mineral inclusions and zircon that crystallized from parental magma that interacted with hydrous fluids and/or crustal materials. Group II zircon have Ti-in-zircon temperature 620 to 850 ℃, in agreement with experiments and simulations of magmatic zircon crystallization temperature. However, their Ce/Ce* ranges over three orders of magnitudes (Figure 9). Hoskin discriminant diagrams suggest the majority of Group II zircon extend from the magmatic zone towards hydrothermal signatures (Figure 10). Such a trend implies that Group II zircon may have had magmatic origins but fluid alteration and/or metamorphism conditions induced post-magmatic trace element diffusion, mineral exsolution, or inclusions that attenuated Ce/Ce* and diminish the original magmatic fingerprint.

Group II includes nearly all zircon derived from kimberlites, detrital zircon in sedimentary rocks (i.e., siltstone), mineral deposits, and ~40% zircon from igneous rocks (**Figure 2.11b**). According to the specific examples in this study, zircon in kimberlite is identified as xenocryst with a source melt of mafic composition (Kostrovitsky et al., 2016), and detrital zircon in siltstone is crystalized from a melt of A-type granitoid or a mixture of felsic and mafic melt (Paulsen et al., 2017). Therefore, despite the range of host rock compositions, most of Group II zircon have a magmatic origin. Petrographic examination via BSE and CL images suggest most Group II samples have magmatic textures but with inclusions or secondary recrystallization features, likely resulting from hydrothermal fluids or metamorphism (**Figure 2.12**).

Group I zircon have scores from 14 to 15, meaning they have the highest confidence level of representing mantle-equilibrated zircon. The most noticeable differences between Group I and II zircon are the reduced variations in temperature and Ce/Ce*, and by extension inferred melt redox state (**Figure 2.11a**).

The majority of Group I zircon are from the rhyolitic Mesa Falls Tuff (MFT) at Yellowstone volcanic field (Rivera et al., 2016). The physical conditions and mechanisms of generating silicic magma at Yellowstone are hotly debated topics because cyclic caldera collapse, magma mixing, and crustal assimilation often overprint and complicate the signatures of minerals and magma (Bindeman and Valley, 2001; Wotzlaw et al., 2015). As a result, Yellowstone zircon often inherit heterogeneous oxygen isotopic compositions and trace element concentrations across zones that reflect variations in melt conditions as a function of time. Based on detailed

petrographic observation, trace element composition, and U-Pb dating, Rivera et al. (2016) identified 4 Chemical Domains in MFT zircon that represent distinctive magma conditions; in general, Domains 1 and 2 represent magma mixing from hydrothermally altered crust (characterized by statistically distinct $\delta^{18}O$ below mantle field); Domain 3 represents the oldest pristine magma in the zircon core; and, Domain 4 represents magma with significant fractionation. Out of >400 zircon analyses, this filter scheme has successfully avoided all zircon from Domains 1, 2, and 4, and only picked up zircon from Domain 3, sourced from the most primitive melt. Similarly, other Group I zircon include magmatic cores from Cu-Zn deposits (Zhu et al, 2017), detrital zircon in siltstone (Paulsen, 2017), and magmatic zircon from Yellowstone rhyolite (Colon et al., 2015). Although there are no oxygen isotopes for selected zircon grains as corroborative validation for their mantle-equilibrated source melt, all of these references have: (1) observed multistage growth of zircon and preservation of magmatic texture for some zircon; and, (2) highlighted the likelihood of maintaining mantle-equilibrated source melt based on Hf and oxygen isotopes on limited zircon grains and whole rock.

Accordingly, this filter scheme has demonstrated the capacity to isolate non-magmatic zircon (Group III zircon), magmatic zircon with inclusions or enriched source melt sources (Group II zircon), and mantle-equilibrated zircon (Group I). Based on the Group I zircon population from the compiled zircon dataset interrogated here, we infer a local mantle redox state of ΔFMQ 2.5 ± 1.1 (n= 23). This result overlaps with upper mantle redox values Δ FMQ $\pm$ 2 via $Fe^{3+}/\Sigma Fe$ on spinel peridotites (Frost and McCammon, 2008). Such variation could be a true variation of natural melt redox state,

or due to uncertainty of trace element measurements. The main sources of uncertainty of this inferred redox state are: (1) unfiltered inclusions; (2) variable source melt sources and compositions (and potential magma mixing); (3) extreme fractional crystallization; (4) analytical uncertainty in trace element measurements; and, (5) inaccuracy of the melt redox state calculation (Eqn. 2.1).

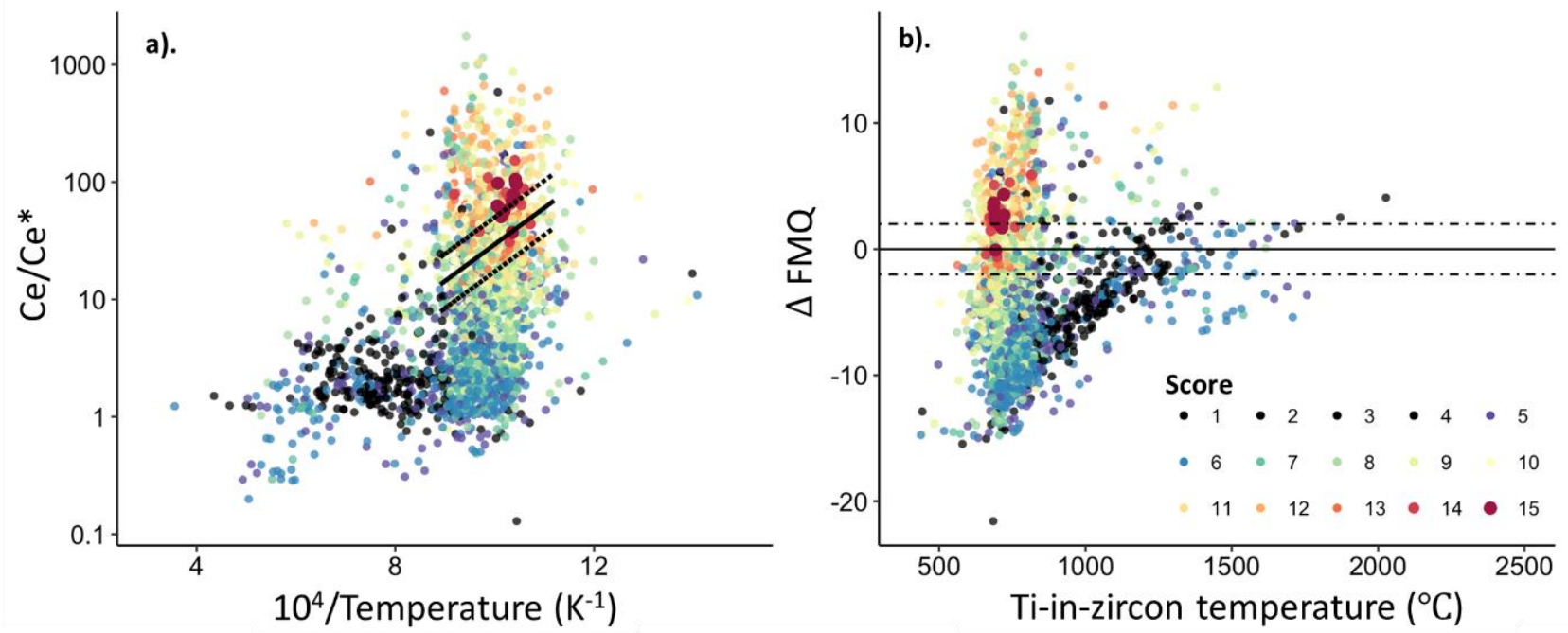


Figure 2.9. Application of the proposed filter scheme and scoring system to the zircon dataset compiled here (n = 2173). In general, the large spread in observed Ce/Ce* may be attributed to distinct sources of source melts, magmatic differentiation prior to zircon saturation, and/or mineral inclusions in trace element measurements. The black dotted and solid lines represent ΔFMQ±2.

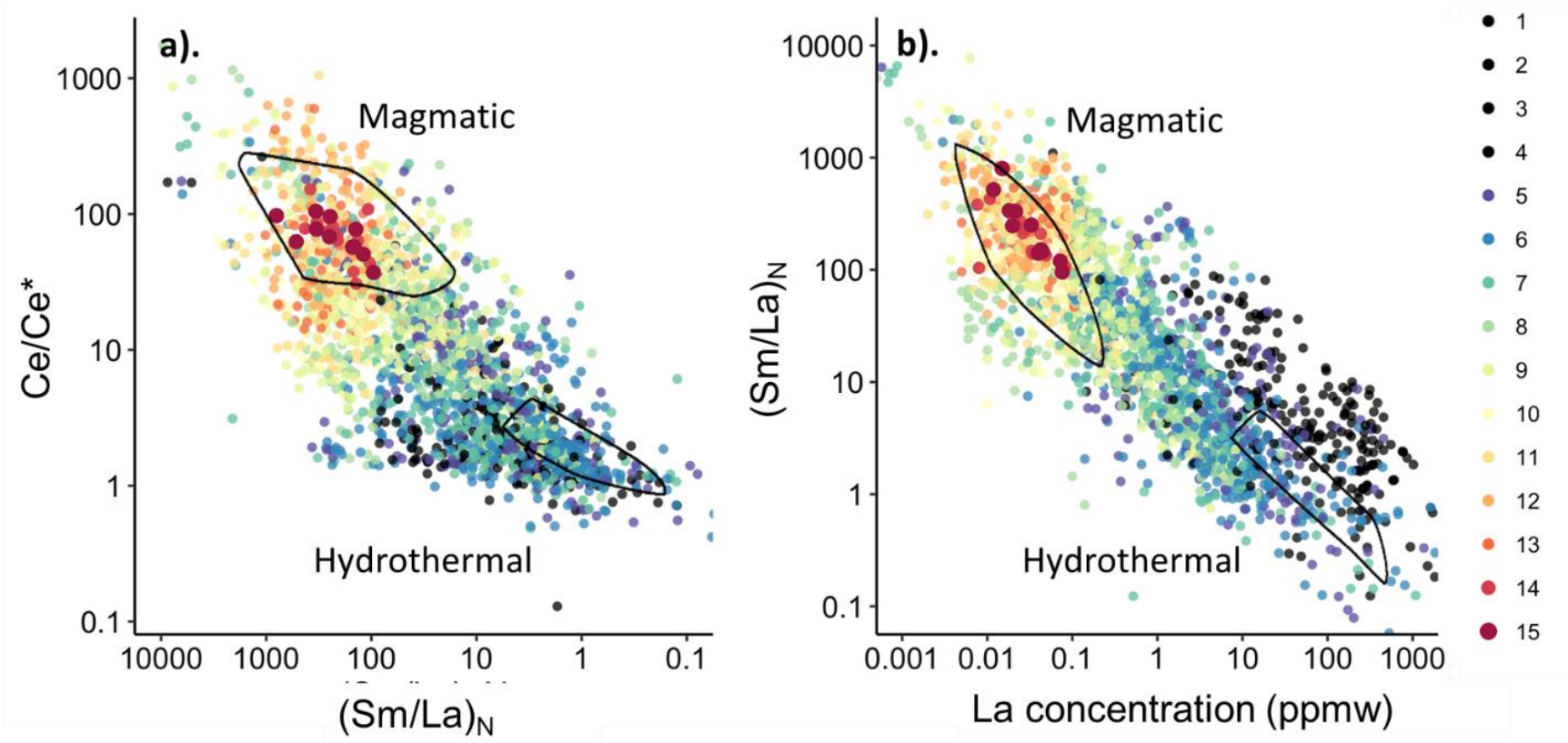


Figure 2.10. Distribution of the compiled zircon data set within the pre-established magmatic and hydrothermal fields proposed by Hoskin (2005).

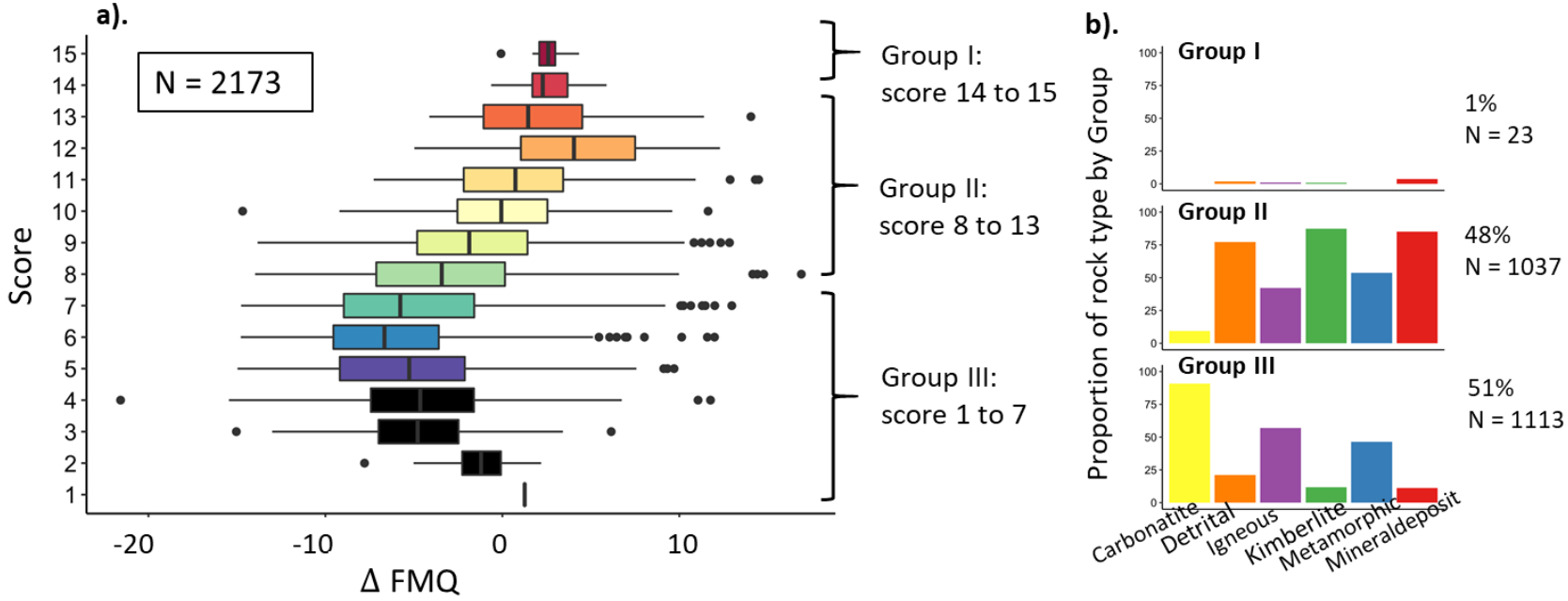


Figure 2.11. (a) Boxplots of zircon populations for each score. Based on the median and variance of inferred melt redox states, these populations may be divided into three groups: Group I (score 14 to 15), Group II (score 8 to 13), and Group III (score 1 to 7). (b) Selectivity and distribution of different rock types within each group (the percentage represents the proportion of all zircon that fall within each group).

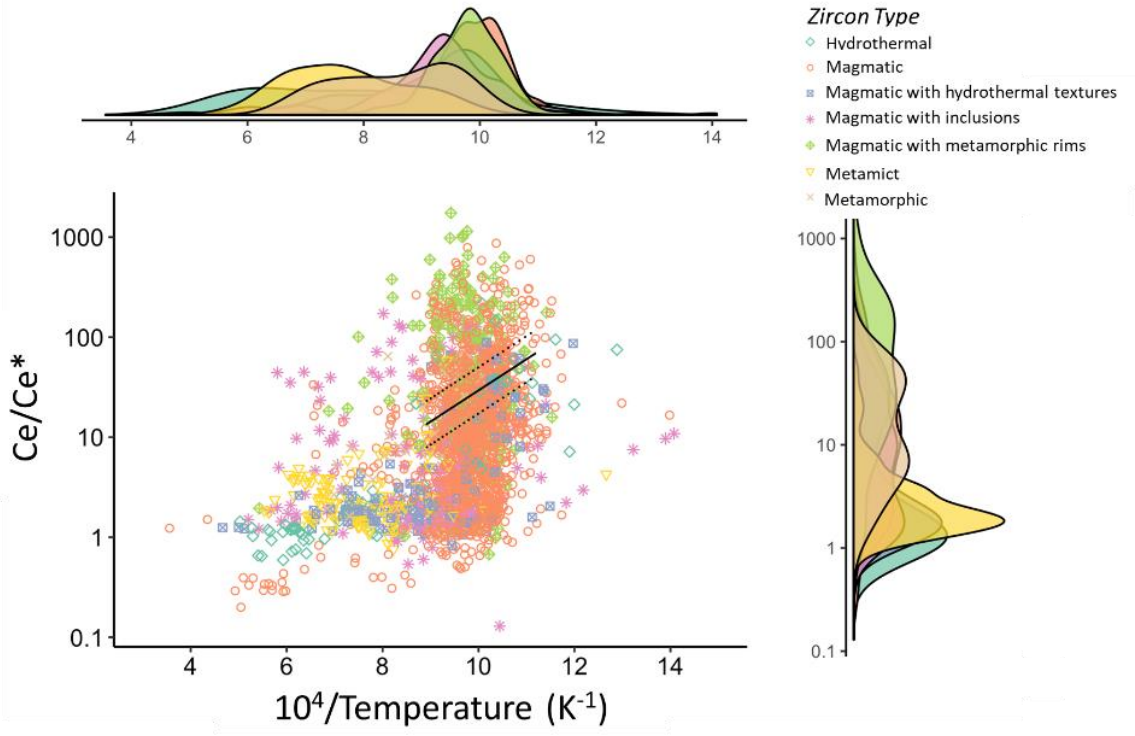


Figure 2.12. Plot of Ce/Ce* and temperature in zircon as a function of morphology and textures. The texture of each individual zircon was deduced from BSE and CL imaging from source references. The black dotted and solid lines represent ΔFMQ±2.

2.4.2. Validation of filter scheme with Jack Hills zircons

We evaluate the performance of filter scheme using Jack Hill zircon (n=53) (data from Cavosie et al., 2006 and Fu et al., 2008). Because this dataset lacks of Nb measurement, U/Yb and Th/Yb are used as alternatives (see 2.3.2.2 for justification).

This filter scheme has selected 7 mantled equilibrated zircon and 6 of them are characterized by mantle-equilbrated oxygen isotopic signatures, oscillatory magmatic zoning, typical magmatic trace element patterns (commonly referred as type 1 zircon), and classified to have a source melt composition of granitoid > 70% SiO_2 according to CART classification (**Figure 2.13**). The five unselected zircon (score < 14) but with mantle-equilibrated oxygen isotope signature, exhibit patchy zoning and/or estimated to have a source melt composition of granitoid < 65% SiO_2, implying potential post-magmatic replacement and early zircon saturation in a less felsic source melt due to Zr enrichment from an open magmatic system. In sum, this validation study highlights the sensitivity of this filter scheme at selecting mantle-equilibrated zircon.

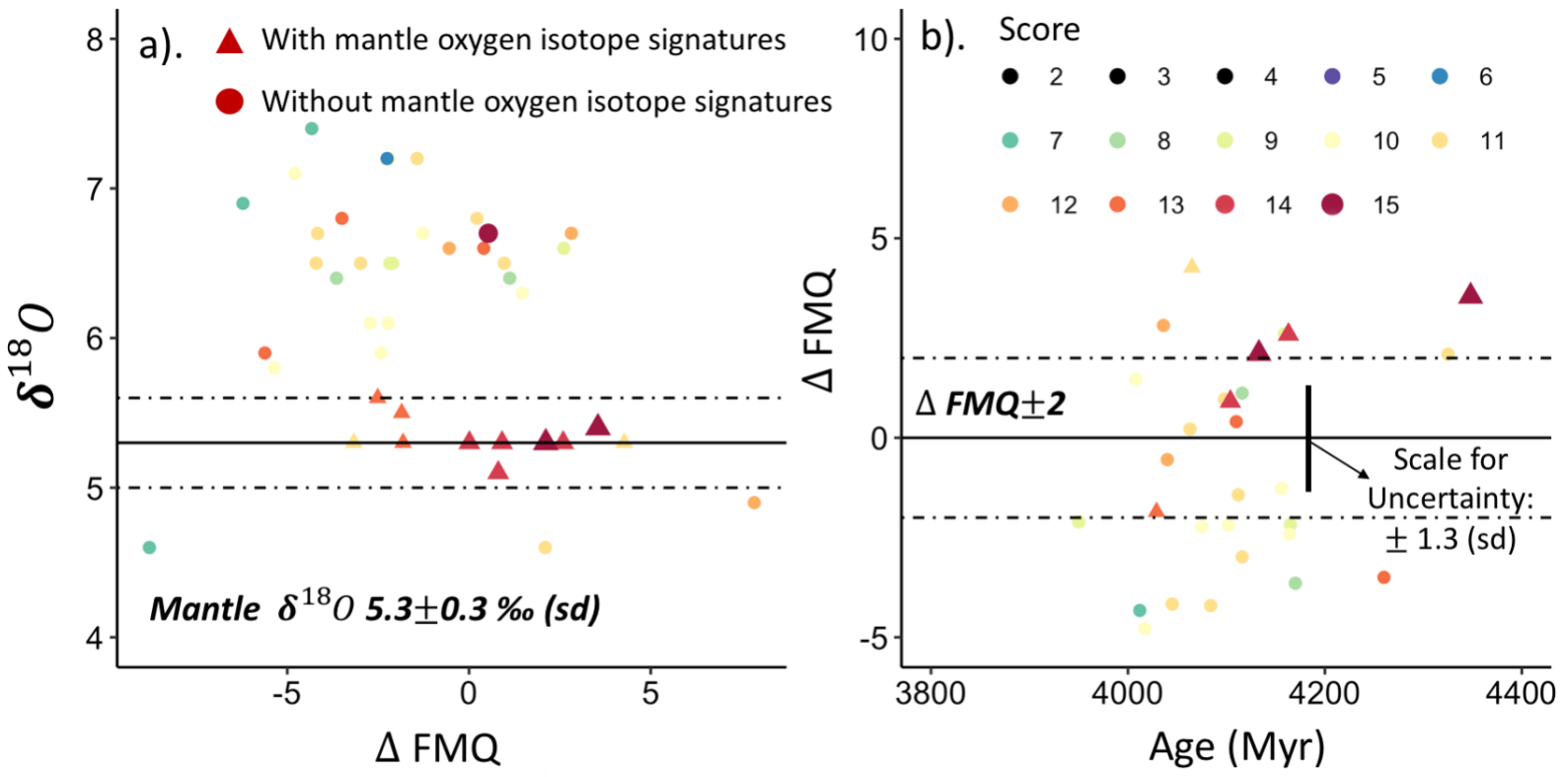


Figure 2.13. Application of filter scheme to Jack Hill zircons (n=53). (a). The black dotted and solid lines represent mantle equilibrium state and has $\delta^{18}O$ 5.3±0.3 ‰ (sd). This filter scheme has selected 7 mantle-equilibrated zircons and 6 of them are supported by mantle-equilibrated oxygen isotopes. The average Hadean magma oxygen fugacity is determined to be ΔFMQ 1.5 ± 1.3 (sd). (b). The black dotted and solid lines represent ΔFMQ±2. The Hadean mantle has oxidized to FMQ region since 4400 Ma.

The average inferred fO_2 of these 7 selected mantle-equilibrated zircon is ΔFMQ 1.5 ± 1.3 (sd), overlapping with the average oxygen fugacity ΔFMQ 1.4 ± 2.3 (sd) determined from the same dataset in previous study Trail et al., 2011. Though some data lack of U-Pb measurements, the Hadean mantle has oxidized to FMQ region since 4400 Ma (**Figure 2.13b**). Such agreement validates the performance of this proposed method and corroborates the accuracy of oxygen fugacity determination in this study with an uncertainty of 1.3 log unit.

2.4.3. Case study: Application of filter scheme to TTG zircon from Greenland

Further, this filter scheme is applied to TTG zircon (n=76) from three rock samples (524906, 524907,524994), as a case study to demonstrate the use of this filter scheme at promoting zircon research from multiple levels. First and foremost, based on distribution of the score, presence of all three classes of zircon (Groups I, II, and III) implies multiple zircon populations in the host rocks. However, most zircon in these three samples are characterized as Group III, suggesting that majority of samples have experienced significant degrees of hydrothermal alterations and/or metamorphism that led to either recrystallization as non-magmatic zircon (**Figure 2.14b**). In particular, rock sample 524994 has the least amount Group III zircon, implying that this rock sample has experienced the least amount of modification. By using the trace element composition of the highest score zircon grain (the most likely mantle-equilibrated candidate), we infer a local melt redox state of ΔFMQ -0.5±1.3 since 2950 Ma. The

uncertainty of melt redox state determination is adopted from section 2.5.2. Again, such result agrees with the inferred redox state of upper mantle redox via other proxies (ΔFMQ $\pm$ 2, Frost and McCammon, (2008)).

In summary, this filter scheme not only helps to select mantle-equilibrated zircon, but also provides quick insights about of zircon saturation, petrogenesis of zircon textures and morphology, and source melt conditions that facilitate zircon research by prioritizing magmatic zircon crystals or whole rock samples that are least likely to be contaminated, or chemically overprinted. Other obvious applications of this filter scheme include studies of both out-of-context and in-context zircon, providing a quick guideline for directing more comprehensive and lab-intensive whole rock, imaging, and isotopic investigations.

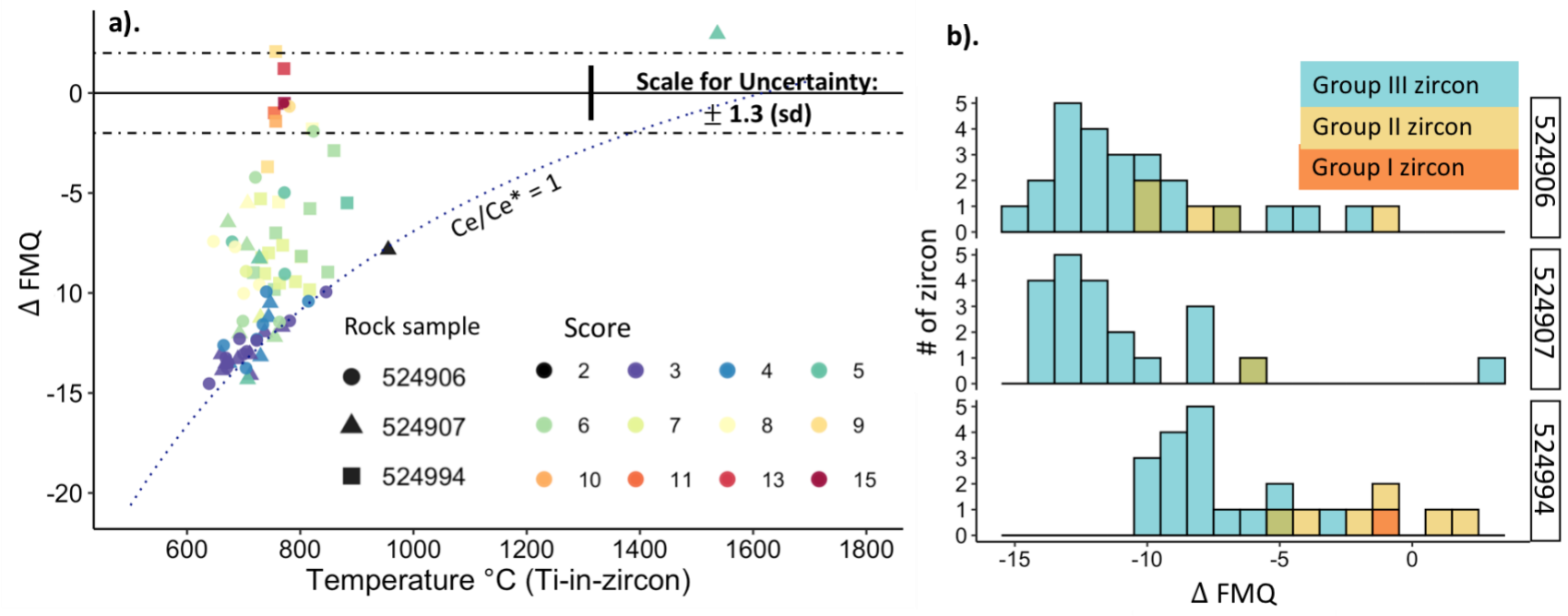


Figure 2.14. Application of the filter scheme proposed here to a case study of TTG zircon from three independent samples (524906, 524907, 524994) from Greenland. (a) Inferred melt redox state as a function of temperature. (b) Distribution of zircon by group.

2.5.Conclusion

We have established a simple filter scheme for selecting of mantle-equilibrated zircon using only trace element systematics (P, Ti, Y, Nb, REE, Hf, Th, U) as a way to circumvent more extensive analytical meausrements/techniques when time, access to lab facilities, and financial resources are limited. This filter scheme has demonstrated the capability of isolating non-magmatic zircon (Group III scores ≤ 7), magmatic zircon with inclusions or derived from enriched source melt (Group II scores between 8 and 13), and mantle-equilibrated zircon (Group I score ≥ 14). The causes of variation in inferred redox states Include: (1) unfiltered inclusions; (2) distinct source melt sources and potential magma mixing; (3) variable degrees of magma differentiation; (4) analytical uncertainty in trace element measurements; and, (5) inaccuracy of the melt redox state calculation. Despite these uncertainties, this filter scheme is validated by identifying mantle-equilibrated zircon from a well-characterized Jack Hill zircon dataset without requiring traditional protocols (i.e., petrographic studies, whole rocks analysis, and oxygen isotopes), and deriving agreeable mantle oxygen fugacity to previous studies with uncertainty at 1.3 log unit (sd). Further, without contextual information, the selected mantle-equilibrated zircon from a case study of Greenland TTG zircon using this filter scheme implicate a mantle source ΔFMQ -0.5 since 2950 Ma. In addition to select mantle-equilibrated zircon, the filter scheme is capable of facilitating zircon research at multiple levels. The case study demonstrated that this filter scheme provides quick insights about petrogenesis of zircon, including zircon saturation, textures, and morphology, and source melt conditions, thus selecting/prioritizing magmatic zircon crystals or whole rock sample that are least

likely to be chemically overprinted. Altogether, the validation test and the case study suggest this filter scheme can be extended to studies of out-of-context and in-context zircon in the future.

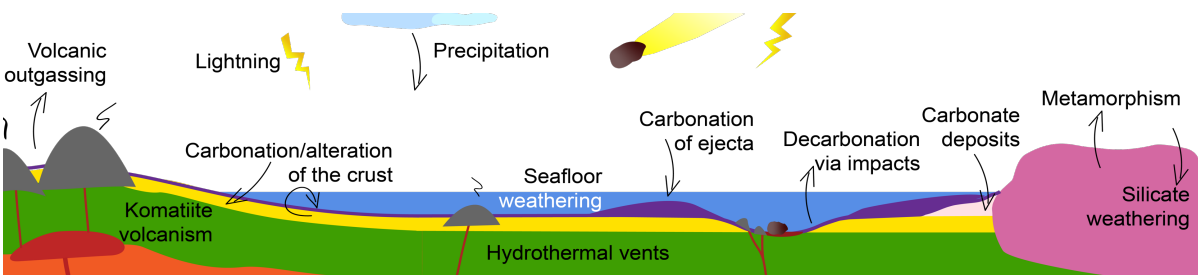
2.6.Acknowledgment

This study was financially supported by the NASA grant (80NSSC18K1612). Special thanks are given to my parents, professors, and groupmates for their guidance and unconditional supports on this study. Thank you, Dr. Dustin Trail for his kind encouragements. The compiled dataset was acquired from 30 previously published works, which are listed in the supporting information file (**Appendix A1**). The self-collected Greenland dataset was submitted to Earthchem (under review) and also accessible in the supporting information (**Appendix A2**).

Production of Organic Precursors via Crater-forming impacts and Its Implication to Early Earth, Mars, and Ceres

Ziqin Ni[1*], Benjamin Farcy[1], Ricardo Arevalo Jr.[1], Michael Eller[2], Veronica T. Pinnick[3], Emile A. Schweikert[4], William B. Brinckerhoff[3]

[1]University of Maryland, College Park, MD, 20742.
[2]California State University Northridge, CA, 91330.
[3]NASA Goddard Space Flight Center, Greenbelt, MD, 20771.
[4]Texas A&M University, College Station, TX, 77843.



Chapter 3: Production of Organic Precursors via Crater-forming impacts and Its Implication to Early Earth, Mars, and Ceres

Quick overview of the chapter:

Crater-forming impacts on planetary surfaces, though often considered as catastrophic events, could have produced a significant amount of critical precursor compounds that enable organic synthesis in a lifeless world. Inorganic matrices of carbonates and nitrogen salts have been observed on the surface of Mars, Ceres, and are expected to be present in large quantities on the surface of early Earth. The motivation of this research is to investigate the ability to generate new molecules via impact processing on carbonates and nitrogen salts by high power laser and projectile bombardment of C60+ onto N-rich substrate in the laboratory. The production yields in the microscopic laser-generated vapor plume were scaled up to replicate the volume and duration of vapor plume generated by planetary-scale impacts, and a total production yield was estimated based on numeric modelling. The precursor compound studied in this research, CN-, is not stable by itself and requires polymerization above an abundance threshold to produce more stable molecules. Do these impact-generated organic compounds degrade or undergo chemical evolution in the planetary setting? This chapter discusses the specific cases of early Earth, Mars, and Ceres.

In this chaper, I acquired and processed the data, interpreted the results, made the figures, and wrote the paper. This chapter has been submitted to the Astrobiology journal and it is currently under review.

3. Abstract

Crater-forming impacts on planetary surfaces deliver a significant amount of energy that is capable of producing precursor compounds like cyanides that are critical to formation of organic molecules. To understand the chemical processes of impact-induced organic chemistry, we use both momentum-based and photon-based impact experiments to study the post-impact recombination products. A beam of mm-size C_{60}^+ projectiles colliding onto an N-rich substrate was used to simulate the shock process and plume generation of planetary impacts. Organic precursors (specifically CN^- and CNO^-) produced during the experiment indicate recombination of C and N in the plasma plume. Additionally, a high-power laser directed onto carbonate and N-salt was also used to simulate post-impact plasma generation during planetary impacts and corroborate the production of organic precursors. The production rate of CN^- determined from the laser simulation experiments was used to infer HCN yields as a function of impactor size, and the volume and longevity of vapor plumes. In the case of early Earth, with an N_2-dominated atmosphere and a shallow marine, a total of 10^{15}-10^{16} kg HCN could have been produced via surface impacts during a 100 Ma interval in the Hadean eon. The CO_2-rich atmosphere and carbonate-rich surface of early Mars could have supported the production of ~10^{12} kg HCN via impact processes during a 100 Ma interval in the Noachian Period. In Ceres that is lack of significant atmosphere and is covered with carbonate and N salt, an impactor > 10^{12} kg could have melted the subsurface icy layer and concentrate 10^5-10^{11} kg impact-derived HCN per event in shallow liquidus reservoirs that would support further polymerization. Future studies could use the model provided here to estimate the production of HCN via planetary

impacts in different hypothetical settings and explore the possibility of HCN polymerization to facilitate molecular evolution.

3.1.Introduction

Meteoritic impacts are universal processes that transport exogenous and endogenous matters between inner and outer solar system and reprocess (sub)surface materials in planetary settings. Massive impactors (e.g., diameter > 200 m) traveling at 18 km/s can deliver substantial amount of kinetic energy (> 10^{19} joules) that would be transferred to shake, fracture, and compress the ground. The enormous tension built up at the impact site can lead to vaporization and ionization of surrouding materials. Exogenous matters from the impactor and endogenous materials from the planetary environment would be ultimately mixed and recombine to produce new molecules as the vapor plume cools (Managadze et al., 2003a). New molecules generated from such far-from-equilibrium impact process could be highly reactive and catalyze the synthesis of essential organic molecules from inanimate substances in the planetary environments.

The majority of life essential elements carbon (C) and nitrogen (N) are stored in geological reservoirs including atmospheres and rocks. The simplest heterogenous product of C and N is cyanide (nominally as the anion CN^-), which can be oxidized to produce cyanate (CNO^-), or neutralized to form hydrogen cyanide (HCN) and/or fulminic acid (HCNO). Cyanide ions (and their derivatives) are critical organic precursors because they can polymerize spontaneously under a wide range of conditions (temperature, pressure, type of solvents, etc.) to form HCN polymers -$(H_xC_yN_z)_n$- (Das et al., 2019). Under various physical and chemical conditions that

were regulated in the laboratory, HCN polymers were observed to produce essential organic monomers in the development of life such as amino acids (Oró & Kamat, 1961), nucleobases (Borquez et al., 2005), sugars (Botta et al., 2017), and fatty acids (Takahashi et al., 1990). These monomers could be polymerized into a variety of informational polymers, especially peptides/proteins (Matthews & Moser, 1967), RNA/DNA (Kitadai & Maruyama, 2018), to lipids (Saladino et al., 2012) that serve specific biological functions. To date, HCN and HCN polymers have been detected in comets (Fray et al., 2005; Matthews & Ludicky, 1992), the atmospheres of Jupiter and Titan (Hanel et al., 1981; Matthews, 1997; Tokunaga et al., 1981), molecular clouds (Rank et al., 1971), and star forming regions (Graninger et al., 2014). The Murchison meteorite has been shown to contain at least 600 different HCN polymers with chain lengths ranging from C_6 to up to C_{45} (Naraoka et al., 2017). Thus, understanding the synthesis of HCN can provide constraints on the origin and distribution of primordial organics, chemical evolution of complex organic molecules in absence of biology, and may inform the statistical probability of developing life in the planetary environments.

In the laboratory, HCN and a multitude of cyanide radicals have been produced via ultraviolet photolysis (Ferris & Ishikawa, 1988; Gerakines et al., 2004), electrical discharge (Ardaseva et al., 2017; Chameides & Walker, 1981; Ferus et al., 2017), and interaction of the atmospheric gases with energetic particles (Airapetian et al., 2016). HCN and amine have also been produced via intense shocks created by a high-power laser in a gas mixture of N-bearing gas (N_2, NH_3), C-bearing gas (e.g., CO, CO_2), and H-bearing gas (e.g., H_2, H_2O), suggesting the feasibility of producing organic molecules via impact shocks of atompshere (McKay & Borucki, 1997). Similar organic

molecules were observed when a carbonaceous projectile is entering a chamber filled with nitrogen gas molecules, suggesting such energetic impact process could harness C and N from inert species in the atmosphere, as well as those in the projectile (Kurosawa et al., 2013; Parkos et al., 2018).

The early Earth may have contained significant deposits of (bi)carbonates and nitrogen-rich salts (or N salts) on the surface. For example, chemical weathering of ultramafic rocks in a CO_2-rich atmosphere would have produced significant quantities of calcite ($CaCO_3$), sodium bicarbonate ($NaHCO_3$), siderite ($FeCO_3$), ankerite ($Ca(Fe,Mg,Mn)(CO_3)_2$), and magnesite ($MgCO_3$) (Frondini et al., 2019; Kelemen et al., 2020). Any N_2 in the atmsphere would have faciliatetd the fixation of ntitrogenous salts via electric discharge and/or atmospheric shock (Summers, 2005; Summers & Chang, 1993). Comparable potential exists for abiotic impact reprocessing of carbonates and N salts on other planetary surfaces, such as weight percent level carbonates (calcite, siderite, magnesite) (Bandfield et al., 2003) and ppm-level N on Martian surface (Stern et al., 2015), and collocation of NH_4-phyllosilicates and Mg,Ca-carbonates observed on the surface of Ceres (Ammannito et al., 2016; Carrozzo et al., 2018; De Sanctis & Ammannito, 2021). Thus, solid planetary surfaces likely supplied significant sources of C and N to produce organic precursors via crater-forming impacts.

The primary goal of this study is to investigate the capability of synthesizing organic precursors (specifically cyanide ions) via laboratory simulation of hypervelocity impacts on a solid surface of carbonate and/or N salt. We used both oxidized nitrate

salts (NO_3^-) and reduced ammonium salts (NH_4^+) to explore the role of redox conditions in molecular synthesis via crater-forming impacts. Empirical yields of cyanide ions are then translated into estimates of HCN production derived from impact-driven plumes on the surfaces of early Earth, early Mars, and Ceres. These estimates provide a baseline to infer the abundance of impact-produced organic precursors, and the possibility of them being concentrated enough to enable polymerization and advance chemical evolution in planetary environments.

3.2.Materials and Methods

The physical and chemical processes associated with planetary impacts have been simulated experimentally via projectile bombardment, which reproduces the transfer of momentum between the impactor and target (DeBord et al., 2013; Kurosawa et al., 2013), and high energy laser processing, which simulates the plasma state of vapor plumes created by crater-forming impacts (Geohegan & Puretzky, 1995; Kushwaha & Thareja, 2008; Managadze et al., 2003b; McKay & Borucki, 1997). To increase the confidence of our results, we simulated the crater-forming impacts using both techniques: physical impact via a beam of isotopically enriched, charged fullerene projectiles ($^{13}C_{60}^+$) and high-power pulsed laser ablation (**Figure 3.1**). Both laboratory experiments investigated the chemical products via mass spectrometry in negative ion mode; thus, the formation of HCN was detected as deprotonated CN^-.

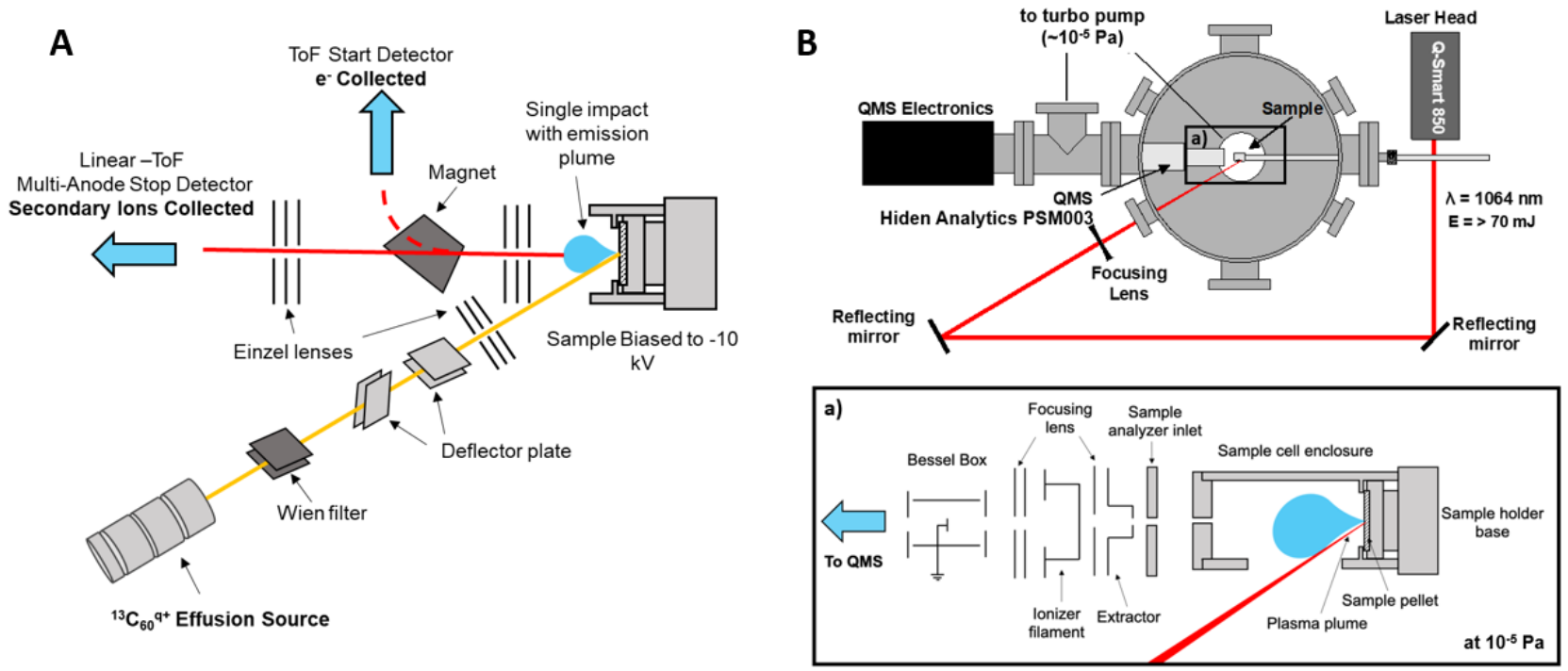


Figure 3.1. A. Simulation of a crater-forming impact on planetary surface using a beam of fast-moving $^{13}C_{60}^{+}$ projectiles that accelerated towards a substrate rich in nitrogen salts. **B.** simulation of vapor plume generated by a planetary-scale impact via high-power laser irradiation on a substrate rich in nitrogen salts and carbonates.

In the projectile bombardment experiment, we accelerated a beam of isotopically enriched buckminsterfullerene $^{13}C_{60}^{+}$ (98 atom% ^{13}C; MER), an analog to a carbonaceous impactor, onto powdered ammonium nitrate NH_4NO_3 (> 98 atom% purity; CIL, CAS: 6484-52-2) that represents an N-rich planetary surface. The experiments were conducted using the Secondary Ion Mass Spectrometer (SIMS) at Texas A&M University. Isotopically doped $^{15}NH_4NO_3$ (> 98 atom% ^{15}N; CIL, CAS: 31432-48-1) and $NH_4{}^{15}NO_3$ (> 98 atom% ^{15}N; CIL, CAS: 31432-46-9) were used as reactants in follow-on experiments to distinguish the sources of C and N from any potential sample surface contaminants and/or residual gases from the background.

For each experimental trial, ≤ 100 mg powered sample was deposited onto a customized stainless-steel sample holder that was compacted and secured using a set screw. A 90% transmission nickel grid was added to the surface of the sample holder to eliminate surface charging. A stainless-steel collar that mates to the grid was also added through a friction connection to secure the sample. Ions produced in the plasma plume were transmitted into a time-of-flight (ToF) mass analyzer and detected via multichannel plate (**Figure 3.1A**; see (Eller et al., 2013) for more details about the instrument setup). The simulation chamber was held at near vacuum ($0.5\text{-}1.0\times10^{-4}$ Pa) to reduce the contribution of C and N from atmospheric gases.

In the laser simulation experiment, we irradiated a solid substrate of carbonates and N salts using a Q-switched pulsed laser (Q-smart 850, Quantel Corp.) at 1064 nm to simulate vapor plumes generated via crater-forming impacts on a C- and N-rich planetary surface. This laser setup offered a minimum laser energy of 70 mJ and a pulse

width of 9 ns, enabling irradiances at least 10^9 W/cm^2 that can vaporize and ionize targets of carbonates and dry ice (Schultz, 1996), graphite (Bulgakova & Bulgakov, 2001), silicates and metals (Lu et al., 2002). This laser-generated vapor plume is modeled to have the similar pressure and thermodynamic states of vapor generated by hypervelocity impacts at ≥ 20 km/s independent of impactor mass (Fletcher et al., 2015; Kadono et al., 2002), and thus providing an efficient approach to study the chemical species that could be produced from crater-forming impacts on planetary surface.

Each experimental trial used a 300 mg sample mixture that consisted of reagent-grade calcium carbonate $CaCO_3$ (> 99.95-100.05% dry basis; Sigma Aldrich, 398101) and one of two N-salt components: ammonium chloride $N[-3]H_4Cl$ (MP Biomedicals 02150109-CF); or, sodium nitrate $NaN[+5]O_3$ (>99 % chemical purity; Sigma Aldrich, 221341). The powered mixture was admixed with a minimum amount of isopropanol to create a slurry and homogenized using a vortex mixer (Lab Dancer; IKA) and transferred to a pellet die via spatula. After the residual isopropanol evaporated, the homogenized mixture was compressed into a flat sample disk with uniform thickness (1.5 mm) using a hydraulic pellet press (25-ton Ring Press; Research & Industrial Instruments Co.). The solid sample disk was then transferred onto a sample holder, which was then inserted into the ablation chamber was held at near vacuum (1.0×10^{-5} Pa) as the ablation target (**Figure 3.1B;** see Appendix B for more details about the instrument setup of the laser simulation experiment). Similar to the projectile bombardment experiment, isotopically doped reactants $Ca^{13}CO_3$ (>99 atom% ^{13}C; Sigma Aldrich, 492027), $^{15}NH_4Cl$ (> 98 atom% ^{15}N; Sigma Aldrich, 299251), and $Na^{15}NO_3$ (>98 atom% ^{15}N; Sigma Aldrich, 364606) were used in follow-on

experiments. In between experimental runs, sample containers were cleaned with detergent, then rinsed and washed repeatedly using acetone, isopropanol, and dichloromethane before use.

3.3.Results

3.3.1. Molecular recombination in $^{13}C_{60}^{+}$ bombardment experiment

We observed $^{13}CN^{-}$ and $^{13}CNO^{-}$ peaks at mass to charge ratio (m/z) 27 and 43 respectively when the $^{13}C_{60}^{+}$ beam bombarded the NH_4NO_3 rich target (**Figure 3.2**). The peaks shifted predictably by one atomic mass unit when isotopically doped $^{15}NH_4NO_3$ or $NH_4{}^{15}NO_3$ was used, indicating the recombination of ^{13}C and ^{15}N derived from the solid reactants. Migration of NO_2^- peak by one atomic unit as the $NH_4{}^{15}NO_3$ was used suggests that NO_2^- was primarily the decomposition product of nitrate salts, instead of recombination products of N and O atoms in the plume. However, peaks at m/z 25, 26, 27, and 42 were observed across all experiments, signaling the presence of a measurable amount of ^{12}C contamination in the vacuum chamber, potentially derived from hydrocarbon pump fluid (i.e., polyphenyl ether and polymerizing ethylene). Despite this contamination, the generation of double-labeled $^{13}C^{15}N^{-}$ and $^{13}C^{15}NO^{-}$ nonetheless represents the recombination of reactants derived from the solid starting materials.

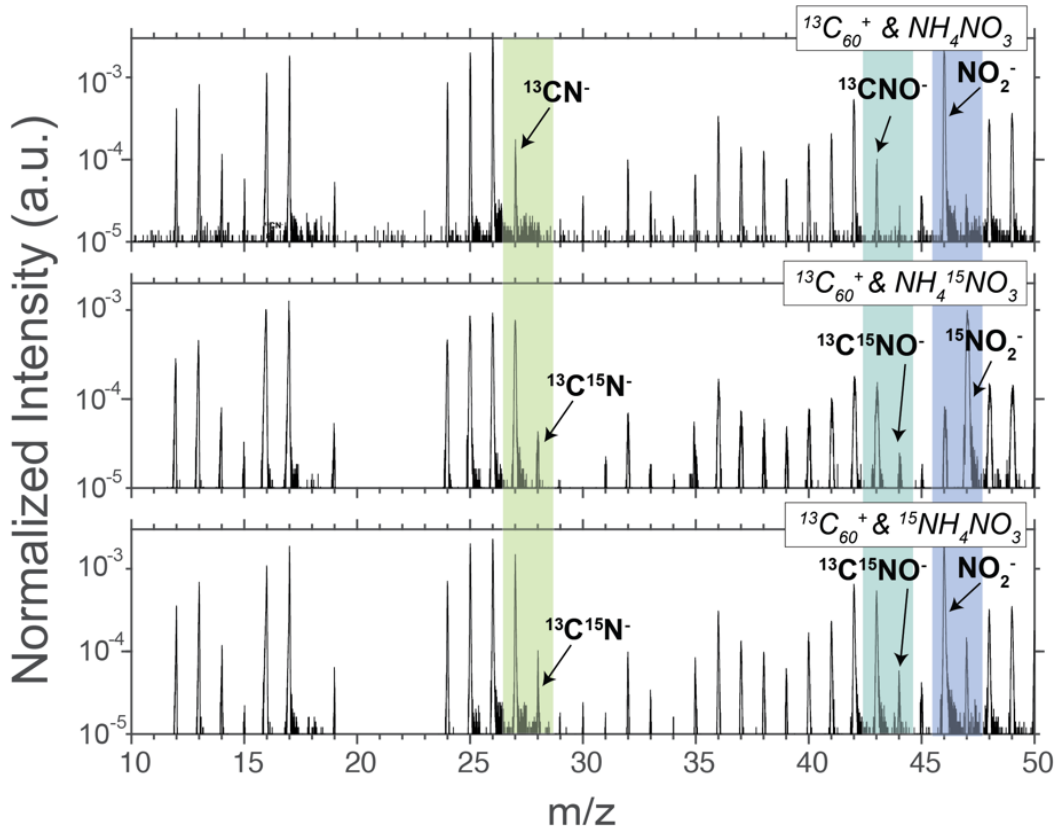


Figure 3.2. Synthesis of $^{13}CN^-$ and $^{13}CNO^-$ using a beam of fast-moving $^{13}C_{60}^+$ projectiles bombarding towards a solid substrate of ammonium nitrate (NH_4NO_3) at velocities >70 km/s. The uppermost panel shows mass spectra from experiments using unlabeled NH_4NO_3, while the middle panel and bottom panel are mass spectra from experiments using $NH_4{}^{15}NO_3$ and $^{15}NH_4NO_3$, respectively. The intensity on the y-axis is normalized to the total number of counts and reported as the average of the total number of measurements.

3.3.1. Molecular recombination in laser simulation experiment

Laser ablation of the sample target consisting of $CaCO_3$ and $NaNO_3$ yielded mass spectral peaks at m/z 26, 42, 46, consistent with the nominal masses of CN^-, CNO^-, and NO_2^- (**Figure 3.3A**). When the ablation target was prepared using isotopically doped $Ca^{13}CO_3$, the peaks at m/z 26 and 42 shifted to m/z 27 and 43. Similar shifts of the nominal CN^-, CNO^-, and NO_2^- peaks by one atomic mass unit were observed when using isotopically labeled $Na^{15}NO_3$. When both isotopically labeled materials were mixed together, the characteristic peak CN^- and CNO^- shifted by two atomic mass units, and NO_2^- shifted by one atomic mass unit. Other prominent peaks observed in the mass spectra include O^- (m/z 16), OH^- (m/z 17), and $^{35}Cl^-$ and $^{37}Cl^-$ from the following NH_4Cl experiments. Despite potential contamination from various sources, these three isotopically labeled experiments together reinforce the recombination of C and N sourced from the solid starting materials.

In experiments of $CaCO_3$ and NH_4Cl, comparable mass peaks were observed at m/z 26 and 42, which migrated predictably with the application of isotopically doped reactants, demonstrating the formation of CN^- and CNO^- from the carbonate and ammonium-bearing ablation target (**Figure 3.3B**). Two Cl isotopes at m/z 35 and 37 decomposed from NH_4Cl were also observed. The detection and identification of CN^- and CNO^- in both $NaNO_3$ and NH_4Cl experiments in near vacuum condition (1.0×10^{-5} Pa) demonstrated the feasibility of synthesizing new molecules irrespective of the redox state of the target material and without the need to invoke an atmosphere.

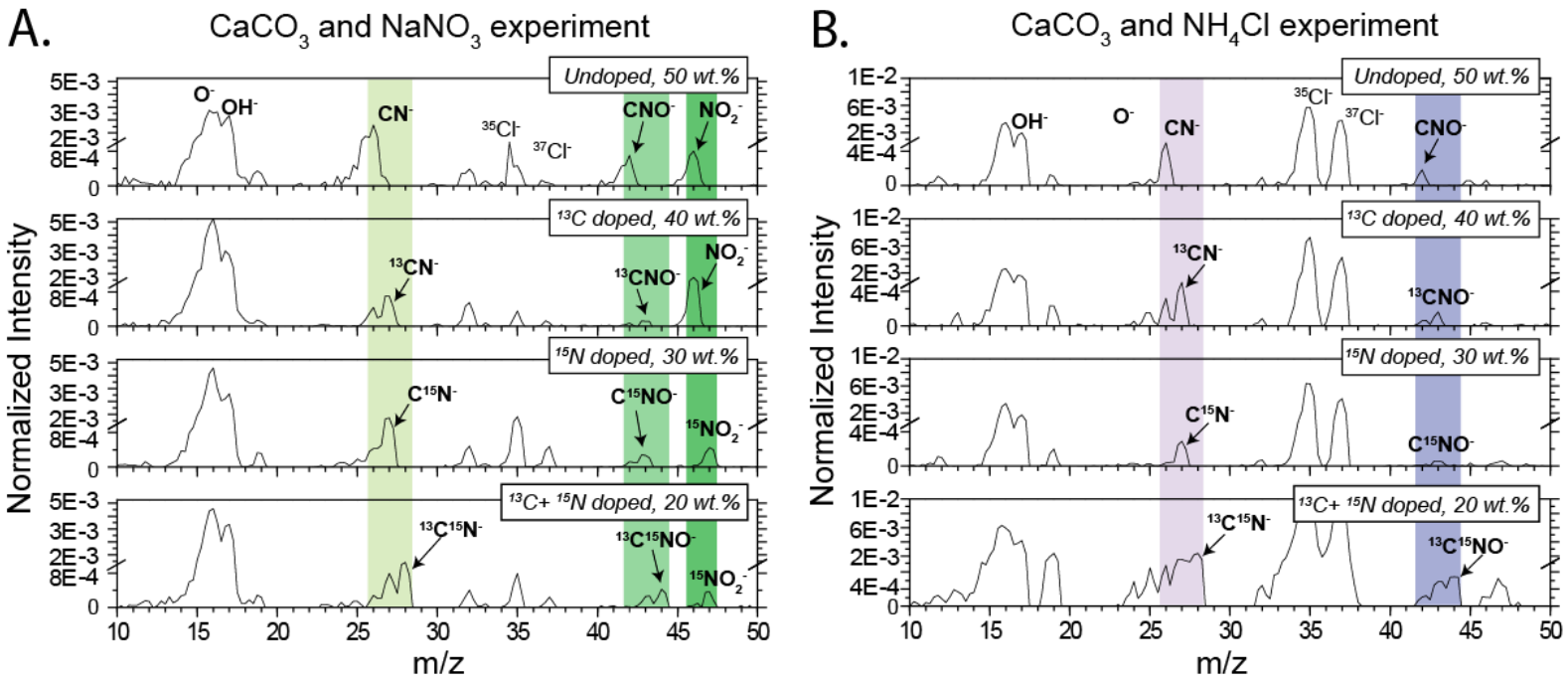


Figure 3.3. Mass spectra of ions generated in the vapor plume created via high power pulsed laser ablation on a solid target consisting of (A) $CaCO_3$ and $NaNO_3$ and (B) $CaCO_3$ and NH_4Cl powder mixture under near vacuum conditions (1.0×10^{-5} Pa). The intensity on the y-axis is the summation of 20 mass spectra recorded at a specific ion energy, between 1 to 20 eV; above 20 eV, the number of ions decayed exponentially. The 20 spectra were summed into a single spectrum to increase the signal to noise ratio and integrate ions of different energies, and then normalized to the total number of counts in the experiments. The background of each spectrum has been subtracted. The predictable peak shifts starting at m/z 26 (CN^-) and 42 (CNO^-) highlighted in both experiments demonstrate the recombination of C and N from solid. The raw data are accessible in the Appendix B.

3.3.2. Linear regression analysis of CN^- from laser simulation experiments

Peak intensities of impact-derived CN^- at the corresponding mass station were extracted to establish calibration curves (i.e., linear regression analyses) of CN^- production yields at various weight percentages of N salts in the substrate (**Figure 3.4**). The intercepts of both linear regression models were set to zero under the assumption that no CN^- should be produced from the solid starting materials when N salts are absent in the target. Under the same weight percentage of N salts, the regression results suggest that 2× more CN^- is produced when nitrate (the oxidized form of N-salt) is used in the sample target in place of ammonium. However, caution is warranted as the detectability of CN^- is not equivalent to the production rate of neutrals like HCN and HCNO because neutrals fail to be detected by the mass analyzer.

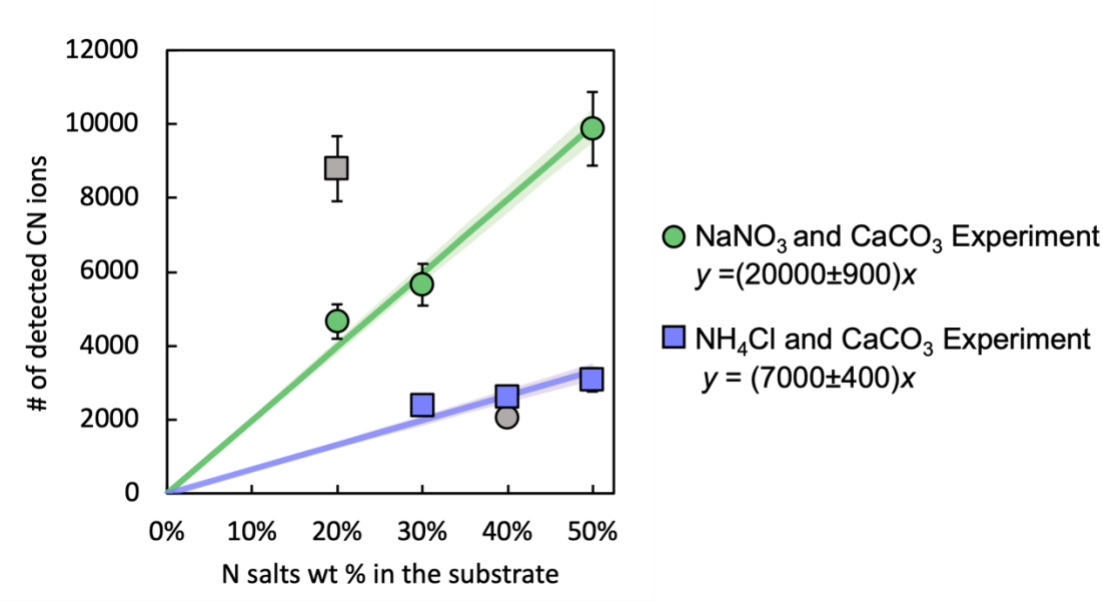


Figure 3.4. Calibration curves between the number of detected CN^- ions and the weight percentage of N salts as starting materials in the laser simulation experiments. All extracted peak intensities, reported as counts per second, were first subtracted from the noise floor and then normalized by the 20 ms data acquisition time at each mass station. The intercepts of the linear regression curves were set to be zero. The errors of measurements are estimated using 1 relative standard deviation (1 RSD) of the noise floor. The shaded areas represent the predicted number of CN^- based on linear regression results at 1 SD. Outliers are identified in gray; the outliers in $NaNH_3$ and NH_4Cl experiments was likely due to incomplete dissociation of nitrate salt and contamination of other carbon- and nitrogen-rich residuals, respectively. The raw datasets from laser simulation experiments are available in the Appendix B.

3.4. Discussion

The detection and identification of recombined compounds especially CN^- in both simulation studies – physical bombardment with $^{13}C_{60}^+$ and high-power pulsed laser irradiation – demonstrate the capability of synthesizing new molecules via impact events on a planetary surface containing carbonate and/or N salt. Calibration curves of CN^- peak intensities detected at various weight percentages of N salts indicate CN^-

production yields increase with the abundance of N in the substrate, and by extrapolation the availability of N salts (and carbonates) on the planetary surface.

While efforts have been made using momentum-based experiments and laser ablation to estimate the size, density, and chemistry of an impact-generated plume, the relationship of these experiments to natural crater-forming impact events requires scaling up. The following discussion aims to: (1) constrain CN^- production rates as a function of the volume and longevity of laser-generated vapor plumes, (2) estimating crater morphology and longevity of meteoritic impact-generated vapor plumes on planetary surfaces, and ultimately (3) extrapolate CN^- production yields via crater-forming impacts in different planetary settings.

3.4.1. CN- production rates in laser-generated vapor plumes

The first step to quantify the CN^- production rate in laser-generated vapor plumes is to estimate the amount of carbonates and N salts that are vaporized and atomized per laser shot. High irradiance laser ablation (>10^9 W/cm^2) typically removes the uppermost 10^{-7} m of the sample per laser shot due to an explosive boiling of the substrate materials (Bulgakova & Bulgakov, 2001; Lu et al., 2002), indicating 10^{-13} m^3 (or 10^{-10} kg) of sample materials was excavated per single shot in our laser-based experiments. We assume 5 % of ejected masses would be vaporized (Schultz, 1996) and a 1-10 % atomization efficiency of vaporized materials. Given 5-50 wt.% of N salts in a target mixture, the plumes were estimated to have (0.07-9)×10^{-12} moles of N atoms and (0.5-10)×10^{-12} moles of C atoms in our experiments.

If the recombination rate of C and N atoms in the vapor plume following a logarithmic growth function, the production of CN^- would be:

$$A_t = \frac{A_o}{1+e^{-\beta t}} \qquad \textit{Eqn (3.1)}$$

where A_o is the initial concentration of limiting reagents in the vapor plume (i.e., C or N atoms in moles/m^3), A_t is the concentration of CN^- in produced in the plume (in moles/m^3) after time t in second, and β is the production rate of CN^- in the vapor plume. The explosion of mass ejection and generation of vapor plume have been observed to delay and could last from hundreds of nanoseconds to tens of microseconds in near-vacuum conditions (LaHaye et al., 2013; Yoo et al., 2000), and thus we assumed a t =1 μs as a medium value for C and N atoms from the target mixture to recombine in the laser-generated vapor plume. The detection efficiency of recombined ions was on the order of 1%, meaning the other 99% of CN^- were not detected due to the limitation of ion optics in the setup of laser simulation experiment. Based on the linear regression models constructed from the laser simulation experiment (**Figure 3.4**), about $(1.5\pm0.2)\times10^7$ CN^- ions would be produced per second. In the following calculations, we use CN^- production rate 1.5×10^7 to estimate the production yields of HCN via crater-forming impacts in different planetary settings.

3.4.2. Crater morphology and plume volume/longevity during crater-forming impacts

In the simplest scenario (i.e., a planetary target void of atmosphere and surface ocean), the mass and speed of the incoming projectile does not decrease prior to the impact

event, and thus the energy imparted to an impacted surface is equal to the kinetic energy of the impacting projectile. The diameter of a transient crater created on the planetary surface (D_{tc}) can be estimated using

$$D_{tc} = a\left(\frac{\rho_{impactor}}{\rho_{target}}\right)^{\frac{1}{3}} D_{impactor}^{0.78} v_{impactor}^{0.44} g E_{impact}^{-0.22} \left(sin\left(\theta_{impact}\right)\right)^{\frac{1}{3}}, \qquad \textit{Eqn. (3.2)}$$

where $\rho_{impactor}$ and ρ_{target} are the density of impactor and target respectively, $D_{impactor}$ is the diameter of impactor, g is the gravitational acceleration of the planetary body, and θ_{impact} is the impact angle in degrees (G. Collins et al., 2012; G. S. Collins et al., 2005; Schmidt & Housen, 1987). A proportional constant, a, is set to 1.2 or 1.4 for a target surface with water or without water, respectively. The volume of a transient crater (V_{tc}) can be estimated from D_{tc},

$$V_{tc} = \pi D_{tc}^3/(16\sqrt{2}). \qquad \textit{Eqn. (3.3)}$$

The vaporized mass fraction of a transient crater was measured to be 5% and 10% during a crater-forming impact event on a carbonate-rich and ice-rich target, respectively (Schultz, 1996). Similar to laser simulation experiments, about 1-10% of vaporized materials would be atomized for molecular recombination in the post-impact plumes. The vapor plume would expand at 10-15 times the size of impactor based on two-dimensional numerical simulations under near-vacuum condition (Melosh et al., 1993), thus the volume of vapor plume was estimated as the volume of a sphere (r_{plume}= 15$r_{impactor}$). The internal energy of the vapor plume was measured to be 2-5 % of the impact energy on a carbonate or icy target at the most probable impact angle

45° (Schultz, 1996; Svetsov & Shuvalov, 2019). This superheated vapor plume would remain its plasma state until it cooled to reach a transparency temperature capable of molecular recombination. We predicted the longevity of vapor plume by calculating the time it took to radiate the absorbed heat (i.e., 2% of impact energy) from its transparency temperature (i.e., 6,000K for atmosphere-free silicate vapor (Melosh et al., 1993)) following the Planck's law.

The presence of an atmosphere would have imposed negligible influence on the shape, energy, or momentum of projectiles because the mass of displaced atmosphere is much smaller than the mass of projectiles considered in this study. Therefore, we expect minimum changes in estimates of impact energy and crater morphology at the impact surface. The presence of an atmosphere, however, could suppress the expansion of vapor plume (Schultz, 1992, 1996) and slow down the process of radiative cooling (Sugita & Schultz, 2003a, 2003b), producing a smaller plume volume and a longer plume longevity compared to the estimates of vapor plume and longevity in near-vacuum condition. An ocean layer of water depth > 100m would slow down the projectile significantly, and thus reducing the impact energy at the surface and undermining its capability to excavate surface materials. A shallow water layer of depth ~1 m, however, can also accelerate the crater formation, allowing up to 60 vol.% more surface materials to be excavated. Details about the role of atmospheres and oceans in crater morphology and plume generation are discussed in Appendix B.

3.4.3. CN^- production yields via crater-forming impacts on early Earth, Mars, and Ceres

Based on the basic modeling assumptions described above, we can estimate the crater morphology and constrain plume dynamics of crater-forming impacts on early Earth, Mars, and Ceres. The range of physical and chemical parameters for these calculations are summarized in Appendix B.

3.4.3.1.Early Earth

The physical and chemical environment on the early Earth remain under debated but critical to the estimates of plume generation and molecular production via crater-forming impacts. Due to chemical weathering of ultramafic crust and impact ejecta in a CO_2-rich atmosphere, the Hadean crust was likely to harness ~10^{19}-10^{21} moles of carbonates to sustain a global carbon cycle (Sleep & Zahnle, 2001), equivalent to 0.1-10 wt.% of carbonate minerals on Hadean surface. Abiotic fixation of N_2 gas in atmosphere would have produced ~10^9-10^{12} moles of N salts per year ((Brandes et al., 1998) and references therein), leaving ~0.001-1 wt.% of N salts on the crustal surface over an 100 Myr interval in Hadean Earth. Frequent large asteroid impacts on Hadean Earth would have evaporated majority of ocean water in the global scale (Maher & Stevenson, 1988; Sleep et al., 1989), and thus we assumed a shallow marine of 10 m in the following calculation.

The range of carbonaceous projectiles considered for the surface impacts in early Earth scenario is 10^{16}-10^{20} kg (diameter size of ~20 to 400 km). This range of projectiles would be massive enough to strike on the planetary surface without a significant

deacceleration due to presence of an atmosphere and a shallow marine. With a 10 wt.% carbonates and 1 wt.% N salts on the crustal surface, this range of projectiles would produce 10^{12}-10^{15} moles of CN^- in a single crater-forming impact event, or an upper bound of 10^{10}-10^{14} kg HCN assuming a 1% atomization efficiency and 50% neutralization efficiency. A crustal surface with 0.1 wt.% carbonates and 0.001 wt.% N salts would produce a lower bound, > 10^9 kg HCN produced per single impact event. Considered the amount of impact energy delivered by a single carbonaceous projectile of 10^{16}-10^{20} kg, the production efficiency of HCN is on the order of 10^{11}-10^{13} molecules per joule of impact energy, 4-6 orders of magnitude lower than the production efficiency of HCN via atmospheric shock in a methane-rich atmosphere (McKay & Borucki, 1997; Scattergood et al., 1989).

We also computed the total HCN production yields from surface impacts by integrating the frequency of impact events during a 100 Myr interval in early Earth following the mass-frequency curve of the impactors established in . With a shallow marine about 10 m in depth, the total HCN production via surface impacts was estimated to be ~10^{15} kg. If the marine is shallower than 10 m, the HCN production yields would be inflated by at most 60% due to generation of additional acoustic pressure to excavate crustal materials into the vapor plume (Chemin et al., 2022; Zhu et al., 2001). If the ocean layer was >500 m, even a 10^{20} kg projectile could not penetrate the thick water column and make sufficient ejecta on crustal surface for HCN production; however, such massive projectile would evaporate the entire ocean layer of present size (Citron & Stewart, 2022) and thus creating shallow marine environments for subsequent crater-forming impact events. We therefore conclude a total HCN production yield of 10^{15}-

10^{16} kg during a 100 Myr interval in early Earth. This value is 2-4 orders higher than the estimated total of organic matter delivered by exogenous material (~10^{13} kg carbon) (Anders, 1989).

Despite such a high yield, impact-generated HCN would have undergone quick decomposition via photolysis, thermal decomposition, oxidative dissociation by metal catalysts on the ocean floor minerals, and other prebiotic processes in early Earth (Ferus et al., 2020). These impact-produced organic precursors would be more stable if they could polymerize and chemically evolve in the planetary environments. Here we assume the impact-produced HCN will be circulated in the atmosphere on a global scale, which eventually fall out and dissolve in the ocean with a residence time of 1 year (Q. Li et al., 2000). We explored whether the molarity of HCN produced per single impact event in the ocean would surpass the threshold concentration (0.01 M) to achieve HCN polymerization instead of hydrolysis (Sanchez et al., 1967). Our model suggests that HCN polymerization is only feasible if the ocean depth is less than 10 m and impactor size > 10^{18} kg, suggesting the low probability of achieving HCN polymerization on early Earth using impact-produced HCN alone.

3.4.3.2.Mars

Following a similar approach, we estimated the yields of HCN by crater-forming impacts on Noachian Mars. The Noachian Mars, which spanned from 4.1-3.7 Ga, was likely to experience an elevated rates of cratering on a presumably warm and wet surface that was covered by hydrous weathering products such as phyllosilicates and carbonate deposits from a CO_2-H_2O atmosphere (Carr & Head, 2010; Ehlmann et al.,

2008). Based on the outcrops of hydrous-altered mineral phases on ancient Noachian surface, early Mars was likely to sustain water for an extended period of time (Ramirez & Craddock, 2018) and thus we used a marine depth of 10 m for the following calculation. The Martian surface was estimated to have 2-5 wt.% carbonates (calcite, siderite, magnesite) (Bandfield et al., 2003) and ~0.03 wt.% N salt (Stern et al., 2015), as indicated by the chemical measurements of Martian soil from the Mars Science Laboratory (MSL) rover. Our estimates of HCN production is based on the assumption of a stable early Martian atmosphere that is cold enough to sustain surface water and similar to present Mars, about 10^{16} kg and is dominated by 95 wt. % CO_2 and 0.28 wt.% N_2 (Mahaffy et al., 2013; Segura et al., 2012).

Mars has a lower gravity and smaller cross sections compared to Earth, attracting a range of less massive carbonaceous projectiles (10^{14}-10^{17} kg) travelling at 10 km/s (Strom et al., 2005). This range of projectiles would produce 10^{8}-10^{11} kg HCN per impact event. Using the flux of Martian impactors based on the mass-frequency curve of terrestrial impactors (Anbar et al., 2001) and assuming an Earth/Mars impact flux ratio of 10 (Bottke et al., 2010), the total HCN production yield via crater-forming impacts on Mars is ~10^{12} kg during an 100 Ma interval in Noachian Mars. Assuming impact-derived HCN will be circulated in the atmosphere on a global scale and dissolve in the ocean with a residence time of 1 year, the concentration of HCN in ocean would exceed the polymerization threshold (i.e., 0.01 M) when the ocean depth is less than 0.02 m and impactor size > 10^{17} kg, highlighting a small likelihood of polymerizing HCN in early Mars.

3.4.3.3.Ceres

Ceres is an ice-rich body that lacks any significant atmosphere or surface ocean. The surface of Ceres has been observed to have NH_4-phyllosilicates and Mg,Ca-carbonates, which are estimated to be ~10% for both (Ammannito et al., 2016; Carrozzo et al., 2018; De Sanctis & Ammannito, 2021). Ceres' craters are most typically in the 1 to 100 km diameter range, and likely created by carbonaceous impactors of 10^7-10^{15} kg (Hiesinger et al., 2016). This range of projectile masses can produce 10^5-10^{11} kg HCN per crater-forming impact based on our model. Ceres is also observed to have a subsurface icy shell that is modelled be at < 5 km depth from its surface (Bland et al., 2016; Sizemore et al., 2019). Projectiles > 10^{12} kg can penetrate the uppermost 5 km of crustal materials, melt the subsurface icy layer, and sustain a liquid reservoir via long-term impact-induced heating. Without the presence of a significant atmosphere to circulate impact produced HCN on a global scale, this surface liquid reservoir could contain up to 0.9 M of HCN, sufficiently concentrated to initiate HCN polymerization and facilitate molecular evolution inside the transient crater. Therefore, crater-forming impacts could be a critical mechanism to enable the abiotic formation of complex molecules, following the path of HCN polymerization on planetary surface.

3.5. Conclusion

Experimental simulations of the vapor plume generated via hypervelocity impacts have shown that the recombination of atomized C and N-bearing minerals can produce molecular precursors of biomolecules. Specifically, CN^- and CNO^- ions were observed in the recombination vapor generated via C_{60}^+ projectile bombardments and laser

ablation experiments targeting carbonate, nitrate, and ammonium starting materials. The production yields of new CN-bearing molecules varies according to the volume and size of the impact plume generated in varying planetary environments. Crater-forming impacts could have played a major role in increasing the organic inventory of a young planet and facilitated the chemical evolution of early planetary environments. In the case of Hadean Earth and Noachian Mars, few projectiles would be able to impact the surface and generate sufficient HCN to enable molecular polymerization assuming a global coverage of atmosphere and shallow marines. In contrast, crater-forming impacts on Ceres could have created the conditions and organic inventory necessary to enable HCN polymerization at a more regular interval, facilitating further chemical evolution to more complex organic molecules. Future studies could use the model provided here to explore the contributions of impact induced HCN and the possibility of HCN polymerization in various planetary environments.

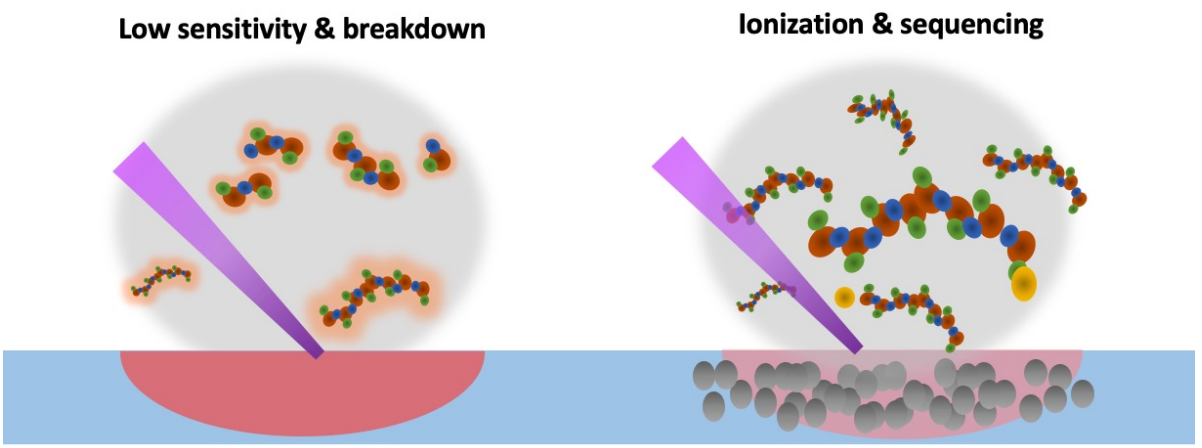
3.6. Acknowledgement

The study at NASA/GSFC was supported by the NASA Grant (80NSSC18K1612) and the Fundamental Laboratory Research (FLaRe) work package, and the work at Texas A&M University was supported by NSF grant CHE-1308312. We acknowledge the innovative and seminal work by Prof. G. G. Managadze (dec.) as inspiration for this study.

Detection of Short Peptides as Putative Biosignatures of Psychrophiles via Laser Desorption Mass Spectrometry

Ziqin Ni[1,2], Ricardo Arevalo Jr.[2], Anais Bardyn[2], Lori Willhite[2], Soumya Ray[2], Adrian Southard[3], Ryan Danell[4], Jacob Graham[5], Xiang Li[5], Luoth Chou[5,6], Christelle Briois[7], Laurent Thirkell[7], Alexander Makarov[8], William Brinckerhoff[5], Jennifer Eigenbrode[5], Brook Nunn[9], and Karen Junge[9]

1 Corresponding Author
2 University of Maryland, College Park, MD USA
3 CRESST II, College Park, MD USA
4 Danell Consulting, Winterville, NC USA
5 NASA Goddard Space Flight Center, Greenbelt, MD USA
6 Georgetown University, Washington, DC USA
7 Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Orléans, FR
8 Thermo Fisher Scientific, Bremen, DE
9 University of Washington, Seattle, WA USA



Chapter 4: Detection of Short Peptides as Putative Biosignatures of Psychrophiles via Laser Desorption Mass Spectrometry

Quick overview of the chapter:

Conventional method to ionize nonvolatile organic molecules via an ultraviolet (UV) laser is to mix analytes with UV-absorbing proton-rich organic matrices, a method known as matrix-assisted laser desorption ionization mass spectrometry (MALDI). However, introducing organic matrices to space exploration would introduce potential risks of forward contamination. Without these UV-absorbing matrices, it would be challenging to detect and sequence peptides/proteins, such as the select 3-mer and 4-mer oligopeptides of psychrophiles as putative biosignatures in ocean worlds. What are the alternative strategies to detect and sequence peptides during the actual mission investigation without significant modification of the instrument design and comply to the planetary protection regulation? What would these putative peptide signatures look like in the mass spectrum and how should we identify them? In this chapter, we explored the capabilities of detecting and identifying peptides via LDMS with the addition of silicon nanoparticles. The outcome of this research enables detection and sequencing of peptides via single-shot LDMS analysis, and thus providing a solution to characterize putative biosignatures via in situ operation of laser enabling mass spectrometer.

In this chaper, I processed the dataset, interpreted the results, made the figures, and wrote the paper. This chapter has been published.

Ni, Z., Arevalo Jr, R., Bardyn, A., Willhite, L., Ray, S., Southard, A., et al. (2023). Detection of Short Peptides as Putative Biosignatures of Psychrophiles via Laser Desorption Mass Spectrometry. *Astrobiology*, *23*(6), 657-669.

4. Abstract

Studies of psychrophilic life on Earth provide chemical clues as to how extraterrestrial life could maintain viability in cryogenic environments. If living systems in ocean worlds (e.g., Enceladus) share a similar set of 3-mer and 4-mer peptides to the psychrophile (*Colwellia psychrerythraea*) on Earth, spaceflight technologies and analytical methods need to be developed to detect and sequence these putative biosignatures. We demonstrate that laser desorption mass spectrometry, as implemented by the CORALS spaceflight prototype instrument, enables the detection of protonated peptides, their dimers and metal adducts. The addition of silicon nanoparticles promotes the ionization efficiency, improves mass resolving power and mass accuracies via reduction of metastable decay, and facilitates peptide *de novo* sequencing. The CORALS instrument, which integrates a pulsed UV laser source and an Orbitrap^{TM} mass analyzer capable of ultrahigh mass resolving powers and mass accuracies, represents an emerging technology for planetary exploration and a pathfinder for advanced technique development for astrobiological objectives.

4.1. Introduction

The search for signs of life elsewhere in the Solar System is a fundamental science goal receiving increasing attention in recent decades. Saturn's moon Enceladus and Jupiter's satellite Europa have been elevated to prime targets for astrobiology and life detection missions in the following decades, owing to the presence of large subsurface water reservoirs (Beuthe et al., 2016; Čadek et al., 2016; Carr et al., 1998; Glein et al., 2015; Hemingway & Mittal, 2019; Iess et al., 2014; Khurana et al., 1998; McKinnon, 2015;

Postberg et al., 2011; Thomas et al., 2016), active energy and material exchange across surface and interior (Greenberg et al., 1998; Patterson et al., 2018; Postberg et al., 2009; Reynolds et al., 1987; Smith et al., 1982; Squyres et al., 1983; Waite et al., 2006, 2017), and sustained geological activity and bioavailable nutrients sufficient to power biological activities over a long period of time (Choblet et al., 2017; Greenberg et al., 1998; Hsu et al., 2015; Nimmo et al., 2018; Reynolds et al., 1987; Squyres et al., 1983). Astrobiological objectives pertaining to these top priority ocean worlds are not solely limited to the detection of simple organic molecules that are generic to planetary targets across the solar system. It is strategically also important to develop capabilities to detect and recognize chemical signatures of living systems thriving in cryogenic environments akin to these potentially habitable ocean worlds (Cable et al., 2021; Hendrix et al., 2018; McKay et al., 2008, 2014).

In a recent experimental study, cultures of *Colwellia psychrerythraea* strain 34H (*Cp*34H) were exposed to progressively colder temperatures, higher salinities, and more depleted nutrient levels. Distinguished novel metabolic strategies responsible for maintaining viability in response to each experimental variable were identified via proteomics analysis (Mudge et al., 2021). A suite of select short peptides comprising 3 or 4 amino acids (referred to as 3- and 4-mer peptides, respectively) were found to be statistically enriched in the populations that survived the harshest environmental conditions. If such enrichments are ubiquitously found in terrestrial psychrophilic organisms, then such specific 3-mers and 4-mers may represent intrinsic protective functionality. Thus, identified short peptide populations could represent putative biosignatures of psychrophiles in future efforts to search for chemical evidence of

living systems in ocean worlds (Mudge et al., 2021). However, such promising chemical indicators must be susceptible to analytical techniques with a path to spaceflight implementation in order to serve as practical biosignatures.

Miniaturized mass spectrometers have been developed for planetary exploration dating as far back as the Lunar Atmospheric Composition Experiment (LACE) as a part of Apollo 17 mission in 1972 (Hoffman, 1975; Arevalo et al., 2020 and references therein). In the context of astrobiology life detection missions, spaceflight mass spectrometers have been important in the detection and identification of volatile organic and trace gases derived from the Martian subsurface (Eigenbrode et al., 2018; Freissinet et al., 2015; House et al., 2022; Millan et al., 2022; Ming et al., 2014), the Enceladus plume (Teolis et al., 2010; Waite et al., 2006), and Titan's atmosphere (Niemann et al., 2005). Recently, the Dragonfly Mass Spectrometer (DraMS) instrument, a dual source linear ion trap mass spectrometer that supports both gas-phase analyses and laser desorption measurements, is under development to analyze the nonvolatile organics on Titan's surface (Grubisic et al., 2021). Thus far, no deployed mass spectrometer has yet analyzed a wide range of complex nonvolatile organic molecules and/or characterized their molecular structures on a flight mission. Peptides/proteins generally represent higher-fidelity indicators of living systems because their function-specific sequences and energy-optimized structures are highly unlikely to be produced abiotically (Marshall et al., 2021). However, it is challenging to ionize macromolecular material via conventional spaceflight sampling and ionization methods (e.g., electron ionization) and fully identify molecular compounds without the characterization of their structures.

Laser desorption mass spectrometry (LDMS) uses a focused laser beam to desorb and ionize analytes from the substrate and determines their chemical composition based on the mass to charge ratio (m/z). LDMS can desorb and ionize nonvolatile organic compounds of higher masses without incurring molecular fragmentations, allowing the determination of chemical constituents of short peptides based on their molecular masses. However, in some operational settings (such as a high laser fluence or elevated background pressure), LDMS can induce characteristic cleavages of the peptide backbones during the laser irradiation in the ionization step (Asakawa et al., 2022; Köcher et al., 2005), molecular dissociation due to collision with background neutrals (Ackloo & Loboda, 2005), and breakdown of metastable ions in the ion acceleration region prior to entering the mass analyzer (Spengler et al., 1991, 1992). These diagnostic fragments are informative to the peptide sequence. It would be plausible to enable simultaneous detection and sequencing of short 3-mer and 4-mer peptides in a single measurement. However, it is essential to ensure optimal ionization efficiency without incurring excessive fragmentation of the molecular ions.

To minimize molecular fragmentation during laser microprocessing, laboratories have embraced the matrix-assisted laser desorption/ionization (MALDI) technique, in which a photo-absorptive matrix compound is physically admixed with the analyte to increase the desorption and ionization efficiency (Karas et al., 1987; Karas & Krüger, 2003; Karas & Hillenkamp, 1988; Zenobi & Knochenmuss, 1998). The most common MADLI matrices are solid organic acids such as α-cyano-4-hydroxycinnamic acid (α-CHCA) that are optimal for ultraviolet (UV) LDI. Their application enables reproducible detection and characterization of protein, peptides, oligonucleotides well

over 10,000 Da at high signal to noise ratio (S/N) (Tanaka et al., 1988). Most studies on peptide sequencing via MALDI analysis are limited to solid organic matrices due to their optimal performance and ease of laboratory technique (Castoro et al., 1995; Chaurand et al., 1999; Hettick et al., 2001; Keough et al., 1999; Köcher et al., 2005; Spengler et al., 1991, 1992; Yergey et al., 2002). However, the breakdown of solid organic matrices, induced by laser irradiation, generates extensive fragments in the mass range m/z < 500, overlapping with valuable molecular species such as amino acids and peptide sequence fragments (Leopold et al., 2018; Schmitt-Kopplin et al., 2010). Further, the use of an organic matrix in spaceflight LDMS: 1) increases the complexity of the experimental operation, such as depositing the matrix on top of the sample; 2) creates convoluted spectra by adding a wide range of low mass peaks associated with the matrix to the mass spectrum; 3) ambiguates data detection by introducing reactants to the plasma plume that may induce molecular recombination; and, 4) introduces the risk of contaminating a planetary target of astrobiological interest.

A more feasible class of MALDI matrices for spaceflight application would be inorganic material. A variety of inorganic matrices have been explored to detect biomolecules since the inception of MALDI techniques, including ultrafine metal power (Tanaka et al., 1988), graphite particles in glycerol solution (Dale et al., 1996; Sunner et al., 1995), thin layers of activated carbon particles on an aluminum plate (Han & Sunner, 2000), and porous silicon and silica gel (Alimpiev et al., 2001; Wei et al., 1999; Zhang et al., 2001). Silicon nanoparticles is one example of inorganic matrices that have been recently shown to promote the ionization of large organic molecules at lower fluences, improve sample homogeneity, reduces shot-to-shot

variability, and suppress the salt inferences from matrices than neat samples (Abdelhamid, 2018; Cohen & Gusev, 2002; Wen et al., 2007). Silicon nanoparticles have been applied to detect various types of small molecules including 3-mer and 4-mer peptides using commercial instruments (Wen et al., 2007), but they have not been used to sequence peptides. Silicon nanoparticles also have not been validated with spaceflight technologies yet. The application of inorganic matrices like silicon nanoparticles in laser desorption mass spectrometry may facilitate the detection of a wider range of molecular biosignatures in future life detection missions, such as those complex organic molecules called out in the Europa Lander Science Definition Team Report (Hand, 2017) or the refractory organic aerosols floating in Titan's atmosphere and on Titan's surface (Cable et al., 2012).

This study aims to investigate the feasibility of detection and *de novo* sequencing of selected 3-mer and 4-mer peptides expressed in the most resilient strains of Cp34H, those exposed to the harshest environmental conditions (Mudge et al., 2021), via the advanced prototype of CORALS (Characterization of Ocean Residues and Life Signatures), a laser desorption mass spectrometry (LDMS) instrument developed for the exploration of Europa and other ocean worlds (Arevalo Jr et al., 2018; Briois et al., 2016; Willhite et al., 2021). The CORALS instrument is composed of a solid-state laser system that emits 266 nm radiation (up to 450 µJ) (Fahey et al., 2020) and a high performance Orbitrap analyzer that can measure the m/z of charged analytes to the 4^{th} decimal places. The precise and accurate measurement of m/z can efficiently separate ions of similar m/z and allows the determination of chemical constituents of unknown molecules even if they are in a sample mixture.

Our investigation began by exploring the capacity of the CORALS instrument to detect and sequence select 3-mer and 4-mer peptides in the simplest LDMS experiments, without application of any MALDI matrices. Such "neat" samples rely primarily on the UV absorption characteristics of the sample (and also the sample plate for thin residues) to facilitate ionization of the analyte and require fewer sample handling considerations than those involving physical admixture of solid organic MALDI matrices. Early work on neat samples with the CORALS prototype demonstrated the detection of amino acids down to fmol/mm^2 concentrations and measurements of trace elements down to ppmw levels (Arevalo Jr et al., 2018; Briois et al., 2016; Willhite et al., 2021), but have not examined the complex refractory organic molecules. We then explored the use of silicon nanoparticles, inorganic alternatives to conventional organic compounds routinely used for matrix assisted laser desorption/ionization (MALDI) techniques, to enhance ionization of the refractory organic molecules. These experiments will advance our fundamental understanding of laser desorption processes and fragmentation of relatively large organic molecules via a deep UV high power laser. The peak distributions and fragmentation patterns obtained in each LDMS spectrum indicate plausible biosignature fingerprints expected in the search for cryogenic life in planetary environments.

4.2. Materials and Methods

4.2.1. Peptide selection and preparation

A subset of psychrophilic peptides was selected to demonstrate the detection and identification of molecular ions and diagnostic fragments at higher laser energy with

the CORALS instrument. These four different short peptides are Ala-Ala-Gly (AAG) (monoisotopic mass: 217.095), Ser-His-Asp (SHD) (monoisotopic mass: 357.117), Gly-Asp-Ala-Glu (GDAE) (monoisotopic mass: 390.128), Ser-Lys-Ser-Arg (SKSR) (monoisotopic mass: 476.260) listed according to their increasing molecular weight. These four peptides were selected because they constitute amino acids of different basicity. Aspartic acid and glutamic acid have acidic side chains while arginine, histidine and lysine have basic side chains. The selected peptides have a mixture of amino acids of different weights and sizes. High molecular weight amino acids like histidine is rarely found in abiotic systems, while low molecular weight amino acids like glycine and alanine have been found in high abundance in meteorites and are thus more likely to be used as starting materials to construct functional polymers (Wen et al., 2007). These peptides were synthesized from Biomatik with purity > 95%. Each peptide was dissolved in 0.1 wt.% formic acid solution (Fisher Chemical™, ≥ 99.0%) to form ~5 wt.% peptide stock solutions, and they were used without further purification.

Matrix-free peptide samples: About 200 uL of aqueous solution was directly drop-cast on the aluminum (Al) alloy metal plate (Goodfellow AL7075 foil) to form a droplet of ~ 60 mm^2 in size and achieve surface analyte concentration ~5e5 $pmol/mm^2$. The Al alloy plate was pretreated with acetone and isopropyl alcohol. The droplet was allowed to dry on the hotplate at 75 °C until a thin residue of peptide was formed on the Al alloy plate. The plate was then transferred and taped on the stainless-steel sample target, ready for sample introduction to CORALS prototype instrument at NASA GSFC. Though such sample preparation method does not create a homogenous surface for

quantitative measurements, it is representative of realistic sample handling and preparation during *in situ* investigation of icy worlds.

Matrix-assisted peptide samples: The silicon nanoparticles were purchased from Sigma Aldrich (<100 nm particle size, ≥98% trace metals basis) and used without further purification. Acetone (Sigma-Aldrich, ≥ 99.5%) was used to mix with a few mg of the silicon nanoparticles to form a liquid slurry. 2uL of the slurry was deposited onto the sample plate and dried, forming a very thin layer of the nano powder on the surface. The peptide solution then was deposited on top of the nanoparticle layer and dried. Because of the high volatility of acetone in the room temperature, acetone would be completely evaporated before the sample was introduced to the instrument. No signature of acetone has been observed in the mass spectra.

4.2.2. Instrument setup

The sample was analyzed by the CORALS prototype instrument at NASA GSFC (**Figure 4.1**). The CORALS prototype is composed of a breadboard of the CORALS laser, integrated to an Orbitrap analyzer and a custom-built ion inlet assembly through a Kimball Physics vacuum chamber. The chamber maintains its low-pressure condition at typically $\leq 10^{-6}$-10^{-7} Pa and the pressure is monitored by a high-vacuum cold-cathode vacuum gauge. Sample plate is introduced into the low-pressure simulation chamber through a load lock via a linear-rotary actuator. The laser beam enters the simulation chamber through a deep UV viewport that offers high transmission of 266 nm radiation and irradiates the sample at a 45° angle. Ions generated from the sample plate are propelled by a high voltage supplied to the sample plate, collimated by the ion optics,

accelerated, and injected into the mass analyzer. The image current induced by ions oscillating inside the orbitrap was detected, amplified, digitized, and exported for offline data processing. All instrument control, voltage and timing tuning, and data acquisition is accomplished with custom LabVIEW based software (RITS, Danell Consulting, Inc.) (Danell, 2010).

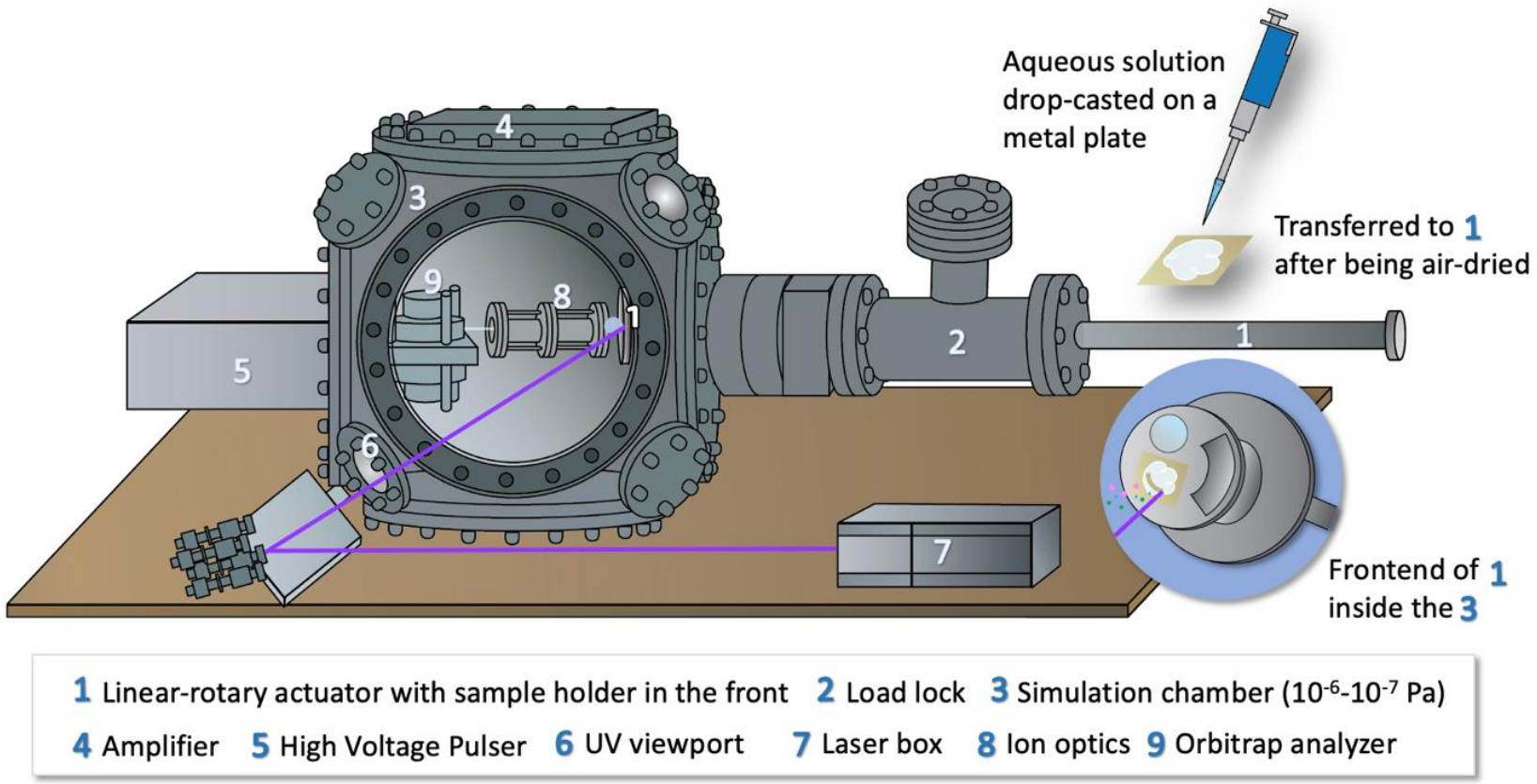


Figure 4.1. Sample preparation and instrument setup at NASA Goddard Spaceflight center.

4.2.3. Data collection and processing

Each transient spectrum (200 ms in length) was collected at a sampling rate of 5 MHz, and then processed to generate a frequency spectrum via Fast Fourier Transform (FFT) in MATLAB. Three times zero-filling to $3X + N = 2^Y$ (where X is the original transient length, N the remaining amount needed to achieve the nearest power of 2, and Y the final zero-filled transient length) was applied prior to FFT to improve peak shape and mass accuracy. No apodization has been applied to the transient prior to FFT analysis and the reasons have been discussed in Appendix C. The frequency spectrum was calibrated into a mass spectrum using an internal standard, typically Cs^+ or Al^+. Identified peaks were fitted with a gaussian distribution using Origin (OriginLab Corporation). The derived peak height, peak width, and center mass are used to calculate signal to noise ratio (S/N), mass resolution, and mass accuracy, respectively. Peak identification and characterization are reported in Appendix C.

4.3. Results

4.3.1. LDMS analysis of neat peptides

The peptides Ser-His-Asp (SHD) and Gly-Asp-Ala-Glu (GDAE) were 3-mer and 4-mer peptides chosen from peptides expressed in the most resilient strains of Cp34H that survived under the harshest environmental conditions (Mudge et al., 2021). Histidine (H) is rarely found in abiotic systems and has a basic residue that has shown to improve protonation efficiency in the gas phase. Aspartic acid (D) and glutamic acid (E) have an acidic residue. SHD and GDAE were each analyzed via LDMS with the CORALS prototype (see Methods) as neat samples deposited onto a flat Al alloy plate.

In the spectra collected, Cs^+ derived from a collocated finely ground disc of CsI (7.49 mm diameter) produced by Almaz Optics, Inc. was used as an internal mass calibrant. The 3-mer SHD was identified by its protonated molecular ion (i.e., $[M+H]^+$) and dimer (i.e., $[2M+H]^+$), both of which were observed below the limits of quantitation (signal-to-noise ratios (S/N) < 10) despite exploring a wide range of fluences between 0.06 and 6 J/cm^2 (**Figure 4.2a**). The most abundant organic peaks were observed up to three orders of magnitude lower than the intensities of Cs^+ (from the CsI disc), Al^+ (from the sample plate), and Na^+ and K^+ (from the sample solution), implying limited ionization efficiency. In contrast, GDAE was identified solely by its sodiated molecular ion (i.e., $[M+Na]^+$), suggesting a distinct ionization pathway (**Figure 4.2b**). The sodiated molecular ion was observed below the limits of quantitation, at about two orders of magnitude lower than the intensities of Al^+, Na^+, and K^+.

Due to the ultrahigh mass resolution and ppm-level mass accuracy provided by the CORALS analyzer, the detection of the molecular ions of SHD and GDAE enables the unambiguous determination of their respective molecular stoichiometries. However, specific diagnostic fragments are essential to reveal the entire sequence of amino acid monomers, a process known as *de novo* sequencing. Peptides are lineages of proteinogenic amino acids, which share a fundamental structure that consists of an alpha carbon ($-C_n$) at the center, bonded to an amino group ($-NH_2$), a carboxylic group (-COOH), a residue ($-R_n$) and a hydrogen atom (**Figure 4.3**). Conventional fragmentation mechanisms induce backbone cleavages at N-C, C-C, and C-N bonds in between two consecutive amino acids. Accepted nomenclature refers to these backbone cleavages containing the N terminal of precursor peptide ions as *a*-, *b*-, and *c*- ions,

respectively. The complementary sets of fragments containing the C terminal of the precursor peptide ions are called *x*-, *y*-, and *z*-, ions. A subscript *n* indicates the number of residues in the fragments, implying the location of the bond dissociation in the precursor peptide ions. This study defines *a*-/*x*-, *b*-/*y*-, *c*-/*z*- sequence ions as the diagnostic peptide fragments needed to achieve *de novo* sequencing. Generation of diagnostic fragments at all possible *n* locations is considered to achieve a full sequence coverage.

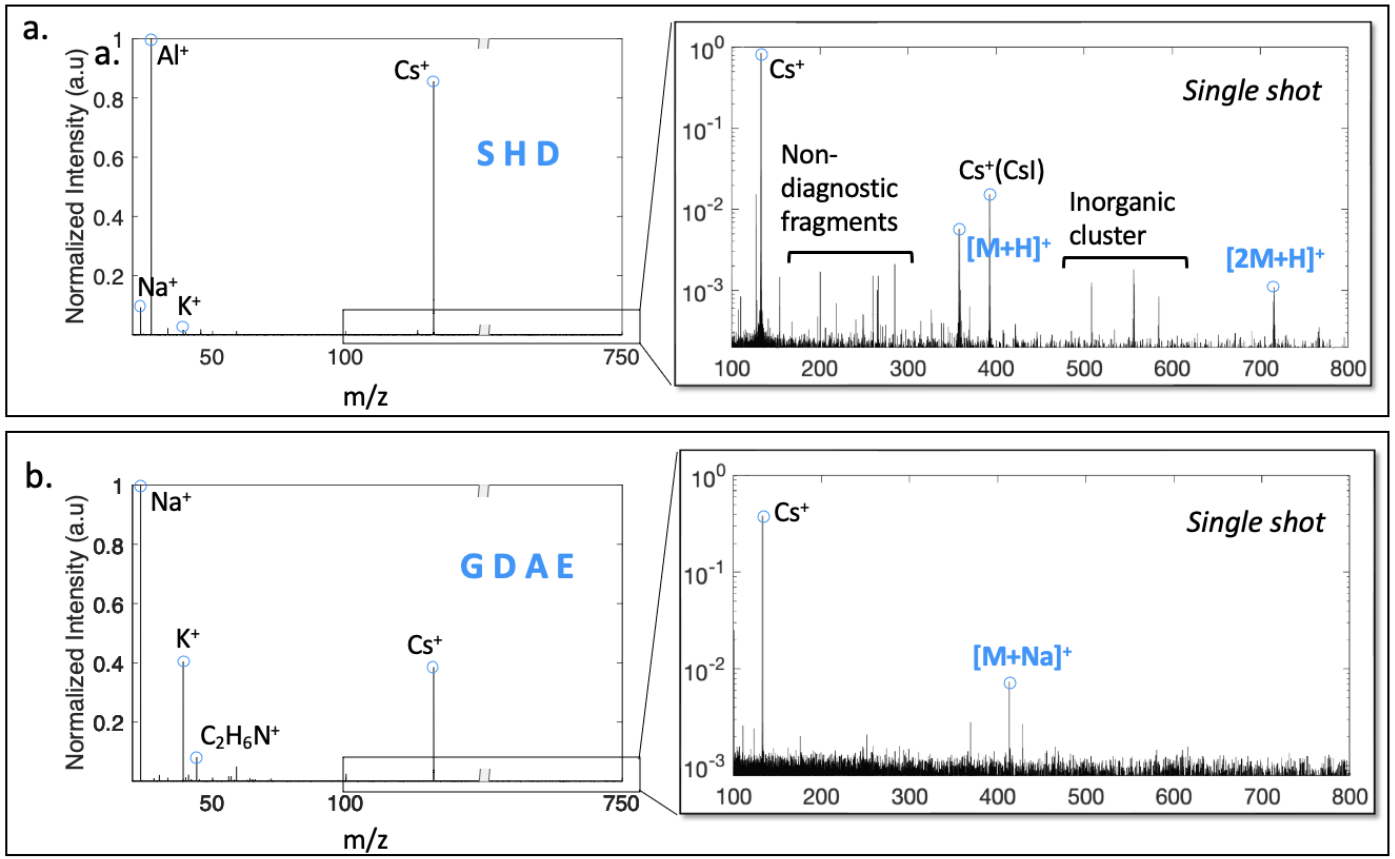


Figure 4.2. Single shot LDMS analysis of neat (a) SHD and (b) GDAE on flat Al plate shows detection of molecular ions at S/N < 100 but lack of diagnostic fragments for peptide *de novo* sequencing. Spectral intensities are normalized to the maximum peak intensity observed in each scan. The characteristics of all identified peaks are listed in Table C1. The raw transient data will be accessible in Appendix C.

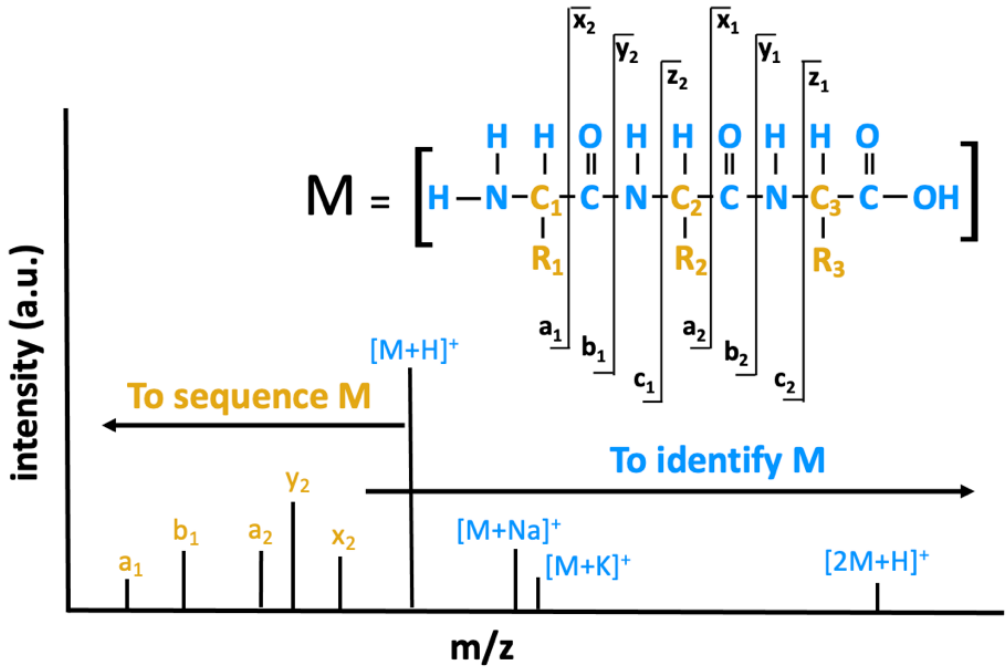


Figure 4.3. A schematic diagram that shows the nomenclature of fragments in a 3-mer peptide to enable *de novo* sequencing. The alpha carbon (-C_n) and the residues (-R_n) are highlighted in the peptide structure. A list of general types of molecular adducts and diagnostic fragments are also illustrated to demonstrate the identification and sequencing of a peptide in a single mass spectrum.

Because the molecular ions of both SHD and GDAE were observed at low intensities (S/N < 100), no diagnostic fragments above the limit of detection (LOD, S/N > 3) were recorded. Although a variety of peaks were seen in the spectrum of SHD, none of them match to the exact masses of expected sequence ions. The positive mass defects of peaks in the m/z 150-300 range suggest the ionic species are organics composed of multiple hydrogen and nitrogen atoms. Consequently, matrix-free LDMS analysis of SHD and GDAE on an Al plate successfully demonstrates the ionization and detection of neat 3-mer and 4-mer peptides, but low sensitivities inhibit peptide *de novo* sequencing.

4.3.2. LDMS analysis of single peptides with silicon nanoparticles

Silicon nanoparticles serve to enhance ionization of large refractory organic molecules without excessive fragmentation. Previous studies have shown the desorption of large organic molecules at a lower laser fluences with higher S/N, greater reproducibility, and with less interference from salt contamination when analyzing compounds doped with silicon nanoparticles (Abdelhamid, 2018; Cohen & Gusev, 2002; Wen et al., 2007). However, to the best of our knowledge, silicon nanoparticle matrix has not yet been used to facilitate the detection and sequencing of short peptides via LDMS, nor been introduced to spaceflight technologies.

To facilitate a semi-quantitative comparison with the neat measurements described above, separate aliquots of SHD and GDAE were each doped with silicon nanoparticles and analyzed via LDMS with the CORALS prototype (**Figure 4.4**). Although the highest intensity peaks in the collected spectra were sourced from the nanoparticles, the protonated molecular ions of both SHD and GDAE were observed at intensities equal to or higher than alkali metals and Al. At higher masses, protonated dimers and molecular ions with metal adducts derived from the analyte solution (i.e., $[M+Na]^+$ and $[M+K]^+$), sample plate (i.e., $[M+Al]^+$), and nanoparticles (i.e., $[M+Si]^+$) were recorded. Dehydrated peaks derived from protonated SHD and GDAE were also observed. The exact masses of multiple molecular ions moieties provided corroborative identification of the molecular formula of each peptide.

Both peptide datasets revealed various types of diagnostic fragments. Although characteristic b-/y- sequence ions were observed for GDAE, access to only b_2/y_2 and

b_3/y_1 provided insufficient information to sequence the entire peptide (i.e., unable to distinguish GDAE or DGAE). In contrast, the spectra for SHD offered a full breadth of fragments, including c_1 and c_2 ions, successfully enabling *de novo* sequencing of SHD. Generation of various types of sequence ions at the same bond location, such as c_1 and $(x,y,z)_2$ ions generated from S-HD sequence breakage, provide redundancy to confirm the peptide sequence. Thus, these experiments corroborate the benefits of adding silicon nanoparticles to enhance the ionization efficiency of nonvolatile organic compounds, in this case short peptides, leading to improved sequencing capabilities.

4.3.3. LDMS of peptide mixtures with and without silicon nanoparticles

To increase the fidelity of our findings, we also explored the detectability of 3-mer and 4-mer peptides in multicomponent samples with and without silicon nanoparticles. We focused on a mixture of two peptides, Ala-Ala-Gly (AAG) and Ser-Lys-Ser-Arg (SKSR) that were statistically linked with the survivability of Cp34H under harsh conditions (Mudge et al., 2021). Because the concentrations of both AAG and SKSR were maintained at the same level as SHD and GDAE in the single peptide experiments described above (see Methods), the two-component mixture contained twice the total concentration of peptide.

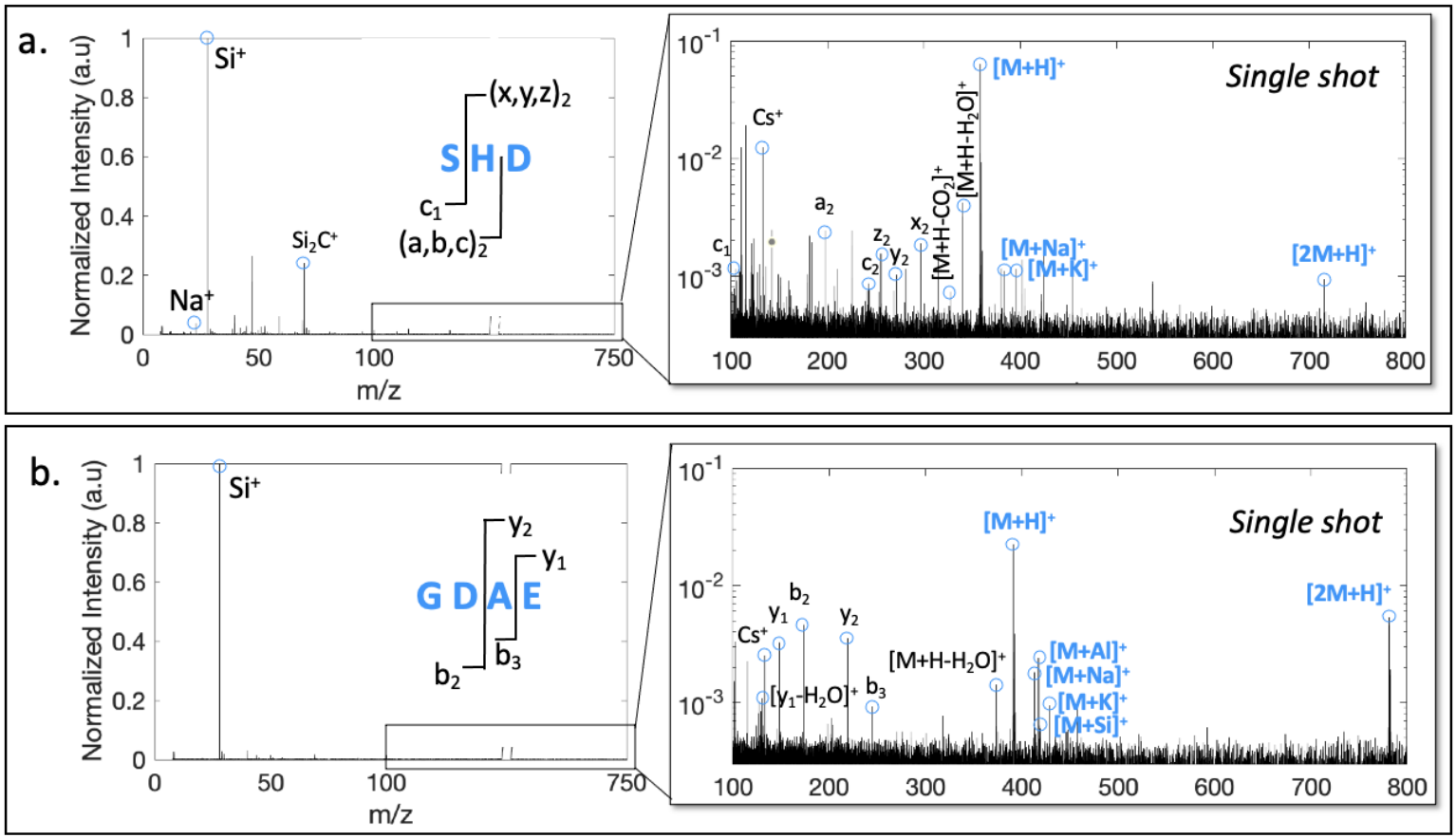


Figure 4.4. Single shot LDMS analysis of neat (a) SHD and (b) GDAE on a flat Al alloy plate with silicon nanoparticles demonstrates detection of various forms of molecular ion moieties and diverse diagnostic fragments that support peptide *de novo* sequencing. Spectral intensities are normalized to the maximum peak intensity observed in each scan. The characteristics of all identified peaks are listed in Table C1. The raw transient data will be accessible in Appendix C.

In analysis of this peptide mixture in absence of silicon nanoparticles (**Figure 4.5a**), SKSR was identified with a diverse assortment of molecular ion moieties (i.e., $[M+H]^+$, $[M+Na]^+$, and $[M+K]^+$) and a full sequence coverage. AAG was seen with various forms of molecular ions (i.e., $[M+H]^+$, $[M+Al]^+$, $[M+K]^+$, and $[2M+H]^+$) and diagnostic fragments at a full peptide sequence coverage. With the addition of silicon nanoparticles (**Figure 4.5b**), AAG was seen with a similar set of molecular ion adducts and fragmentation pattern to neat sample analysis, but SKSR was observed with a reduced diversity of molecular ion adducts and peptide fragments (i.e., y_1, z_1 are confined to proximity of arginine). Thus, the effective role of silicon nanoparticles appears to be dependent on the specific analyte(s) of interest, as discussed further below. Nonetheless, the detection of multiple molecular ion adducts and critical peptide fragments suggest that the CORALS instrument is able to discriminate, identify, and sequence SKSR and AAG simultaneously in a mixture, implying its capability to detect and characterize multiple peptides in an unknown sample from an ocean world.

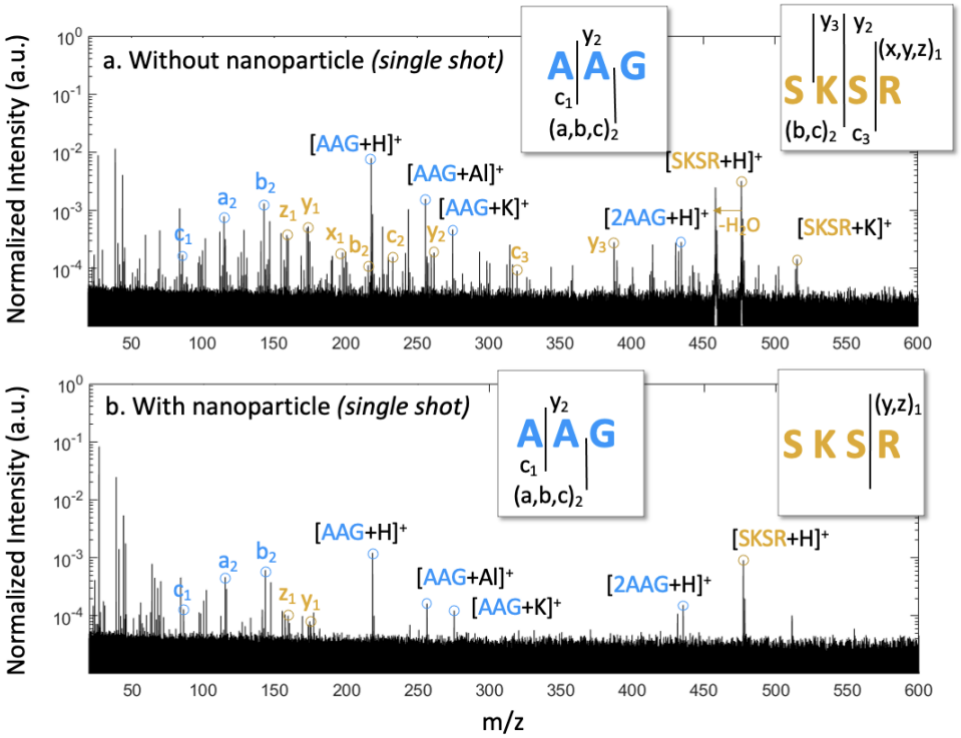


Figure 4.5. Single shot LDMS analysis of peptide mixture AAG and SKSR obtained via LDMS (a) without nanoparticle and (b) with silicon nanoparticle shows detection of various forms of molecular ion adducts and diverse diagnostic fragments that support peptide *de novo* sequencing. Spectral intensities are normalized to the maximum peak intensity observed in each scan. Spectrum obtained without silicon nanoparticles was identified with a wider spread and more diverse forms of sequence ions of SKSR compared to analysis with silicon nanoparticles. The characteristics of all identified peaks are listed in Table C1. The raw transient data will be accessible in Appendix C.

4.4. Discussion

The data presented here demonstrate that LDMS analytical techniques, as validated by the CORALS advanced prototype instrument, enable the detection and sequencing of putative peptide biosignatures of one representative psychrophile without the application of traditional organic MALDI matrices. Distinctive species of molecular ion adducts are identified to enable peptide detection and different fragmentation mechanisms to produce peptide fragments are discussed below.

4.4.1. Distinctive species of molecular ion adducts to enable peptide detection

We examined a series of consecutive scans of SHD and GDAE obtained in matrix-free and matrix-assisted LDMS analysis. The molecular ion (i.e., M^+) was not seen for any of the peptides examined here, with or without the application of silicon nanoparticles. This observation implies that photoionization (i.e., bond gap transition via direct absorption of UV photons) is not a dominant process during LDMS analysis of 3-mer and 4-mer peptides. Most of the peptide molecules were detected as molecular ion adducts in positive ion detection mode, implying the preferred ionization pathway is to attach a proton or a cation to the analyte molecule. The sources of positively charged species could be protons derived from the solvent during sample preparation, protons transferred from analyte or matrices, residual salts from the substrate, or metals emitted from the sample plate (Bae et al., 2012; Jaskolla & Karas, 2011).

In particular, silicon-based inorganic matrices or substrate, such as etched porous silicon surface (Alimpiev et al., 2001; Kruse et al., 2001) and porous silicon powder (Zhang et al., 2001), have shown to promote the protonation of analyte due to migration of proton from the surface Si-OH group as a result of increasing surface acidity upon laser irradiation. Similar ionization processes could also happen to LDI analysis in presence of silicon nanoparticles, but detailed mechanisms contributed to these ionization processes are beyond the scope of this study.

An effective scan to enable peptide detection is defined as a single-shot mass spectrum that contains at least one form of molecular ion adducts (i.e., $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+Al]^{+,}$ $[2M+H]^+$) with S/N > 3. The exact masses of target 3-mer or 4-mer peptides can be deduced from any one form of molecular ion adducts because the exact masses of proton and metals are known and have diagnostic mass defeats. Figure 5 summarized the number of effective scans in multiple consecutive single shot measurements in pie charts. The addition of silicon nanoparticles increased the rate of effective scans by 75% and 58% for SHD and GDAE, respectively. We further characterized the types of molecular ion adducts identified in each effective scan and summarized them in the corresponding histograms (**Figure 4.6**). The compilation of matrix-free LDMS measurements revealed that only protonated SHD and cationized GDAE were observed. Such preference was maintained with addition of silicon nanoparticles and increasing types of molecular ion adducts were identified at higher S/N for both peptides. This observation suggests that if SHD and GDAE were mixed in an unknown mixture, they would be most likely identified as protonated SHD and

sodiated GDAE by the CORALS instrument. Overall, the use of silicon nanoparticles helps to promote ionization efficiency, increase reproducibility of measurements, and improve surface homogeneity, so the probability of identifying SHD and GDAE in an unknown mixture would increase if silicon nanoparticles were used. This conclusion is in agreement with a previous study on small molecule analysis via MALDI with silicon nanoparticles using commercial instruments (Wen et al., 2007).

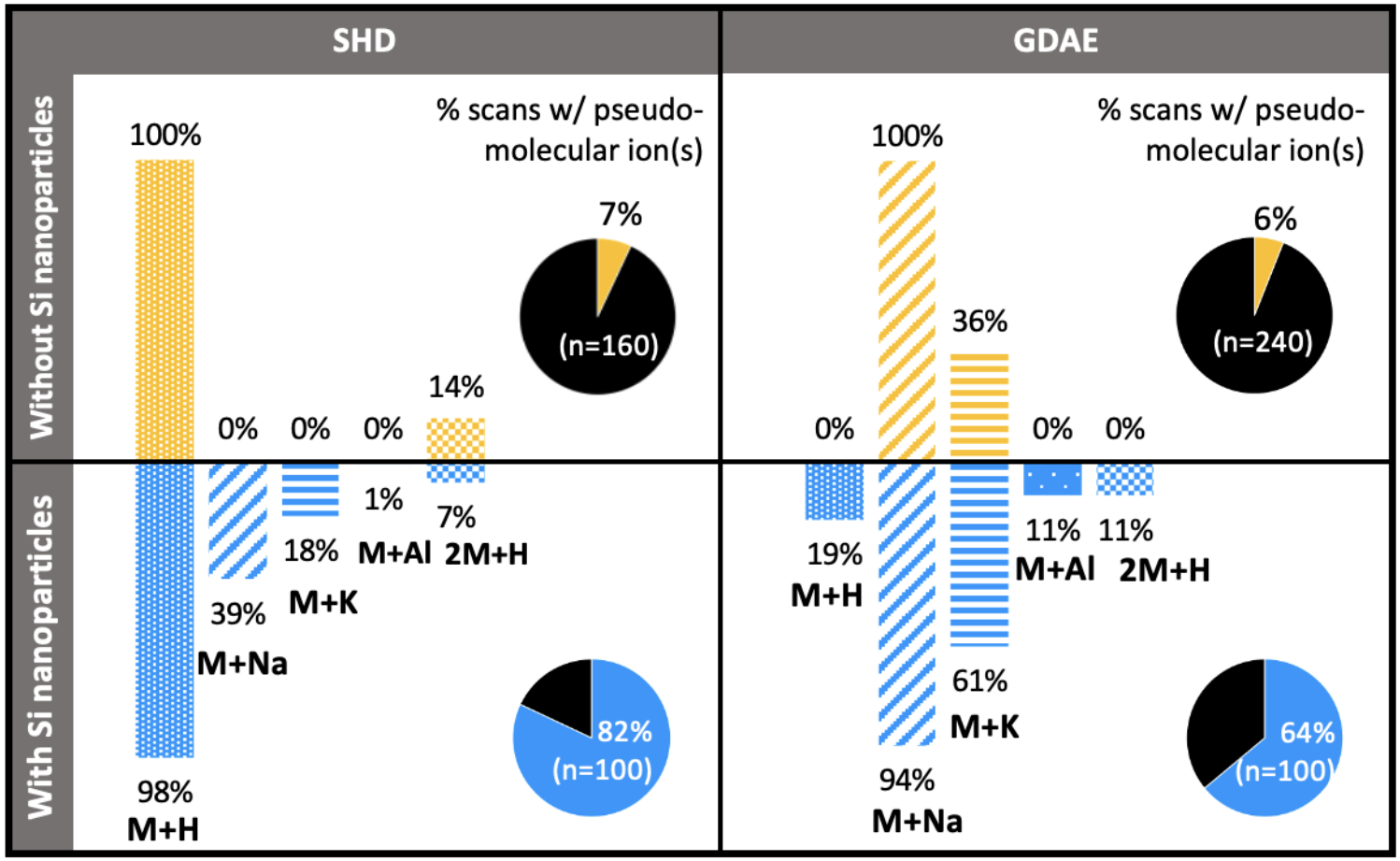


Figure 4.6. The pie charts to the right of each panel summarize the percent of total scans with at least one molecular ion adduct to identify SHD and GDAE without (top) and with (below) nanoparticles. The corresponding histograms show the frequency distribution of identified molecular ion adducts among *n* number of consecutive scans.

4.4.2. Various forms of peptide fragments to enable de novo sequencing

Electron transfer/capture dissociation (ETD or ECD), ultraviolet photodissociation (UVPD), and collision-induced dissociation (CID) are three primary dissociation methods used in conventional tandem mass spectrometry to achieve peptide *de novo* sequencing. Similar charge-driven, collision-induced, and photon-activated dissociation could have occurred during LDI measurements. Observation of ETD/ECD-like and CID-like fragments, as well as UVPD-specific fragments, provides insights into mechanisms associated with the peptide fragmentation prior to entering the mass analyzers.

ETD/ECD-like fragments, or c-/z- sequence ions, have been observed in this study, suggesting that charges are introduced to trigger peptide fragmentation during LDI measurements (Syka et al., 2004). Potential sources of charges in the ionization source include electrons in the plume (Karas et al., 2000; Nielsen et al., 2003), plasmon-induced electrons on the metal plate surface or metallic nanoparticles (Clavero, 2014; Mukherjee et al., 2013; Szczerbiński et al., 2020), and/or hydrogen radicals derived from solvents or matrices (Demeure et al., 2010; Köcher et al., 2005). Though this study is incapable of distinguishing the primary donor of charges during LDI analysis, observation of c-/z- sequence ions in measurements with and without matrices suggest charge-driven dissociation is not limited to the silicon nanoparticles itself. The use of silicon nanoparticles does not strongly favor the generation of c-/z- sequence ions, like some of the solid organic matrices (e.g., 1,5-DAN) would do due to their high yields of hydrogen radical production (Memboeuf et al., 2010).

UVPD-specific fragments, or a-/x- sequence ions, are observed at S/N > 10, suggesting that peptide molecules dissociate as they interact with 266 nm radiation. In analysis of peptide mixture AAG and SKSR studied here, a-/x- sequence ions are identified exclusively to AAG, but no such fragments were identified for SKSR, suggesting the effectiveness of UVPD is specific to analytes themselves, especially their UV absorbance. The use of silicon nanoparticles promotes the generation of UVPD-specific fragments from AAG.

CID-like fragments, namely a-, b-, and y- sequence ions, are the dominant sequence ions observed in this study, suggesting that thermal dissociation plays a critical role to fragment peptides via LDI analyses (Falick et al., 1993; Medzihradszky et al., 2000). Peptide molecular ions could be thermally activated by ion-neutral collisions in the ionization source. Ion-neutral collisions have been observed at the front of expanding laser-induced plasma plumes via fast photography (Gusarov et al., 2000; Kushwaha & Thareja, 2008; Valverde et al., 2014), suggesting CID-like fragments could be generated as the expanding plume materials collide with the background neutrals in atmospheric pressure. Similar effects albeit at a much lower pressure level occur upon expansion into vacuum (Spengler et al., 1991).

Peptide molecular ions could also be thermally activated by the heat conducted from the UV-absorbing silicon nanoparticles and/or the Al alloy sample plate. Controlled amount of thermal energy absorbed by molecular ions can facilitate desorption of intact molecules and produce CID-like fragments to facilitate peptide *de novo* sequencing (Kim et al., 2012, 2013). These fragments were generated from metastable decay and

were shown with reduced spectral resolution due to peak broadening and low S/N (Spengler et al., 1991). If the thermal energy is excessive, dissociation of protonated peptides could occur at any bond via an energy randomization process, generating a sea of fragments that are not unique to the peptide backbone nor comply to commonly-accepted mobile proton frameworks, and even generation of recombination products as the plume cools (Paizs & Suhai, 2005; Wysocki et al., 2000). For example, the non-diagnostic fragments and inorganic clusters observed in the neat analysis of SHD challenge spectral interpretation and hamper peptide *de novo* sequencing. Therefore, optimal performance of peptide detection and sequencing via LDMS analysis requires active control of thermal-induced molecular desorption and breakdown.

Silicon nanoparticles appear to form an effective matrix to facilitate the reproducible formation of molecular ion adducts at increasing S/N and reduced metastable fragmentation. Future work needs to explore how different inorganic matrices (e.g., nanoparticle versus nanowire, distinct composition, particle size distribution) control ionization efficiency, sensitivity, and fragmentation potential for specific classes of molecules. Different data acquisition and data processing strategies should be explored to maximize science return given the limited data storage and processing power onboard; for example, summing multiple scans can reduce noise floor and improve the detection limits when time and energy are available, and use of machine learning algorithms to extract and identify molecules from sample mixtures (Theiling et al., 2022). Based on findings present here, a notional concept of operation for astrobiology landed mission to a cryogenic ocean world might include sample acquisition,

deposition and sublimation of icy samples onto a suite of sample plates preloaded with a variety of inorganic matrices.

4.5. Conclusion

In summary, this study demonstrates successful peptide identification via simple LDMS techniques, and *de novo* sequencing of select peptides with the addition of silicon nanoparticles. Coupled to a high power UV laser, the silicon nanoparticles enhanced ionization efficiency and promoted the generation of diagnostic b-/y-/a-sequence ions in a single scan, similar to CID techniques. This preliminary analysis opens the door for future investigations into peptide detection and sequencing via inorganic matrices like silicon nanoparticles, alternatives to conventional solid organic matrices that are incompatible with planetary protection measures and represent risks for forward contamination for missions to astrobiological destinations.

Further, this study showcases the analytical capabilities of spaceflight prototype CORALS to detect and sequence putative peptide biosignatures of psychrophiles for *in situ* investigation of ocean worlds. Even with the simplest sample preparation method, a single spectrum collected from the CORALS prototype reveals: (1) various forms of molecular ion adducts across a wide mass range, providing for robust identification and cross-validation to the chemical composition of molecular ions with high mass accuracy measurements, and; (2) diagnostic peptide fragments at a full sequence coverage adequate to elucidate the monomer sequence in samples with silicon nanoparticles added; (3) improved resolution and mass accuracy due to reduction of metastable decay with silicon nanoparticles. The benefits of incorporating silicon or

other metal/metalloid nanoparticles into LDMS techniques could increase the science return of future spaceflight investigations.

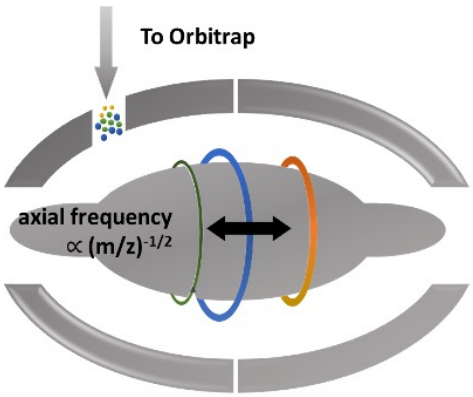
4.6. Acknowledgment

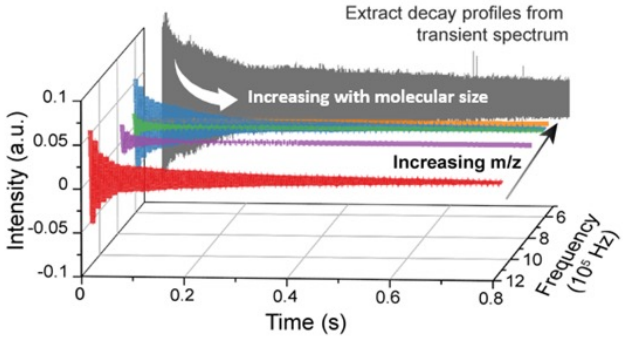
We appreciate the thorough review from Paul Mahaffy. This study was supported financially by NASA ROSES ICEE 2 Grant 80NSSC19K0610, ROSES DALI Grant 80NSSC19K0768, and GSFC Grant 80NSSC18K1612. LC was supported by an appointment to the NASA Postdoctoral Program at the NASA Goddard Space Flight Center administered by USRA and ORAU through a contract with NASA, by the NASA-funded Laboratory for Agnostic Biosignatures Project (NASA Grant 80NSSC18K1140), and by the Mars Science Laboratory mission.

Simultaneous Measurement of Mass and Collision Cross Section via Orbitrap Mass Spectrometry

Ziqin Ni & Ricardo Arevalo Jr.

[1]University of Maryland, College Park, MD, 20742.





Chapter 5: Simultaneous Measurement of Mass and Collision Cross Section via Orbitrap Mass Spectrometry

Quick overview of the chapter:

Development of using transient decay from an orbitrap analyzer to infer molecular collision cross section (CCS) has enabled characterization of unknown compounds with both mass and structure. However, current method looks at pure samples and require exact P, T measurements of the Orbitrap analyzer for accurate CCS calculation, which limits its practical implication in studying low abundance small molecules in sample mixtures. Can we make accurate and reproducible measurements of CCS values of small molecules via Orbitrap mass spectrometry? Can this method be applied to identify low abundance chemical species in a sample mixture? If so, what are the threshold conditions that need to be satisfied? In this chapter, we establish new data processing procedure and identify threshold conditions to enable accurate and reproducible CCS measurements via Orbitrap mass spectrometry. The outcome of this research enables orthogonal identification of molecules via both mass and CCS values, and thus providing additional toolbox to characterize potential chemical signature of life via mass spectrometry.

In this chapter, I conceived the experiment, acquired and processed the dataset, interpreted the results, made the figures, and wrote the papers. This chapter has been submitted for publication.

5. Abstract

Increasingly precise and accurate mass measurements via Orbitrap™ mass spectrometry have been instrumental to detect and identify the molecular stoichiometry of chemical species in a sample mixture, yet it remains challenging to reveal their identities without knowledge of their functional groups, 3D structures, and rotational orientation. Collision cross section (CCS) can be used as a complementary descriptor specific to the conformation (i.e., size and shape) of the analyte. Here, we acquired spectra of pure analytes (oligopeptides KDA and GDAE), a 16 amino acid mixture, and a multicomponent sample containing both linear and cyclic compounds (GDAE, MRFA, and Ultramark 1621) using two commercial Orbitrap mass spectrometers (LTQ Orbitrap XL™ and Q Exactive™ Basic from Thermo Scientific™) with electrospray ionization source (ESI) in positive ion mode. The transient signals were processed using Fast Fourier Transform (FFT) to produce mass spectra, and then short-time Fourier Transform (StFFT) to derive decay profiles of select ionic species. Transient decay rates were extrapolated from the decay profiles and translated into CCS values using two internal standards in the same mass spectrum. Limitations in the precision and accuracy of CCS measurements reflect local space charge effects in the analyzer. Threshold conditions to derive CCS values with $\leq$ *0.5%* uncertainty include identification of peaks in mass spectrum ($S/N \geq 120$), precise determination of decay constants ($RSD_c \leq 8\%$), accurate exponential fit to decay profiles ($\sigma_{R^2} \geq 0.98$) and high reproducibility ($RSD_{R^2} \leq 0.2\%$). Though the accuracies of CCS measurements remain insufficient to enable standalone identification of chemical compounds, the

established CCS-m/z distribution was found to be efficient at distinguishing molecular class, structure, and composition in the sample mixture.

5.1.Introduction

Mass spectrometry (MS) engenders a wide range of analytical techniques that rely on the measurement of mass to charge ratios (m/z) of ionized species to infer the chemical composition of the targeted analyte. Orbitrap MS, a specific branch of high resolution MS, offers ultrahigh mass resolving powers up to $m/\Delta m > 10^6$ Full Width Half Maximum (FWHM) at m/z 100; (Zubarev & Makarov, 2013), ppm-level mass accuracy (Olsen et al., 2005), pronounced sensitivities to trace single ion (Wörner et al., 2022), and parallel analysis of ions across a wide range of m/z (Makarov, 2000). Orbitrap MS has been used to investigate some of the most complex sample compositions, such as the nonvolatile fraction of petroleum (e.g., Niyonsaba et al., 2019), insoluble organic matter in meteorites (e.g., Hashiguchi & Naraoka, 2019), and omics extracted from cells (e.g., Wang & Bodovitz, 2010). Consequently, Orbitrap mass analyzers are routinely used for clinical research (e.g., Brown et al., 2020), pharmaceutical development (e.g., Hernández et al., 2014), aerospace exploration (e.g., Arevalo et al., 2023), and other demanding applications.

Precise and accurate measurements of m/z have been indispensable for deconvolving isobars and informing the atomic composition of unknown chemical species, particularly those present in low abundances within multicomponent sample mixtures. However, atomic composition alone is insufficient to elucidate the connectivity of atoms, the spatial arrangement and orientation of bonds in three dimensions. To

facilitate the identification of chemical unknowns, complementary information is needed to elucidate the molecular structures (i.e., linear or aromatic), molecular classes (i.e., lipids or peptide), functional groups (i.e., amine or aldehyde), etc.

Collision cross section (CCS) has been used as a complementary parameter, in addition to m/z, for chemical identification. Collision cross section is defined as the effective area around an ion within which its momentum can be transferred or lost upon collision with a neutral gas molecule. By measuring the conformation of the ionic species, a function of molecular size and shape, CCS values provide insights in the chemical constitution (including functional groups and molecular class), structure (including molecular geometries, conformation, and stereochemistry), and geometry (such as protein folding and complex formation) (A. Picache et al., 2019; Burnum-Johnson et al., 2019; May et al., 2014). Today, CCS values of chemicals are commonly measured via ion mobility mass spectrometry (IMS) (e.g., (Beegle et al., 2001)), a technique that can achieve chemical identification by itself or serve as a hyphenated front end to a mass spectrometer. CCS values can also be predicted via computational models and machine learning algorithms, such as the MetCCS predictor model (Zhou et al., 2016). Complementary measurements of CCS and m/z narrow down various prospective molecules to a smaller set of highly probable candidates in existing databases, thereby increasing the capability and confidence of identification.

There is an ongoing effort to achieve simultaneous measurements of CCS and m/z with a single analytical system, to maximize the identification power of chemical unknowns without multiple platforms. Orbitrap MS is placed at the center of this initiative. The

Orbitrap mass analyzer consists of a spindle-like central electrode (CE), a pair of barrel-like outer detection electrodes (DE), and a deflector electrode adjacent to the ion injection slit (**Figure 5.1**). The CE and DE are shaped such that together they form an electrostatic field (E_{CE}) with a quadro-logarithmic trapping potential that promotes axial oscillation along the CE solely dependent on the m/z of analyte ions. During data acquisition, the CE is ramped to a stabilized voltage (e.g., ±3.5 kV (Perry et al., 2008)), resulting in ions of different m/z would accelerate to arrive at different concentric orbitals around the CE, a process known as electrodynamic squeezing. Low m/z ions orbit closer to the CE and oscillate axially along the CE at a higher frequency than high m/z ions. The DE collects image current induced by the axial oscillation of ion packets along the CE, which is then amplified, digitized, and output as transient signal.

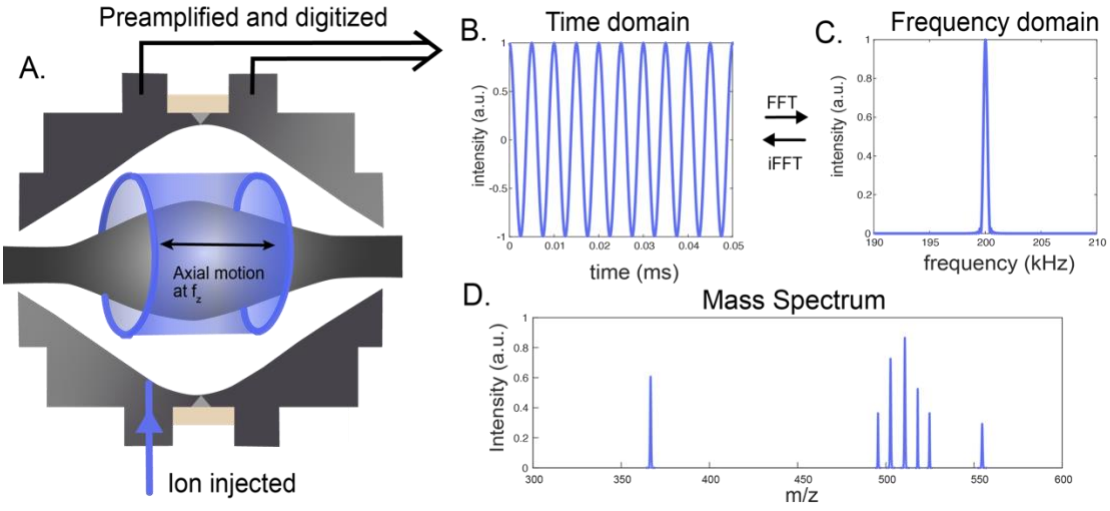


Figure 5.1. Schematics to demonstrate relationship of ion motion and ion distribution in phase domain to frequency domain.

A typical data processing routine in Orbitrap MS includes: i) Fast Fourier Transform (FFT) of the transient signal to produce a frequency spectrum, ii) calibration of the frequency spectrum and transmutation into a mass spectrum using internal standards; and, iii) peak analysis in the mass domain to enable chemical identification and quantification. FFT has been a powerful tool to process transient signals. FFT deconvolve transient signals into components of harmonic waveforms at different frequencies. The implementation of FFT returns the average magnitude and frequency of individual waveforms over the data acquisition time, providing information about the signal intensity (or relative abundance) and axial frequency of ionic species within the sample mixture respectively.

Transient signals are observed to decay over time. Because ions inside the Orbitrap analyzer are accelerated to move under a high voltage potential (i.e., >3.5 keV), a single ion-neutral collision is assumed to remove ions from coherent ion packets, leading to decay in the transient signal or known as transient decay. Therefore, the rate of transient decay is informative about the CCS values.

Previous studies have successfully made CCS measurements of ions via Orbitrap MS (James et al., 2022; Makarov & Denisov, 2009; Sanders et al., 2018b). These studies were able to determine CCS values of multiply charged kDa-sized proteins within 7% of ground-truth values derived from IMS experiments. Although these demonstrations highlight the feasibility of measuring CCS values via Orbitrap MS, an open question remains whether such techniques can determine CCS values for low mass, singly-charged ions with smaller relative variations in CCS, such as individual amino acids. A

higher CCS resolution is required to discern ions of similar structures and/or masses due to greater similarities in spatial arrangements of atoms. New data processing technique and method is explored to provide a greater sensitivity in CCS measurements via Orbitrap MS.

This research aims to explore empirically the capability of making CCS measurements of small molecules (m/z < 500) in organic-rich samples of varying complexity and concentration via Orbitrap MS. In particular, the CCS measurements were evaluated based on its reproducibility and accuracy. Threshold conditions, such as the signal-to-noise ratio (S/N) of the peak and goodness-of-the-fit to the decay profiles, were identified as guideline for practical implementation of the method. Further, characteristic trends in m/z and CCS distribution were used to distinguish molecule of different structures and classes.

5.2.Materials and methods

5.2.1. Materials preparation and data collection

Chemicals and Reagents. The powdered Lys-Asp-Ala (KDA) and Gly-Asp-Ala-Glu (GDAE) analytes were procured from Biomatik (purity >95%) and dissolved in 0.1 wt.% formic acid (Fisher Chemical^{TM}, ≥ 99.0% purity), resulting in ~200 μM peptide solutions. These single component solutions were used to test various methods to derive m/z and CCS from transient signals under the near ideal conditions.

A 10 mM amino acid solution procured from Waters (Item WAT088122) was diluted by a factor of 10 in MilliQ water to produce a final solution with 100 μM of each amino

acid. This low concentration solution mixture was prepared to explore the threshold conditions of deriving m/z and CCS from a sample containing multiple organic compounds, and thus a higher density of peaks in the mass spectrum. The amino acids comprising this solution, in order of increasing molecular weight, are: glycine, L-alanine, L-serine, L-proline, L-valine, L-threonine, L-leucine and L-isoleucine, L-aspartic acid, L-lysine, L-glutamic acid, L-methionine, L-histidine, L-phenylalanine, L-arginine, and L-cystine (an oxidized dimer of the α-amino acid L-cysteine).

Finally, a multicomponent solution composed of 10 μM GDAE solution and the Pierce™ LTQ Velos ESI positive ion calibration solution procured from Thermo Scientific™ (Catalog number: 88323), which includes caffeine (100 μM), Met-Arg-Phe-Ala (MRFA; 2 μM), and Ultramark 1621 (50 μM) was prepared to investigate advantages of distinguishing and separating molecular classes and shapes in sample mixtures via simultaneous measurements of m/z and CCS. Both GDAE and MRFA are 4-mer oligopeptides with a backbone consisting of alternating carbon and nitrogen atoms. Compared to GDAE, MRFA has a larger molecular size, and contains heavy S atom, aromatic ring and long carbon chain in the amino acid residues. Ultramark 1621 has a compact and dense ring structure composed of three N and three P atoms with alternating single and double bonds. Each P atom is connected to an O atom and two side chains that have a basic structure of $CH_2(CF_2CF_2)_nH$, where n = 1, 2, or 3. Due to the possible range of n values, a total of six different side chain positions exist in Ultramark 1621, which is characterized by a wide range of diagnostic peaks between m/z 1022 – 1922.

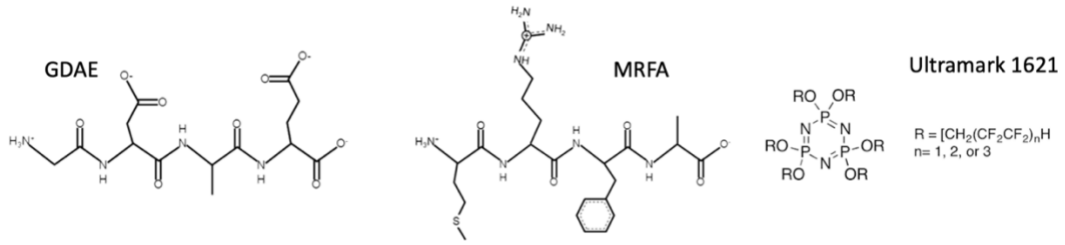


Figure 5.1. The molecular structure of GDAE, MRFA, and Ultramark 1621 in compound mixture solution.

Instruments. We collected spectra in positive mode using an LTQ Orbitrap XL and a Q Exactive commercial mass spectrometer from Thermo Scientific™ (Bremen, Germany). The instruments were directly connected to a heated electrospray ionization (ESI) source without any chromatographic separation. Data were collected across a range of m/z 100-2000 Th with varying transient lengths, which also resulted in varying mass resolving powers (**Table 5.1**). Raw transient data were extracted from both instruments with the support of Thermo Scientific™ for the purpose of this study.

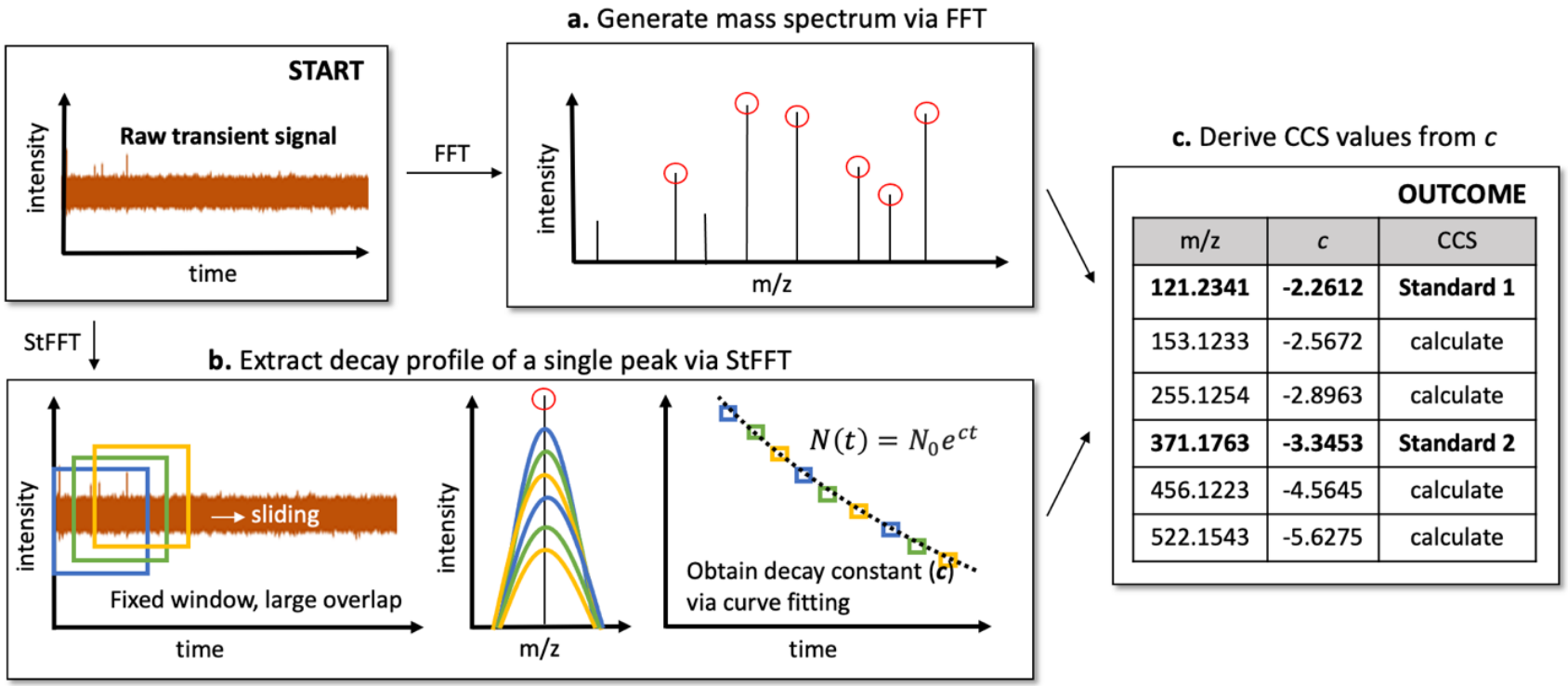
Table 5.1. A summary of samples that have been analyzed by this study.

Sample	Instrument	Time (s)	Mass range	# of scans
KDA	Q-Exactive	0.512	100-800	20
GDAE	Q-Exactive	0.256	100-800	20
Amino acid mixture	LTQ-XL Orbitrap	0.256	100-500	54
Solution mixture	LTQ-XL Orbitrap	0.256	100-2000	14

Data processing. Each transient signal was processed with 10× zero filling and the hanning apodization prior to FFT implementation (**Schematics 5.1**). A single internal standard was used for mass calibration. The measurement of *m/z* and intensity of the interested peaks were obtained via parabolic fitting from the five most abundant points around the selected peak apex. The *S/N* of a selected peak was measured as the ratio of

peak intensity to the noisefloor, which is estimated as 3x standard deviation from the average of noise background in the vicinity of the peak.

Transient signals were also processed via the short-time Fourier Transform (StFFT), a variation of FFT in which the original time-domain transient is divided into shorter segments (or slices) before FFT implementation, to extract transient profile of each ionic species as a function of time. Each slice is independently processed via FFT to generate a corresponding mass spectrum. Extracting peak intensities in these mass spectra in a sequential order provides information about the progressive decay of signal intensity as a function of time. The two critical parameters to implement StFFT are the width of each transient slice and the degree of overlap between consecutive slices (***D1***). As the width of a single slice gets larger, the mass resolving power in the corresponding mass spectrum increases. A 100 ms transient signal can generate mass spectra with a mass resolving power greater than 10,000 at full width half maximum at m/z 100-2000 Th, which is sufficient to identify peaks of interest in this study. As the degree of overlap increases, the observed signal intensity over time becomes more smoothly connected, improving the fit of decay profiles (i.e., higher R^2) and by extension increasing confidence in the inferred decay constant, *c*. For example, a 512 ms transient signal was sliced into approximately 50 segments of 100 ms transient with 90% overlap. Peak intensities in consecutive mass spectra were connected to produce decay profiles of specific ionic species. Each individual decay profile (specific to each peak of interest) was fitted with an exponential decay function with characteristic decay constant *c*, which was then used to determine a representative CCS value, as described further below.



m/z	c	CCS
121.2341	**-2.2612**	**Standard 1**
153.1233	-2.5672	calculate
255.1254	-2.8963	calculate
371.1763	**-3.3453**	**Standard 2**
456.1223	-4.5645	calculate
522.1543	-5.6275	calculate

Schematics 5.1. A schematic flowchart that shows the steps to obtain CCS and m/z measurements of identified peaks in a mass spectrum. The transient signal is processed with (**a**) Fast Fourier Transform (FFT) to produce a mass spectrum and (**b**) short-time Fast Fourier Transform (StFFT) to extract the decay profile that can be fitted to derive the decay constant c using simulated dataset. (**c**) The decay constants of all identified ionic species are converted to CCS values using two identified peaks of known CCS values as internal standards.

Mathematical equation. In the Orbitrap analyzer, the axial oscillation frequency of a single ion packet (f) is related to its m/z at a factor of a restoring factor k,

$$f = k\left(\frac{m}{z}\right)^{-\left(\frac{1}{2}\right)}. \qquad \textit{Eqn (5.1)}$$

Because the Orbitrap analyzer is operated at a high voltage potential (~3.5 keV), ion-neutral collisions are assumed to follow an energetic collision model, meaning a single ion-neutral collision can remove ions completely from the coherent ion packet. Thus, the rate of transient decay can be modeled via an exponential decay function:

$$N(t) = N_0 e^{-ct}, \qquad \textit{Eqn (5.2)}$$

where N_0 and $N(t)$ are the number of ions at time zero and at time t, and c is the decay constant that is specific to the rate of ion-neutral collision for individual ion packets. If the distance of a single axial oscillation is L, the total traveling distance during data acquisition time t would be Lft. The average distance between two successive collisions is defined as mean free path l, equivalent to:

$$l = (n\sigma)^{-1}, \qquad \textit{Eqn (5.3)}$$

where the σ is the CCS of the ion(s) and n is number of gas molecules in the Orbitrap analyzer. The total number of ion-neutral collision expected for a single ion packet would be $Lftn\sigma$. Therefore, the CCS values of ions can be measured by extrapolating the decay constant c from transient signals,

$$c = Lfn\sigma. \qquad \textit{Eqn (5.4)}$$

If we assume ions at different *m/z* travel a similar distance L per single axial oscillation and number of background gas molecules *n* is the same per single analysis, then *Eqn (4)* could be simplified as $c \propto f\sigma$ for different ionic species in the same analysis at constant vacuum conditions. A linear regression is applied to obtain corrected measurements of decay constants (c') for each individual analysis,

$$c' = ac + b. \qquad \textit{Eqn (5.5)}$$

The calibration factors *a* and *b* can be inferred from two internal standards of known *m/z*, by extension *f*, and CCS values σ. Internal standards help to establish the linear correlation between the empirical *c* (from fitting of transient decay) and theoretical c (from $c \propto f\sigma$). The calibration factors *a* and *b* are then used to derive σ for the rest of the ions via:

$$\sigma_n = (c_n{}' - b)/(af_n). \qquad \textit{Eqn (5.6)}$$

5.3. Results

5.3.1. Analysis of oligopeptides KDA and GDAE in pure sample

In order to demonstrate the applicability of the relationships outlined above, this investigation first examined transient signals acquired during the analysis of single component peptide solutions (~200 μM; see Methods). The mass spectra of KDA and GDAE demonstrate reproducible detection of the protonated molecular ion $[M+H]^+$, sodiated molecular ions $[M+Na]^+$, and protonated dimer $[2M+H]^+$ in 20 consecutive analyses. For each scan, the decay profile of each ionic species was extracted using StFFT (**Figure 5.2**). The average R^2 of the exponential fits (μ_{R^2}) is > 0.95, providing statistical confidence to the derivation of decay constant *c*. The average and relative standard deviation (RSD) of decay constant *c* over 20 scans are reported in **D2**. The CCS values of $[M+H]^+$ and $[M+Na]^+$ for both KDA and GDAE, which were assigned via the MetCCS Predictor model (see **D3**), were used as internal standards to estimate the CCS value of $[2M+H]^+$ for KDA with 5% precision (*RSD*) and GDAE with 1% precision (*RSD*).

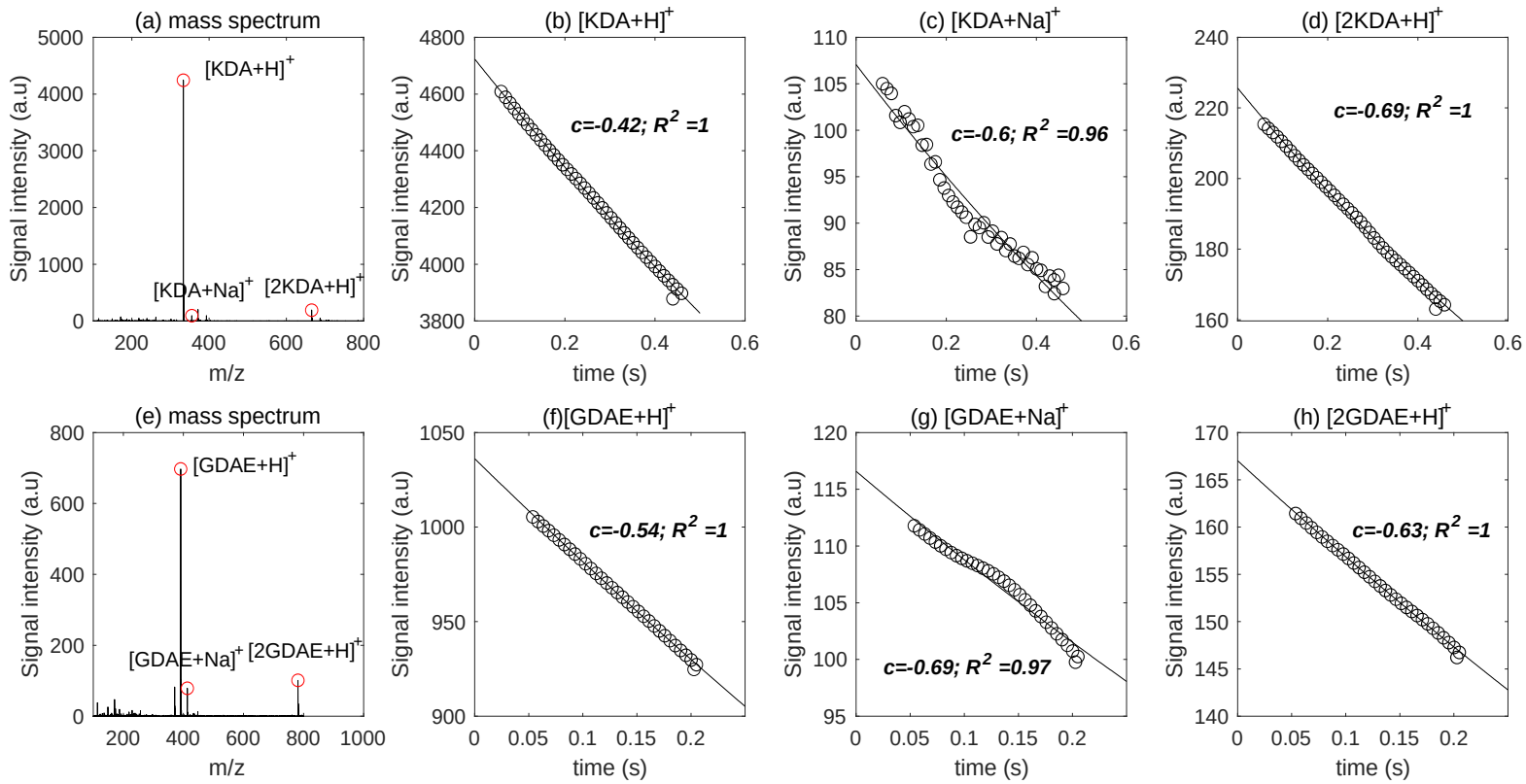


Figure 5.2. The decay constant of identified ionic species in a single analysis of 200 μM oligopeptide KDA (top) and GDAE (bottom) in 0.1% formic acid solution. (**a, e**) A typical mass spectrum of KDA and GDAE, respectively. (**b-d, f-h**) The decay profile of each identified ionic species was extracted via StFFT from the single mass spectrum in panel (a) and (e), respectively. Each decay profile was fitted using an exponential function to derive decay constant c; the goodness of the fit was evaluated by R^2.

The next step of this investigation focused on demonstrating the same methodology to study a mixture of amino acids at low concentration (100 μM). Protonated molecular ions for 12 of the 16 amino acids in the solution were identified at *S/N* 10 – 400; glycine, alanine, serine, and aspartic acid were observed below the limit of quantitation (LOQ at *S/N* < 10; **Figure 5.3**). The decay profile of each protonated amino acid was extracted, and the statistics of the exponential fitting was summarized in **D2**. Protonated amino acids with *S/N* < 30, namely valine, threonine, lysine, and glutamic acid, exhibited modulations in their respective signals. Statistically, these modulations were characterized by an average R^2 (μ_{R^2}) <0.95, relative standard deviation of R^2 (RSD_{R^2}) > 0.1%, and relative standard deviation of decay constants (RSD_c) > 5%, suggesting that randomized modulation in transient signals resulted in poor exponential fits and large uncertainty in decay constants. In contrast, protonated amino acids with *S/N* > 100 were observed to have R^2 > 0.98 and RSD_{R^2} < 0.05 %, resulting in a higher confidence level in curve fitting and reproducibility of decay constants. One exception to this trend is peak at *m/z* 132.1127, which had *S/N* >100 but modulation in the transient signals was observed. This peak could be protonated leucine or protonated isoleucine which are structural isomers to each other but have different CCS values. The CCS values of protonated amino acids have been extensively studied via Fourier Transform Ion Cyclotron Resonance (FTICR) so the accuracy of the CCS measurements enabled by the data collected here will be evaluated in the following discussion.

A similar procedure was used to analyze a solution that contains organic molecules representing multiple distinct classes of molecular structures, including oligopeptides

GDAE and MRFA, and a mixture of fluorinated phosphazines, Ultramark 1621. Whereas GDAE was identified as a pronated molecular ion $[GDAE+H]^+$ and dimer $[2GDAE+H]^+$, MRFA was recognized by singly and doubly charged protonated molecular ions $[MRFA+H]^+$ and $[MRFA+H]^{++}$ (**Figure 5.4**). Ultramark 1621 was identified by a series of peaks across a mass range of m/z 1122 – 1822 Th; peaks at m/z 1022 and 1922 were not observed LOQ. Caffeine was found below LOQ. Following the protocol mentioned above, the decay profile for each peak of interest was extracted and the statistics were summarized in **Table D2**. Peaks with $S/N < 25$ (such as $[MRFA+H]^{++}$) were observed to have modulation in transient signals. The exponential curve fit to these transient signals with modulation resulted in $\mu_{R^2} < 0.97$ and $RSD_{R^2} > 10\%$ across consecutive scans, reinforcing previously observed negative impacts on the confidence level and reproducibility of the inferred decay constants. When $[GDAE+H]^+$ and $[2GDAE+H]^+$ were used as internal standards, the rest of ionic species has reproducible CCS values with $RSD_{CCS} < 1\%$ across different scans.

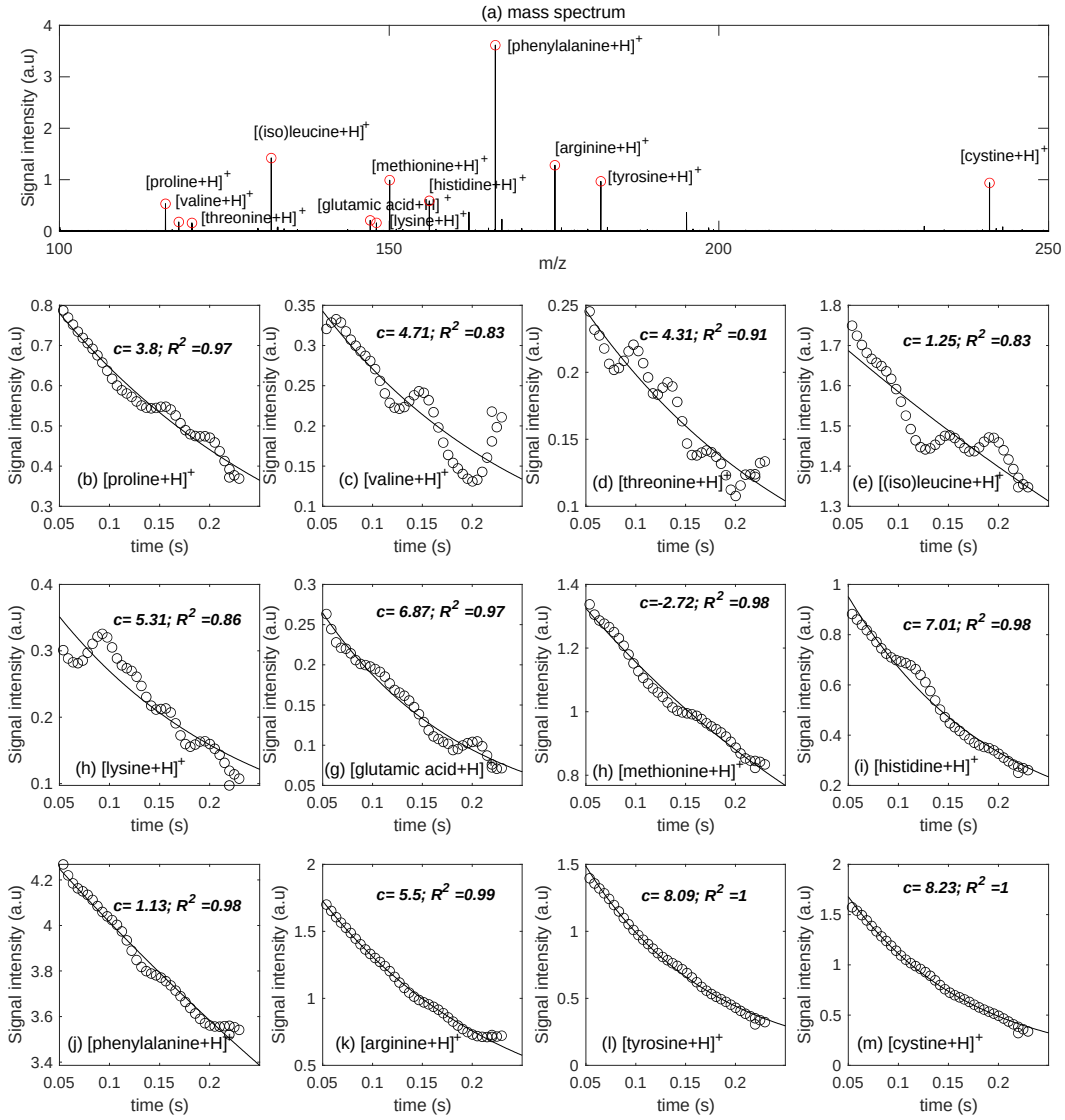


Figure 5.3. The decay constant of identified ionic species in a single analysis of a mixture of amino acids in water solution. (a) A typical mass spectrum of the solution mixture. (b-m) The decay profile of each identified species was extracted via StFFT from the single mass spectrum in panel (a). Each decay profile was fitted using an exponential function to derive decay constant c; the goodness of the fit was evaluated by R^2.

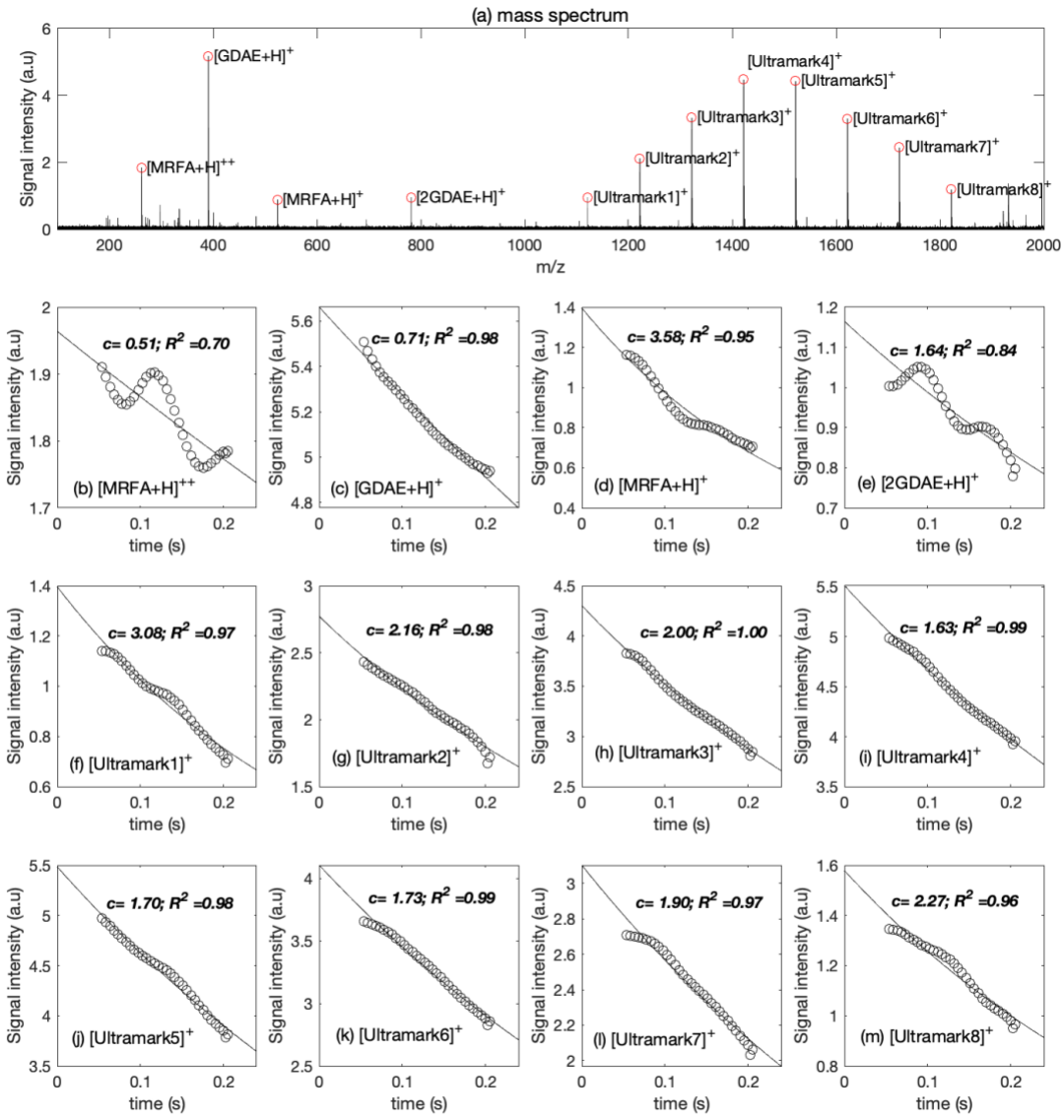


Figure 5.4. The decay constant of identified ionic species in a representative single scan of solution mixture (MRFA+GDAE+Ultramark1621). (a) A typical mass spectrum of the solution. (b-n) The decay profile of each identified species was extracted via StFFT from the single mass spectrum in panel (a). Each decay profile was fitted using an exponential function to derive decay constant c; the goodness of the fit was evaluated by R^2.

5.4. Discussion

5.4.1. Reproducibility and accuracy of CCS measurements

The methods developed here have demonstrated simultaneous measurements of m/z and CCS of targeted ionic species in a single-scan mass spectrum collected via Orbitrap MS. This method has been applied to process transient signals of high concentration analytes, a low concentration amino acid mixture, and a multicomponent solution mixture that contains different molecular structures, implying a wide applicability to investigate unknown samples of varying concentration and complexity. Chemical identification via CCS measurements is successfully shown, with CCS values empirically derived here falling within < 9% of theoretical values and/or previously published IMS determinations. Having reproducible and accurate CCS measurements are critical to identify unknown compounds with high confidence. Yet, nonideal conditions in the Orbitrap analyzer, such as space charge effects and low analyte *S/N*, introduce modulation in the transient signals and/or accelerate ion losses. The following discussion aims to (1) identify the threshold conditions needed to achieve precise and accurate CCS measurements, (2) showcase an application of using CCS-m/z distribution to accelerate identification of multicomponent in sample mixtures, and (3) discuss the causes that contribute to modulation of transient signals.

Threshold conditions to make CCS measurements. A key to make CCS measurements via Orbitrap MS is the capability to derive a precise and accurate decay constant *c* from a transient signal. A poorly fitted transient decay profile (as represented by low R^2) will generate a decay constant *c* with low confidence (as represented by

high RSD_c), which in turn will produce a CCS value with large uncertainty. As shown by the data above, poorly fitted transient decay profiles are often associated with peaks at low *S/N* in mass spectra and transient signals with modulation (as manifested by high RSD_{R^2}). Therefore, the *S/N* of the peak, average goodness of the fit of the transient signals (μ_{R^2}) across multiple scans, and reproducibility of the fits (RSD_{R^2}) are three quantifiable matrices to evaluate the RSD_c and by extension the reproducibility of CCS measurements.

Figure 5.5 compiles the RSD_c of all identified ionic species from all sample analyses in this study. The RSD_c was observed to inversely correlate to the RSD_{R^2}, confirming the importance of the decay function fit to the transient profile. To achieve CCS measurement with less than 0.5% uncertainty (or $RSD_{CCS} \leq 0.5\%$), decay constants need to be determined with sufficient precision ($RSD_c \leq 8\%$), accurate and precise fit to the decay profiles ($\sigma_{R^2} \geq 0.98$, $RSD_{R^2} \leq 0.2\%$), and signal intensity above the threshold ($S/N \geq 120$).

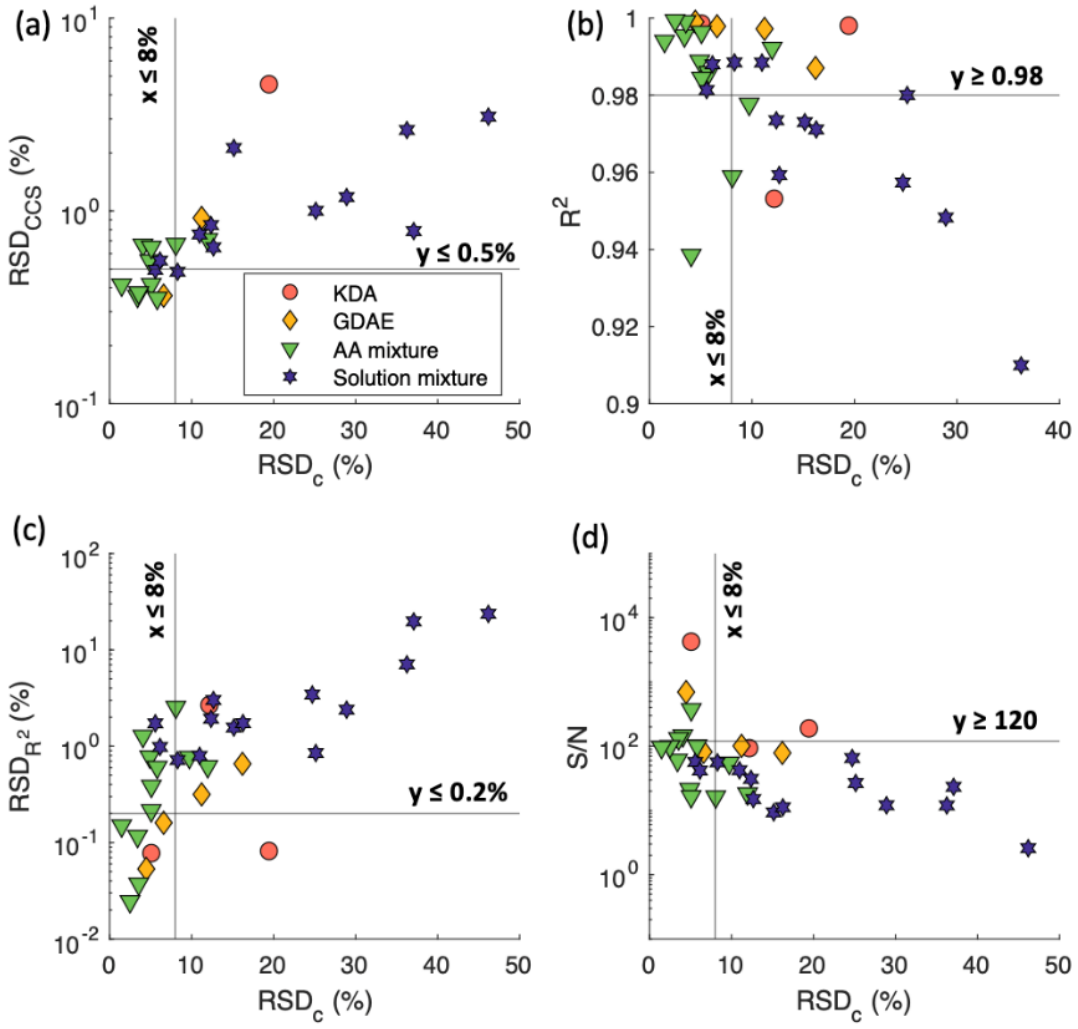


Figure 5.4. *This figure shows the distribution of relative standard deviation (RSD) of the decay rates of each ionic species as a function of: (a) the RSD of CCS values; (b) the R^2 of their decay profiles; (c) the RSD of R^2, and (d) their average S/N in the mass spectra. Threshold conditions are labeled to achieve CCS measurements with >99.5% reproducibility.*

Because CCS values were quantified with respect to two internal standards, the accuracy of CCS measurements derived here were found to be affected by the selection of those internal standards. In the case of the amino acid solution mixture, the median percent error of CCS measurements across 54 scans was -27.7% for rest of protonated amino acids when protonated lysine (m/z 147.1245) and protonated methionine (m/z 150.0703) were used as internal standards. The median percent error dropped to -5.2% when protonated threonine (m/z 120.0907) and protonated arginine (m/z 175.1329) were used as internal standards and further decreased to 0.0% when protonated proline (m/z 116.0803) and protonated tyrosine (m/z 182.0957) were used. The selection of two internal standards that are further apart in the m/z space appears to help constrain the systematic bias in CCS measurements. However, even in the best-case scenario when the median percent error of CCS measurements is 0.0%, the percent error in CCS measurements for individual amino acids is still variable, ranging from -5% to 9%, remaining insufficient to differentiate potential candidates in database using CCS values alone. Making accurate CCS measurements is a difficult task; the challenges are not unique to Orbitrap MS. Even with a conventional IMS platform, the same chemical species could be reported up to ±10% than the smallest CCS values published in literatures (France et al., 2020; Stow et al., 2017).

Characteristic trend in CCS-m/z distribution. Despite limited accuracy in CCS measurements alone, the simultaneous measurements of m/z and CCS via Orbitrap MS still serve as an efficient means to achieve multidimensional separation of molecules in sample mixtures. The characteristic trends in CCS-m/z distribution were observed to vary as a function of molecular structures. For example, in the multicomponent

sample mixture, MS analysis of solution mixture, ionic species sourced from three structurally-distinct chemical compounds, GDAE, MRFA, and Ultramark 1621, defined statistically distinct linear trends in the CCS-m/z distribution (**Figure 5.6**). The linear regression slopes for GDAE and MRFA were determined as 0.192± 0.002 (1 standard deviation (*SD*)) and 0.236± 0.003 (1*SD*), respectively, implying a similarity in structure and composition as 4-mer oligopeptides. In contrast, Ultramark 1621 has a shallower slope of 0.125 ± 0.001 (1*SD*), likely due to incorporation of heavier N and P (versus C) atoms in its cyclic structures. The uncertainties on the slopes, which reflect the reproducibility of 14 replicate measurements, suggests that such trend lines can be reproducible and thus serve as diagnostic indictors (or corroborative measures) of molecular identity. Previous studies have also identified that the distinctive CCS-m/z distributions may be defined by specific molecular structures and functional groups based on a compilation of over 3800 experimentally acquired CCS values (A. Picache et al., 2019). Due to these reasons, we argue that characteristic trends in CCS-m/z distributions allow rapid recognition and separation of molecules based on their molecular structure, shape, and composition, thus providing orthogonal information about unknown chemical species.

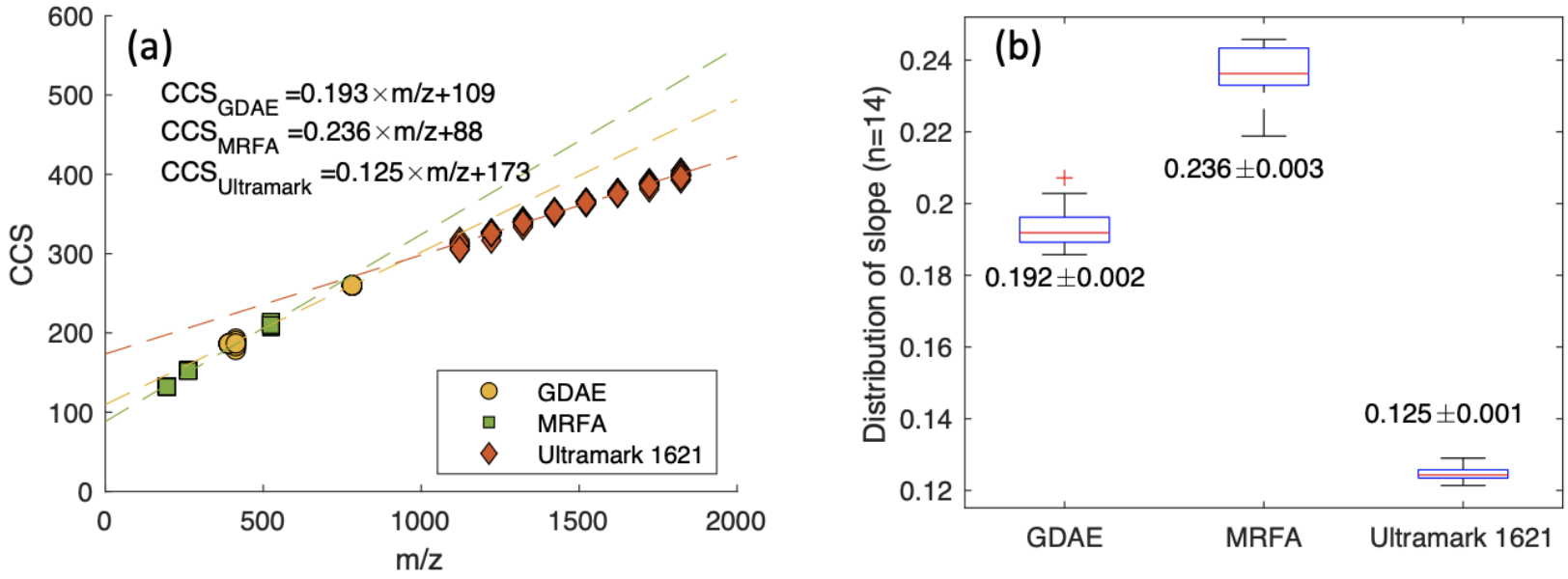


Figure 5.6. (a). A scatter plot and linear regression analyses show the CCS versus m/z distribution of ionic species identified in 14 consecutive scans of a solution mixture containing GDAE, MRFA, and Ultramark 1621. Ionic species from the same compound were grouped and characterized via a linear regression model. (b). Box plots show the distribution of slope of each compound obtained by individual scans (n = 14 total). Average values and standard deviation (SD) of the slopes are labeled.

The modulation in transient signals. Large uncertainties and errors in CCS measurements are often associated with observations of modulation in transient signals manifest as oscillatory drops and rises of signal intensity as a function of time. To further investigate this phenomenon, we compiled the StFFT mass spectra from a single transient signal to track changes in the peak frequency and intensity of monitored ionic species as a function of time. Specifically, we investigated a 0.256s transient signal of $[MRFA+H]^{++}$, $[GDAE+H]^{+}$, and $[MRFA+H]^{+}$ from **Figure 5.4**. While $[GDAE+H]^{+}$ has the highest *S/N* in the mass spectrum, $[MRFA+H]^{++}$ and $[MRFA+H]^{+}$ have similarly low *S/N* and are located on opposite sides of the $[GDAE+H]^{+}$ peak in the mass spectrum (**Figure 5.7**). The compilation of StFFT mass spectra revealed a relationship between frequency shifts and peak intensity. The peaks at low *S/N* were subject to directional frequency shifts, with $[MRFA+H]^{++}$ observed to have a positive frequency shift and $[MRFA+H]^{+}$ observed to have a negative frequency shift. The maximum frequency shift was ± 3 Hz, approximately m/z 0.002-0.008 Th in the mass domain, suggesting a localized disturbance that distorted ion motions and lead to modulation in signal intensity. Irrespective to the direction of the frequency shift, the modulated transient signal reaches its local maximum when the frequency shift is maximized, implying that frequency shifting might lead to a rise of peak intensity as a side effect. Therefore, a close examination of modulated transient signals suggests that (1) ion packets of low *S/N* are prone to frequency shifts, (2) frequency shifts are linked to changes in signal intensity, and (3) frequency shifting is directional as ions at smaller m/z, in relative to ion intensity ion packets, show positive frequency shift and vice

versa. Based on these observation, one hypothesis to explain modulation in transient signals is local space charge.

Local space charge occurs when the frequency, phase, and intensity of ion oscillation are shifted owing to Coulombic repulsion to nearby ionic species. The Columbic repulsion between two ionic species reaches maximum when they are nearest and the magnitude of repulsion intensified with their charge states. Local space charge should be localized (small scale influences), oscillatory (its impact reaches maximum when two ionic species are close and reduces as they move apart), and non-systematic (specific to charges of nearby ionic species). Local space charge effects are distinct from global space charge effects, which impose systematic errors in frequency measurements that can only be corrected during mass calibration or minimized via automatic gain control during data acquisition. Though local space charge has been predicted via theoretically modelled nonideal electric fields inside the Orbitrap analyzer (Kharchenko et al., 2012), the empirical observation, cause, and impact of local space charge, however, have not been studied in detail.

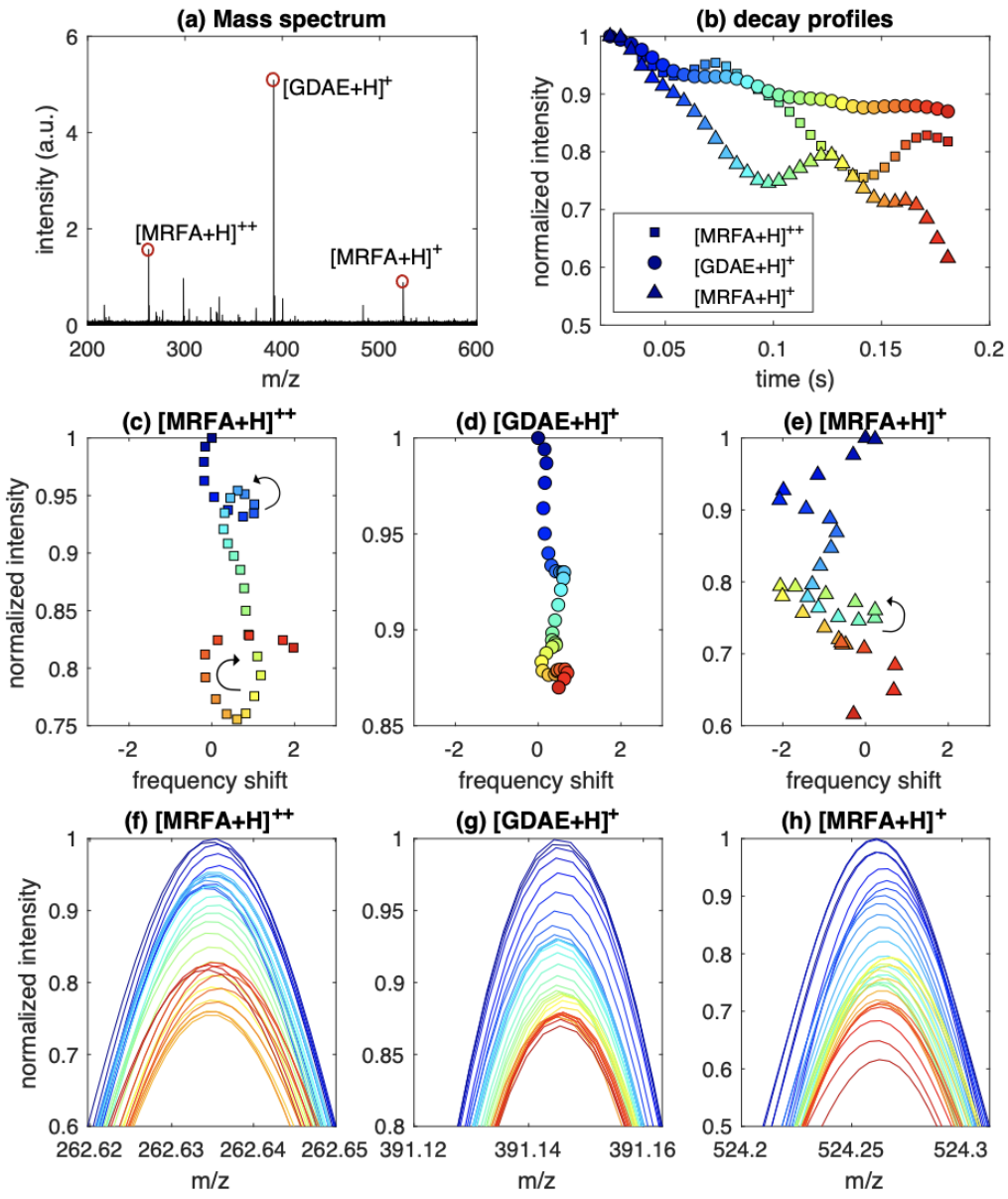


Figure 5.7. (**a**) A typical mass spectrum of $[MRFA+H]^{++}$, $[GDAE+H]^{+}$, and $[MRFA+H]^{+}$ from the multiponent solution mixture. (**b**) The normalized decay profiles of three ionic species extracted from the mass spectrum reveal their modulated patterns in time domain. (**c-e**) The normalized peak intensities (to the maximum value of each ionic species) were plotted as a function of frequency shift at each segment of transient signals via StFFT. The frequency shift is calculated as the difference in frequency from current transient segment to the initial transient segment. (**f-h**) Mass spectra generated from each segment of transient signals via StFFT show the peak shapes and relative intensities of three ionic species through time. The color represents the sequential orders of each segment.

Local space charge explains the modulation patterns in transient signals and the directional frequency shift of ions with low *S/N*. If we assume ion of a single *m/z* have a Gaussian distribution of signal intensity versus frequency, low abundance ions (low *S/N*) will be effectively "pushed" into a skewed distribution by more abundant ions (high *S/N*) during close approach between ion packets. Consequently, the low *S/N* ion packet will experience a shift in frequency and rise in intensity. As the two ion packets continue to oscillate and move away from each other, the magnitude of repulsive force will diminish, and thus the skewed ion profile will relax back to Gaussian distribution with a broader FWHM and lower peak intensity. In contrast, high abundance ions will be less influenced by interactions with other ion packets. $[GDAE+H]^+$, the highest abundance peak in Figure 5.7, was observed with less than 0.5 Hz frequency shift. This phenomenon is consistent with observations of $[MRFA+H]^+$ and $[MRFA+H]^{++}$ shifting in the opposite directions to $[GDAE+H]^+$.

Other common causes to explain nonideal transient decay from Orbitrap analyzer include imperfect machining, global space charge, and poorly tuned ion optics (resulting in unstable ion trajectories). However, these hypotheses would be expected to impose generic disturbances. Therefore, local space charge effects are considered the most probable cause of transient modulations in the Orbitrap analyzer.

5.5.Conclusion

In this work, we have showcased the value of deriving CCS values and measuring *m/z* simultaneously via Orbitrap mass spectrometry. To enable CCS determinations, decay profiles of multiple select peaks in mass spectra can be extracted using short-time

Fourier Transform. The decay constant of each exponential curve fit to the data can be extrapolated and then translated to a predicted CCS value with appropriate choice of internal standards. This method can produce CCS measurements with $RSD_{CCS} \leq 0.5\%$: specifically, the targeted ionic species should have signal intensity above threshold ($S/N \geq 120$), accurate fitting to transient decay ($\sigma_{R^2} \geq 0.98$) with reproducible fitting ($RSD_{R^2} \leq 0.2\%$), and precise determination of decay constants ($RSD_c \leq 8\%$). Though the accuracy of CCS measurements remains insufficient to make standalone identification of chemical compounds and separate structural isomers, the CCS-m/z distribution was found to be characteristic to molecular structure and composition, thereby providing a corroborative measure of molecular identity. The data processing procedure described here can be applied without modification in existing sample preparation and data acquisition routines.

The low reproducibility and accuracy of CCS measurements often stem from decay profiles with modulated patterns in the time domain, which is likely caused by local space charge. Local space charge creates correlated modulation in signal intensity and frequency shift as a function of time, and its impact is most prominent on singly charged ions with low S/N next to high abundance ions. Future research is needed to understand the roles of local space charge in abundance fractionation and signal decay inside the Orbitrap mass analyzer.

5.6.Acknowledgment

This study was supported by the NASA ROSES ICEE 2(Grant 80NSSC19K0610, PI: Ricardo Arevalo Jr.) and DALI Program (Grant 80NSSC19K0768, (PI: R. Arevalo. Jr).

We want to express special thank you to Alexander Makarov and his team at Thermo Fisher Scientific™ to provide direct access to transient signals from commercial Orbitrap instruments.

Chapter 6: Conclusion

In search for chemical signs of life in space, organic molecules are of particular interest because they are the essential components of living organisms found on Earth. However, organic molecules can be produced abiotically, and their abundance could be enhanced(diminished) by the habitable(harsh) planetary environments. A corollary to this realization is the motivation to constrain abiotic synthesis and molecular evolution of organic compounds via laboratory experiments and computational simulation. The specifics of planetary environments, such as the redox condition, is critical to infer the prebiotic pathway that produce organic compounds. Furthermore, analytical techniques need to be developed to characterize organic molecules using space-ready technologies. This thesis applies mass spectrometry techniques to these areas of research, including inferring planetary redox conditions, simulating post-impact vapor plumes for organic molecule production, and characterizing the size, structure, sequence, and mass of chemical compounds. The following sections provide a summary of the key conclusions and implications of each chapter.

- **Chapter 2:** The oxygen fugacity of magmatic environments, by extension the near surface redox environment in distant past, could be inferred from a single shot trace element analysis of zircon grains via laser ablation inductively coupled plasma mass spectrometry. Yet, the reliability of inferred redox values requires detailed petrographic analysis and whole rock geochemistry to constrain their growth and diagenesis history, and crystallization environment. Alternatively, Chapter 2 used a data-driven approach to look at the shared trace

element systematics of 2173 zircon analyses that compiled from 31 independent sources. The proposed filter scheme facilitates the selections of mantle-equilibrated zircon and provides quick insights about redox states using trace element systematics alone (P, Ti, Y, Nb, REE, Hf, Th, U), and therefore representing promising means to infer environmental conditions via geochemical measurements in near-term space missions.

- An important lesson learned from Chapter 2 is the power of harnessing big datasets to simplify conventional procedures and incorporate data-driven insights into practical implementations of geochemical research. Zircon is just one example that embodies such idea; similarly, I can foresee the potential for estimating temperature and pressure conditions through trace element systematics of garnet, a different mineral phase, as a simpler approach to constrain garnet-biotite thermometry and garnet barometry in metamorphic rocks. This is not to say that comprehensive geochemical investigation should be or will be replaced because thorough analysis requires time and attention to detail. However, it is important to recognize the invaluable merits of a data-driven approach in identifying, extracting, and maximizing information about unknown planetary environments when the available analytical techniques are not as comprehensive as desired. I believe that big data analytics will play a critical role in future space exploration by reducing the high dimensionality of expert knowledge into simple sets of measurements that are feasible, executable, and interpretable using space-ready technologies. I think statistical insights and solid field

knowledge are of equal significance in proposing new methods to characterize planetary environments with high efficiency and reliability.

- Areas of research follow by Chapter 2 includes but not limiting to:
    - Apply this method to study zircon in lunar rocks.
    - Use similar data-driven approach to infer temperature and pressure condition using trace element systematics of garnet, to infer redox conditions using trace element systematics (such as Zn/Fe) in basaltic rocks, and etc.

- **Chapter 3:** Hypervelocity impacts (HVIs) have long been considered catastrophic events that can destroy living systems. However, laboratory experiments of HVIs have suggested that they could have facilitated the transition from the abiotic to the biotic world. Chapter 3 investigates the capability of producing new molecules via impact processing on carbonates and nitrogen salts through high-power laser and projectile bombardment of C60+ onto N-rich substrates in the laboratory. Empirical observations, along with numeric estimates, provide constraints on the yield of organic production via crater-forming events on the surface of Earth, Mars, and Ceres.
    - A key takeaway from Chapter 3 is the coevolution of planets and organic compounds. Chapter 3 reveals that the abundance of organic compounds that can be generated via crater-forming impacts heavily relies on the chemical composition of planetary surfaces. Furthermore, the chances of these simple precursors to further evolve are governed by the

physicochemical conditions, such as the presence of oceans. Therefore, the abiotic synthesis and prebiotic evolution of organic molecules are intimately tied to the evolving physical and chemical conditions of planetary environments, including the redox state, temperature, depth of water, etc. Since the specifics of environmental conditions are heterogeneous across the planetary surface and changing with time, interpreting a specific life-detection measurement requires rigorous interdisciplinary collaboration of researchers, including planetary science, astronomy and astrophysics, chemistry, biology, and geoscience. I believe such collaborative work is necessary to build representative computational and laboratory simulations to adequately investigate the complex relationship between an evolving planet and the emergence of life.

- Areas of research follow by Chapter 3 includes but not limiting to:
    - increase the complexity of the substrate in simulation experiments to better represent the composition of planetary surfaces on Earth and targeted planetary bodies.
    - extend the duration of simulation experiments to investigate the molecular evolution of impact-generated molecules in simulated planetary environments.

- **Chapter 4** incorporated silicon nanoparticles (an alternative to traditional organic matrices) to detect and sequence 3-mer/4-mer peptides via in source decay using the CORALS prototype instrument. This study showcases the

analytical capabilities of spaceflight prototype CORALS to detect and sequence putative peptide biosignatures of psychrophiles for *in situ* investigation of ocean worlds. Even with the simplest sample preparation method, a single spectrum collected from the CORALS prototype reveals: (1) various forms of molecular ion adducts across a wide mass range, providing for robust identification and cross-validation to the chemical composition of molecular ions with high mass accuracy measurements, and; (2) diagnostic peptide fragments at a full sequence coverage adequate to elucidate the monomer sequence in samples with silicon nanoparticles added; (3) improved resolution and mass accuracy due to reduction of metastable decay with silicon nanoparticles.

- A critical implication of Chapter 4 is the demonstration of using silicon nanoparticles as an alternative to conventional organic matrices in LDMS, and thus opening the door for the use of inorganic matrices that are incompatible with planetary protection measures and reduced risks for forward contamination for missions to astrobiological destinations. Based on findings present here, a notional concept of operation for astrobiology landed mission to a cryogenic ocean world might include sample acquisition, deposition and sublimation of icy samples onto a suite of sample plates preloaded with a variety of inorganic matrices. Other hyphenated technologies such as capillary electrophoresis, could be used prior to sample deposition to enhance the throughput of analysis and enhance the sensitivity of measurements.

- Areas of research follow by Chapter 4 includes but not limiting to:
    - Explore the effects of various inorganic matrices (such as metalic nanoparticles, silicon-based nanomaterials, etc.) and their impacts at enhancing the ionization efficiency and sensitivity of particular molecular class (e.g., phospholipids, fatty acids, amino acids) in positive or negative ion detection mode.
    - Build a database of LDMS spectra of different class of molecules and summarize their molecular adduct formations and diagnostic fragments in the mass spectra for molecular identification of similar ions in future LDMS analysis.
    - Perform statistical analysis and establish evaluation scheme on the reproducibility (i.e., number of spectra of molecular adducts) and sensitivity (i.e., the S/N of molecular adducts at various instrumental operation settings) of spectra to enable autonomous instrument calibration and optimization, on-board data acquisition, data analysis, and data processing during future space operation.

- **Chapter 5** developed data processing workflow to enable simultaneous measurements of mass and collision cross section (CCS) by processing transient signals via short-time Fourier Transform (StFFT). We identified threshold conditions to derive reproducible and precise CCS measurements. Though the accuracies of CCS measurements remain insufficient to enable standalone identification of chemical compounds, the established m/z-CCS distribution

was found to be efficient at distinguishing molecular class, structure, and composition in sample mixture.

- A critical insight revealed in Chapter 5 is the dynamic ion motions inside the Orbitrap mass analyzer. These complicated ion movements include ion-neutral collision and local and global space charge effects, all of which lead to different rates of decay in transient signals and induce deviation from ideal ion orbital frequencies (or their m/z). Without a clear understanding of the fundamental ion motions inside the Orbitrap, it is challenging to realize the intrinsic errors in peak intensities and mass accuracies in mass spectra generated from Orbitrap MS, and it becomes even harder to achieve quantification via label-free Orbitrap MS and improve the frontier of analytical performance. I believe that the transient signal is the X-ray photography that documents the nanoscopic ion motions inside the Orbitrap mass analyzer, and various data processing techniques (including FFT, StFFT, Hilbert transform) are effective tools to reveal and study these intricate ion motions. I strongly encourage an increase in research focus on the ion motion inside the Orbitrap and fundamental principles of FT processing. I believe these research efforts will lead to significant development in analytical capabilities of Orbitrap MS (such as CCS measurements from a single scan) and advance the quantitative and qualitative capability of Orbitrap MS to a new level.
- Areas of research follow by Chapter 5 includes but not limiting to:

- The signal decay of ions with m/z 50 – 2000 due to ion-neutral collision inside the mass analyzer and their impacts in intra-scan fractionation
- The signal decay of low intensity ions due to the presence of a high intensity ion in proximity (such as many isotopic systems), modulated decay profile of low abundance ions, and the impacts on the quantitative capability

Publications

Published papers:

- **Ni, Z.,** & Ray, S. (2023). A miniaturized laser desorption Orbitrap mass spectrometer for biosignature detection. *Nature Astronomy*, 1–2. https://doi.org/10.1038/s41550-022-01879-6

- Arevalo, R., Willhite, L., Bardyn, A., **Ni, Z.,** Ray, S., Southard, A., Danell, R., Grubisic, A., Gundersen, C., Minasola, N., Yu, A., Fahey, M., Hernandez, E., Briois, C., Thirkell, L., Colin, F., & Makarov, A. (2023). Laser desorption mass spectrometry with an Orbitrap analyser for in situ astrobiology. *Nature Astronomy*, 1–7. https://doi.org/10.1038/s41550-022-01866-x

- Theiling, B. P., Chou, L., Da Poian, V., Battler, M., Raimalwala, K., Arevalo, R., Neveu, M., **Ni, Z.,** Graham, H., Elsila, J., & Thompson, B. (2022). Science Autonomy for Ocean Worlds Astrobiology: A Perspective. *Astrobiology*, *22*(8), 901–913. https://doi.org/10.1089/ast.2021.0062

- Willhite, L., **Ni, Z.,** Arevalo, R., Bardyn, A., Gundersen, C., Minasola, N., Southard, A., Briois, C., Thirkell, L., Colin, F., Grubisic, A., Fahey, M., Yu, A., Hernandez, E., Ersahin, A., Danell, R., & Makarov, A. (2021). CORALS: A Laser Desorption/Ablation Orbitrap Mass Spectrometer for In Situ Exploration of Europa. *2021 IEEE Aerospace Conference (50100)*, 1–13. https://doi.org/10.1109/AERO50100.2021.9438221

- Arevalo, R., **Ni, Z.,** & Danell, R. M. (2020). Mass spectrometry and planetary exploration: A brief review and future projection. *Journal of Mass Spectrometry*, *55*(1), e4454. https://doi.org/10.1002/jms.4454

- **Ni, Z.,** Arevalo, R., Piccoli, P., & Reno, B. L. (2020). A Novel Approach to Identifying Mantle-Equilibrated Zircon by Using Trace Element Chemistry. *Geochemistry, Geophysics, Geosystems*, *21*(11), e2020GC009230. https://doi.org/10.1029/2020GC009230

Submitted papers:

- **Ni, Z.**, Arevalo Jr., R., Bardyn, A., Willhite, L., Ray, S., Southard, A., Danell, R., Jacob Graham, R., Li, X., Chou, L., Briois, C., Thirkell, L., Makarov, A.,

Brinckerhoff, W. B., Eigenbrode, J., Junge, K., Nunn, B. (*In Print*) Detection of Short Peptide Biosignatures of Psychrophiles via Laser Desorption Mass Spectrometry. *Astrobiology*.

- **Ni, Z.**, Farcy, B., Arevalo Jr., R., Eller, M., Pinnick, V. T., Schweikert, E. A., Brinckerhoff, W. B. (*Under review*) Production of Organic Precursors via Crater-forming impacts and Its Implication to Early Earth, Mars, and Ceres. *Astrobiology*

Papers in preparation:

- **Ni, Z.,** Arevalo, R. Simultaneous measurement of mass and collision cross section via Orbitrap mass spectrometry.

Appendice A: Supplementary material to Chapter 2

Supplementary Material to Manuscript

A Novel Approach to Identifying Mantle-Equilibrated Zircon by Using Trace Element Chemistry

Ziqin Ni[1], Ricardo Arevalo Jr.[1], Philip Piccoli[1], Barry Reno[2]

[1]University of Maryland, College Park, MD, 21401
[2]Northern Territory Geological Survey, Darwin, NT, 0800

Corresponding author: Ziqin Ni (zni@umd.edu)

Analytical method to measure zircon (n = 76) from three metasedimentary rocks (524906, 524907, and 524994) that sampled at Kangerdluarssuk Fjord, southern West Greenland:

Zircon U–Pb isotope data were collected on zircon separates, embedded in 2 cm epoxy blocks, by laser ablation inductively coupled plasma mass spectrometry (LA–ICP–MS) at the Geological Survey of Denmark and Greenland (GEUS) following the methods of Frei and Gerdes (2009) and Frei *et al*. (2006), and using analytical conditions described in detail in Dyck *et al*. (2015). A New Wave Research UP213 frequency-quintupled solid state Nd:YAG laser system, employing a two-volume cell technology, was coupled to a Thermo-Fisher Scientific ELEMENT 2 double-focusing single-collector magnetic sector-field ICP–MS. Mass resolution was 300. U–Pb isotope data in zircon were acquired from single spot analysis using a spot diameter of 25 µm at a repetition rate of 10 Hz and a nominal laser fluence of 15–16 J/cm^2.

Trace element compositions in zircon were performed using a 40 µm spot diameter at a repetition rate of 10 Hz with 60% energy and a nominal laser fluence of ~6 J/cm^2. Trace element analysis locations and U–Pb analysis locations were adjacent to each other, within the same backscatter electron zone in each analysed zircon. Trace element concentrations was normalised to a stoichiometric zircon ^{30}Si of 29.75 wt% (Deer *et al*., 1992). Calibration was carried out relative to the NIST 610 glass standard (Pearce *et al*., 1997), which was analysed in brackets of 2 standards and 5 samples.

Appendice B: Supplementary material to Chapter 3

Supplementary Material to Manuscript

Production of Organic Precursors via Crater-forming impacts and Its Implication to Early Earth, Mars, and Ceres

Ziqin Ni[1*], Benjamin Farcy[1], Ricardo Arevalo Jr.[1], Michael Eller[2], Veronica T. Pinnick[3], Emile A. Schweikert[4], William B. Brinckerhoff[3]

[1]University of Maryland, College Park, MD, 20742.
[2]California State University Northridge, CA, 91330.
[3]NASA Goddard Space Flight Center, Greenbelt, MD, 20771.
[4]Texas A&M University, College Station, TX, 77843.

Email: zni@umd.edu

Content

B1. Experimental setup of laser simulation experiment

The laser irradiated the center of the target at a 30° incident angle relative to the sample normal with an elliptical beam profile of approximately 0.8 mm (minor axis) by 1.4 mm (major axis), equating to an area of 0.88 mm^2. Ions formed in the plasma (including recombination products) were extracted and collimated by a simple set of ion optics and detected by a quadrupole mass spectrometer (QMS, Hiden Analytics PSM003). A Bessel box energy filter at the QMS inlet allowed scanning of individual ion energies. This inlet accepted ions emitted from the 1-mm diameter orifice in line with the cylindrical sample holder held at electrical ground. QMS scans were triggered to the laser Q-switch to synchronize ion collection. All experiments were performed at 10^{-5} Pa of chamber pressure to attenuate interferences from ambient CO_2 and N_2. All scans were performed in negative ion mode; thus, the formation of HCN was detected as deprotonated CN^-.

Table B1. Raw dataset to laser simulation experiments

Mass station	NO3				NH4				
	20%	30%	40%	50%	20%	30%	40%	50%	
10	174300	170700	182350	259350	174200	114750	61550	109900	
11	189950	190100	197600	304200	172600	122550	60400	117700	
12	236100	242800	223650	338850	233500	156800	64100	156050	
13	307450	267350	279350	385200	327000	149850	88700	131650	
14	347000	347750	347650	470300	344600	156300	74850	137850	
15	575650	621300	552050	738750	709450	248000	112650	210700	
16	1235500	1206550	1134450	1188950	1651650	1249250	574800	1342850	
17	985200	865700	748900	936200	1122550	671750	366650	724200	
18	272250	227150	246150	321650	314400	146600	71500	114050	
19	218850	214750	197100	295400	637100	178900	109000	136800	
20	143750	134850	156950	185300	156250	122400	66850	92200	
21	139050	127750	150600	166950	121600	119800	75950	82650	
22	137050	129350	147700	152550	133650	116000	71050	79450	
23	155300	140850	160100	193350	131150	117050	63950	80600	
24	167800	169750	154900	231100	226200	129050	77700	98700	
25	199750	180750	155200	341400	318750	130500	95950	97400	
26	218350	237100	208700	648450	326000	135050	129500	229750	$^{12}C^{14}N^-$
27	308300	405100	257700	224500	482150	214250	202000	80400	$^{13}C^{14}N^-$ /$^{12}C^{15}N^-$
28	370900	142500	152000	149750	575900	120950	79250	65700	$^{13}C^{15}N^-$
29	150100	123850	152100	147350	136700	119500	77450	69300	
30	142550	124550	165100	169300	147250	119650	73200	70550	
31	145650	146550	154950	151000	182350	127400	78350	80200	
32	212050	218550	232650	249500	356150	166550	112000	118350	
33	163050	155950	168300	169550	445350	154550	112050	110900	
34	156550	157650	157700	161600	1261700	288650	242650	218350	
35	263850	389150	195150	315750	2145150	2270100	1844600	2180650	
36	144850	148900	155600	150900	884900	199450	172100	134000	
37	178450	207750	170350	171400	1978600	1346150	980350	1247800	
38	138750	121950	157050	130000	245900	115050	77300	81950	
39	133900	120750	160400	125100	130250	101700	75900	73300	
40	141100	126200	161450	137800	143950	100950	69800	74600	
41	146550	122300	163500	167350	140200	99600	69300	80000	
42	146650	144100	171600	304750	184300	107250	83350	118550	CNO^-
43	191300	184150	185850	137000	243300	120700	100050	70700	$^{13}C^{14}NO^-$ /$^{12}C^{15}NO^-$
44	216650	130150	161750	129050	291950	102300	66650	72800	$^{13}C^{15}NO^-$
45	149100	133200	181600	161950	131050	104800	70800	89150	
46	160700	131650	365550	305050	182550	100400	77150	80100	NO_2^-
47	197650	208100	168900	132150	256200	113250	67650	73150	$^{15}NO_2^-$
48	135650	122350	157750	119450	149700	93400	74050	73750	Background
49	136750	125150	156900	118400	128400	99200	68350	71750	Background
50	140550	119550	151250	123000	133200	94700	68500	69800	Background

Table B2. Linear regression models and predicted number of CN^- molecules at various wt% N salts in laser simulation experiments.

Linear regression model		
model: y=kx	$NaNO_3$	NH_4Cl
Slope (k)	19897	6606
Standard error	871	448
R^2	0.9962	0.9909
Predicted CN^- molecules at various wt% N salts		
% N salt	$NaNO_3$	NH_4Cl
5%	995	330
10%	1990	661
15%	2985	991
20%	3979	1321
25%	4974	1651
30%	5969	1982
35%	6964	2312
40%	7959	2642
45%	8954	2973
50%	9949	3303

Table B3. A list of assumption and parameters about laser micro-processing processes to estimate the volume and longevity of the laser-generated plume and the proportion of recombined CN^- that was detected by the instrument.

Laser parameter			**Justification**
Spot size	m^2	9×10^{-7}	Measured
Laser pit depth	m	1×10^{-7}	1
Laser pit volume	m^3	9×10^{-14}	Calculated
Laser out energy	J	0.07	Measured
Energy flux	J/cm^2	8.0	Calculated
Energy irradiance	W/cm^2	8×10^{8}	Calculated
Pulse duration	s	1×10^{-8}	Attribute
Substrate property			
Density of nitrate-carbonate mix	kg/m^3	2560	Attribute
Density of ammonium-carbonate mix	kg/m^3	2200	Attribute
Molar mass calcium carbonate	g/mol	100.0869	Attribute
Molar mass sodium nitrate	g/mol	84.9947	Attribute
Molar mass ammonium chloride	g/mol	53.491	Attribute
Assumption			
Mass fraction of vaporized materials	%	5	2
Atomization efficiency	%	1-10	3
Transmission + detector efficiency	%	1	4
Prediction for the volume and longevity of vapor plume			Calculated
Volume of vaporized materials	m^3	4×10^{-15}	Calculated
Radius of vapor plume	m	8×10^{-3}	5, calculated
Volume of vapor plume	m^3	2×10^{-6}	Calculated
Longevity of plume	s	1×10^{-6}	6

Justification for the parameters in Table B3:

1. High irradiance laser ablation (>10^{10} W/cm^2) typically removes the uppermost 10^{-7} m of the sample per laser shot due to an explosive boiling of the substrate materials (Bulgakova & Bulgakov, 2001; Lu et al., 2002).

2. We assume 5 % of ejected masses would be vaporized for a carbonate-rich targe (Schultz, 1996).

3. We assume a range of atomization efficiency of vaporized materials from 1-10 %.

4. The detection efficiency of recombined ions was assumed to be on the order of 1%, meaning the other 99% of CN^- were not detected due to the limitation of ion optics in the setup of laser simulation experiment.

5. The vapor plume would expand at 10-15 times the size of impactor based on two-dimensional numerical simulations under near-vacuum condition (Melosh et al., 1993), thus the volume of vapor plume was estimated as the volume of a sphere (r_{plume}= 15$r_{impactor}$).

6. The explosion of mass ejection and generation of vapor plume have been observed to delay and could last from hundreds of nanoseconds to tens of microseconds in near-vacuum conditions at laser irradiance >10^{10} W/cm^2(LaHaye et al., 2013; Yoo et al., 2000). We assumed a t =1 μs as a medium value for C and N atoms from the target mixture to recombine in the laser-generated vapor plume in our experimental setup.

B3. Steps of calculation to estimate HCN production yields in planetary scale impact

We start our estimates of crater morphology and plume generation in the simplest scenario, in a planetary surface environment without the presence of atmospheres nor oceans. In this case, the mass and speed of the projectile does not decrease prior to the impact event, and thus the energy imparted to an impacted surface is equal to the kinetic energy of the impacting projectile,

$$E_{impact} = \frac{1}{2} m_{impactor} v_{impactor}^2. \quad \text{Eqn. (B2)}$$

The diameter of a transient crater created on the planetary surface (D_{tc}) can be estimated using

$$D_{tc} = a\left(\frac{\rho_{impactor}}{\rho_{target}}\right)^{\frac{1}{3}} D_{impactor}^{0.78} v_{impactor}^{0.44} g E_{impact}^{-0.22} \left(\sin\left(\theta_{impact}\right)\right)^{\frac{1}{3}}, \quad \text{Eqn. (B3)}$$

where $\rho_{impactor}$ and ρ_{target} are the density of impactor and target respectively, $D_{impactor}$ is the diameter of impactor, g is the gravitational acceleration of the planetary body, and θ_{impact} is the impact angle in degrees (Schmidt and Housen, 1987). A proportional constant, a, is set to 1.2 for a target surface without water. The depth of the transient crater (d_{tc}) that is below the original ground can be estimated from D_{tc} based on observations by Dence et al., (1977),

$$d_{tc} = D_{tc}/2\sqrt{2}. \quad \text{Eqn. (B4)}$$

The volume of a transient crater (V_{tc}) can be estimated from D_{tc},

$$V_{tc} = \pi D_{tc}^3/(16\sqrt{2}). \quad \text{Eqn. (B5)}$$

The vaporized mass fraction of a transient crater was measured to be 5% and 10% during a crater-forming impact event on a carbonate-rich and ice-rich target,

respectively (Schultz, 1996). Similar to laser simulation experiments, about 1-10% of vaporized materials would be atomized for molecular recombination in the post-impact plumes. The vapor plume would expand at 10-15 times the size of impactor based on two-dimensional numerical simulations under near-vacuum condition (Melosh et al., 1993), thus the volume of vapor plume was estimated as the volume of a sphere (r_{plume}= 15$r_{impactor}$). Because the crater-forming impact plume would be generated at a very high pressure (>100 GPa) and temperature (>10,000 K), it would retain its ionizing state, absorb radiation at all wavelengths (Glassstone and Dolan, 1977). The ionizing state comes to an end until the plume expands and cools to reach a critical temperature ($T_{transparency}$), at which matters start to recombine and condense. The typical $T_{transparency}$ of silicate vapor is about 6,000 K, and 3,000 K for atmospheric, water-rich vapor (Melosh et al., 1993). The radius of the plume at $T_{transparency}$ (R_{maxr}) is about 15 times the impactor diameter based on numerical simulation of vapor plume expansion (Melosh et al., 1993; Nemtchinov et al., 1998). If we assume the vapor plume were expanding at the approximately the same velocity as $v_{impactor}$, the duration of the plume remaining in its ionizing state ($\tau_{ionization}$) would be

$$\tau_{ionization} = \frac{R_{maxr}}{v_{impactor}}. \quad \text{Eqn. (B6)}$$

We predicted the longevity of vapor plume by calculating the time it took to radiate the absorbed heat (i.e., 2% of impact energy) from its transparency temperature (i.e., 6,000K for atmosphere-free silicate vapor (Melosh et al., 1993)) following the Planck's law,

$$\tau_{\text{plume}} = \frac{\eta E_{impact}}{2\pi R_{maxr}^2 \sigma T_{transparency}^4} \quad \text{Eqn. (B7)}$$

where σ is the Stefan-Bolzmann constant 5.67×10^{-8} $Wm^{-2}K^{-4}$.

B2. ***Additional considerations: the roles of atmosphere and ocean***

The mass and speed of the projectile would decrease prior to the impact event if an atmosphere and/or ocean layer was present on the planetary target, thereby affecting the dynamics of the plume formation process. However, atmospheric entry has negligible influence on the shape, energy, or momentum of impactors with masses in excess of the atmosphere displaced during penetration. This momentum of an impactor of this size traveling at 18 km/s would be much larger than the amount of atmospheric drag exerted upon its meteoritic entry, thus leading to minimal influences in the shape and energy of projectiles impacting on the surface of the early Earth.

In presence of atmospheres, the impact-generated plume would theoretically interact with the ambient atmosphere, changing the plume dynamics including the internal energy of the plume, the rate of plume expansion and cooling (Sugita & Schultz, 2003b). The general effect of the background gas is reported to be the spatial confinement and slowing down of the expanding plume. Moreover, at high laser irradiance ($>10^9$ W/cm^2), the ionization of vapor and/or the background gas may become important and should be taken into account in the modeling. The ambient atmospheres that entrained into the impact plumes would be shocked, heated, and ionized to contribute an additional amount of carbon and nitrogen atoms for molecular recombination. To make a first-order approximation, the $T_{transparency}$ was set to be 3000 K for plume expansion and cooling in presence of atmospheres, which extends the τ_{plume} for a prolonged molecular synthesis (G. S. Collins et al., 2005). The carbon and nitrogen gas molecules

in the atmosphere (in the form of CO_2 and N_2 for simplicity) could contribute an additional amount of carbon and nitrogen atoms for molecular synthesis in the post-impact plume. Given the weight percentage of CO_2 and N_2 in the atmospheres ($C_{atmo,CO2}$ and $C_{atmo,N2}$), the carbon and nitrogen atoms sourced from the atmospheres could be estimated assuming a spherical post-impact plume with a radius R_{maxr}. If the projectile were to land in an ocean, on the other hand, the velocity of the impactor at the seafloor ($v_{impactor,seafloor}$) would be reduced due to drag:

$$v_{impactor,seafloor} = v_{impactor} \exp\left(\frac{-3\rho_{water}C_D(ocean)d_{water}}{2\rho_{impactor}D_{impactor}\,sin(\theta_{impact})}\right) \quad \text{Eqn. (B8)}$$

, where the $C_D(ocean)$ is the drag coefficient to account for the decrease in velocity due to the presence of ocean, 0.47 for spherical objects in water, and (d_{water}) is the depth of the ocean. The impact energy would be estimated using $v_{impactor,seafloor}$. The $T_{transparency}$ was set to be 3000 K for a water-rich vapor. The proportional constant a used to estimate D_{tc} in Eqn (3) was set to 1.4 for a target that has icy layers or oceans. Because water layer in the target surface has been shown to accelerate the crater-forming process due to the generation of acoustic wave induces a higher pressure and temperature regime close to substrate (Chemin et al., 2022; Zhu et al., 2001). By changing a from 1.2 to 1.4, the transient crater could have a maximum increase by 60 vol.%.

Table B5. A list of assumptions and parameters for estimating HCN production via crater-forming impacts on planetary surfaces

	Parameters	Unit	**Earth**	**Mars**	**Ceres**	**Ref.**
Mass of impactor	$m_{impactor}$	kg	10^{16}-10^{20}	10^{14}-10^{18}	10^{7}-10^{15}	1,2,3

Density of impactor	$\rho_{impactor}$	kg/m^3	3000	3000	3000	4
Speed of impactor	$v_{impactor}$	km/s	18	10	5	5
Density of target	ρ_{target}	kg/m^3	3000	2600	1500	-
Gravitational acceleration	g	m/s^2	9.8	3.7	0.28	Attribute
Impact angle	θ_{impact}	degree	45	45	45	7
Conc. of carbon in impactor	$C_{i,carbon}$	wt.%	2.5	2.5	2.5	4
Conc. of nitrogen in impactor	$C_{i,nitrogen}$	wt.%	0.25	0.25	0.25	4
Conc. of carbon in substrate	$C_{s,carbon}$	wt.%	1-10	5	10	2,6,9
Conc. of nitrogen in substrate	$C_{s,nitrogen}$	wt.%	0.001-0.1	0.03	10	2,6,9
Proportional constant	a	-	1.4	1.4	1.4	18
Depth of ocean layer	d_{ocean}	m	<500	<500	-	8
Depth of subsurface icy layer	d_{ice}	m	-	-	5000	9
Conversion to vapor energy		%	2	2	2	17
Stefan-Boltzmann constant	σ	Wm^{-2}K^{-4}	5.67E-08	5.67E-08	5.67E-08	Attribute
Transparency temperature	T_{maxr}	K	3000	3000	3000	13
Vaporization efficiency	η_{vap}	vol.%	5	5	10	10
Atomization efficiency	η_{atom}	vol.%	1-10	1-10	1-10	16
Neutralization efficiency	η_{neu}	vol.%	50	50	50	16
Density of atmosphere	$\rho_{atmosphere}$	kg/m^3	1.3	0.02	-	2
Conc. of CO_2 in atmosphere	$C_{atmo,CO2}$	wt.%	0.41	95	-	12
Conc. of N_2 in atmosphere	$C_{atmo,N2}$	wt.%	78	0.28	-	12
CN production rate			1.5e+7	1.5e+7	1.5e+7	15

Justification for parameters for Table B5:

1. For early Earth scenario, the mass range of carbonaceous projectiles considered in this study is 10^{16}-10^{20}kg (Anbar et al., 2001). We consider the threshold size to have minimum deacceleration upon atmospheric entry and make sufficient ejecta on a shallow marine is 10^{16} kg, approximately 20 km diameter impactor. The maximum mass of projectile is 10^{20} kg, approximately 400 km diameter impactor.

2. Mars has a lower gravity and smaller cross sections compared to Earth, attracting a range of less massive carbonaceous projectiles (10^{14}-10^{17} kg) (Strom et al., 2005). The Martian surface was estimated to have 2-5 wt.% carbonates (Bandfield et al., 2003) and ~0.03 wt.% N salt (Stern et al., 2015), as indicated by the chemical measurements of Martian soil from the Mars Science Laboratory (MSL) rover. The surface density of Martian surface was assumed to be 1500 kg/m^3.
3. Ceres' craters are most typically in the 1 to 100 km diameter range, and likely created by carbonaceous impactors of 10^7-10^{15} kg (Hiesinger et al., 2016).
4. We assumed an average density of a carbonaceous projectile is 3000 kg/m^3. The average carbon content in a carbonaceous projectile is 2.5 with a C/N ratio measured to be 10.
5. The average impact speed on surface of the Earth is 18 kg/m^2 (G. S. Collins et al., 2005). Mars and Ceres both have lower gravity and smaller cross sections compared to early Earth, which led to reduction in average impactor speed to 10 and 5 km/s, respectively (Strom et al., 2005).
6. Due to chemical weathering of ultramafic crust and impact ejecta in a CO_2-rich atmosphere, the Hadean crust was likely to harness ~10^{19}-10^{21} moles of carbonates to sustain a global carbon cycle (Sleep & Zahnle, 2001), equivalent to 0.1-10 wt.% of carbonate minerals on Hadean surface. Abiotic fixation of N_2 gas in atmosphere would have produced ~10^9-10^{12} moles of N salts per year ((Brandes et al., 1998) and references therein), leaving ~0.001-1 wt.% of N salts on the crustal surface over an 100 Myr interval in Hadean Earth.

7. The most probable impact angle was assumed to be 45 degree.
8. Frequent large asteroid impacts on Hadean Earth would have evaporated majority of ocean water in the global scale (Maher & Stevenson, 1988; Sleep et al., 1989). We explored a range of 0-500 m in the calculation.
9. Ceres has been observed to have NH_4-phyllosilicates and Mg,Ca-carbonates, which are estimated to be ~10% for both (Ammannito et al., 2016; Carrozzo et al., 2018; De Sanctis & Ammannito, 2021). Ceres is also observed to have a subsurface icy shell that is modelled be at < 5 km depth from its surface (Bland et al., 2016; Sizemore et al., 2019).
10. The vaporized mass fraction of a transient crater was measured to be 5% and 10% during a crater-forming impact event on a carbonate-rich and ice-rich target, respectively (Schultz, 1996).
11. We estimated flux of Martian impactors using the mass-frequency curve of the terrestrial impactors (Anbar et al., 2001) and assuming an impact flux ratio between Earth/Mars equivalent to 10 (Bottke et al., 2010).
12. Our estimates of HCN production is based on the assumption of a stable early Martian atmosphere that is cold enough to sustain surface water and similar to present Mars, about 10^{16} kg and is dominated by 95 wt. % CO_2 and 0.28 wt.% N_2 (Mahaffy et al., 2013; Segura et al., 2012).
13. To make a first-order approximation, the $T_{transparency}$ was set to be 3000 K for plume expansion and cooling in presence of atmospheres (G. S. Collins et al., 2005).

14. The radius of the plume at $T_{transparency}$ (R_{maxr}) is about 15 times the impactor diameter based on numerical simulation of vapor plume expansion in vacuum (Melosh et al., 1993)
15. The CN production rate was derived from the laser simulation experiment from our study.
16. We assume a range of atomization efficiency of vaporized materials from 1-10 % and neutralization efficiency 50%.
17. The internal energy of the vapor plume was measured to be 2-5 % of the impact energy on a carbonate or icy target at the most probable impact angle 45° (Schultz, 1996; Svetsov & Shuvalov, 2019); thus we used 2% as a lower bound.
18. A proportional constant, a, is set to 1.2 or 1.4 for a target surface with water or without water, respectively, to estimate the crater morphology on planetary surface (G. Collins et al., 2012; G. S. Collins et al., 2005; Schmidt & Housen, 1987).

Appendice C: Supplementary material to Chapter 4

Supplementary Material to Manuscript

Detection of Short Peptides as Putative Biosignatures of Psychrophiles via Laser Desorption Mass Spectrometry

Ziqin Ni[1*], Benjamin Farcy[1], Ricardo Arevalo Jr.[1], Michael Eller[2], Veronica T. Pinnick[3], Emile A. Schweikert[4], William B. Brinckerhoff[3]

[1]University of Maryland, College Park, MD, 20742.
[2]California State University Northridge, CA, 91330.
[3]NASA Goddard Space Flight Center, Greenbelt, MD, 20771.
[4]Texas A&M University, College Station, TX, 77843.

Email: zni@umd.edu

The characteristics of identified peaks are summarized in the table C1 below. We found asymmetrical peak shape associated with organic-metal ion adducts (e.g., M+K, M+Na) and peptide fragments, especially in LDMS measurement without silicon nanoparticles. However, similar phenomena were not observed with inorganic peaks (e.g., Cs^+, Na^+, Al^+) at high signal to noise ratio (S/N). The following supplementary notes will describe some of the observation associated with LDMS measurements and speculate how S/N and ion generation in the laser source could affect the spectral performance (i.e., mass resolution and mass accuracy) of observed peaks.

Signal to noise ratio and spectral performance

We found that peaks with a low signal to noise ratio (S/N) show greater errors in mass accuracy (**Figure C1 a**) than expected (Makarov et al., 2006). Peak shapes deteriorate at low S/N while random signal spikes from the elevated noise-floor interfere with the peak, alter the peak symmetry, shift the center mass, distort the peak maxima, and broaden the peak. In addition, difference in peak decay constants comparing to calibrants as well as perturbations of electronics result in similar effects. Such non-ideal peak shape induces errors in estimates of center mass, peak maxima, and peak width via peak-fitting procedures. Compared to a three-point parabolic peak fitting procedure used in commercial Fourier-transform mass spectrometry, this study found that gaussian peak fitting of the entire peak generates better estimates of peak maxima because peak asymmetry could be corrected by incorporating the base of the peak.

Ion generation and spectral performance

We found peaks of high theoretical mass associated with low mass accuracy and reduced mass resolution (**Figure C1 c&d**). These higher mass peaks were typically identified as peptide fragments and organic metal adducts (e.g., M+K, M+Na) based on their stoichiometries. We speculate such phenomenon was due to the unique ion generation mechanisms in the laser source and the specific setup of CORALS instrument.

Ions of higher masses and organic fragments were found to have lower mass resolution in general. The high-power laser and thermal desorption are critical to desorb higher

mass species but also induce significant metastable decay, especially in LDMS measurements without silicon nanoparticles before they enter the orbitrap analyzer. Metastable decay helps to generate diagnostic sequence ions for peptide *de novo* sequencing, but the dissociation of metastable ions within the analyzer accelerates the decay of transient signals. FFT of a decaying transient will return a frequency peak with lower peak amplitude and a wider peak width (Mathur & O'Connor, 2009). Therefore, analyte (ions) that are susceptible to increasing amount of metastable decay are observed with lower mass resolution. This could be later used for "quality control" of spectra and rejection of unreliable peaks.

Peaks of ion adducts (e.g., M+K, M+Na, CsI, Cs_2I) were found to have up to 50 ppm errors in mass measurement. CORALS instrument applied a constant voltage on the sample plate to propel ions generated in the source into the mass analyzer. If ions were created off the plate (i.e., generated in the plume), the newly formed ion moieties will be propelled into slightly different orbiting trajectory in the Orbitrap analyzer due to reduced kinetic energies after acceleration, introducing errors in oscillating frequency and mass measurement. Therefore, analyte ions or adducts that are formed in the plume might be susceptible to a higher mass error.

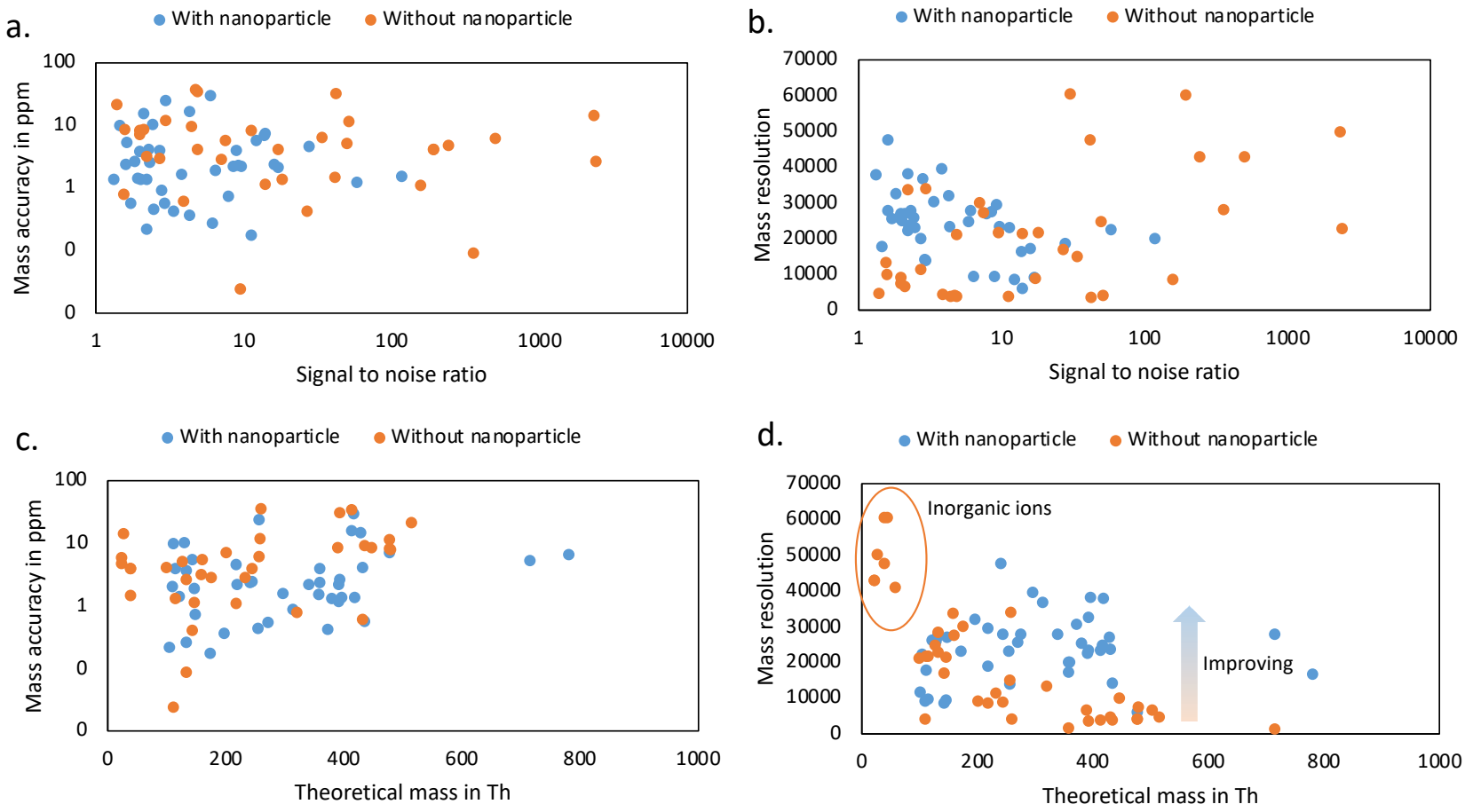


Figure C1. Scatter plot of mass accuracy, mass resolution, theoretical mass, and signal to noise ratio of all identified peaks in spectra presented in this study, using statistics summarized in table C1.

Table C1. Characteristics of peaks identified in the spectra shown in this paper.

SHD without Silicon Nanoparticles (1 scans)									
Gaussian Peak Fit				Peak Assignment				Performance	
Center Max in Da	Max Height in a.u.	FWHM	S/N	Peptide	Ion type	Chemical stochiometry	Theoretical mass in Da	Mass accuracy (ppm)	Mass resolution
22.9889	0.01058	5.36E-04	243	SHD	Na	Na	22.9898	36	42909
26.9803	0.10105	5.39E-04	2323	SHD	Al	Al	26.9815	46	50101
38.9624	0.00179	8.16E-04	41	SHD	K	K	38.9637	33	47759
110.0745	1.16E-04	0.02507	3	SHD	H	C5 H8 N3	110.0713	-29	4391
126.8998	2.15E-03	0.00512	49	SHD	I	I	126.9045	37	24785
132.9009	1.05E-01	0.00577	2411	SHD	Cs	Cs	132.9055	34	23033
259.8112	2.05E-04	0.06039	5	SHD	CsI	Cs I	259.8099	-5	4302
358.1500	8.97E-04	0.19196	21	SHD	MH	C13 H20 N5 O7	358.1357	-40	1866
392.7154	1.82E-03	0.10607	42	SHD	Cs2I	Cs2 I	392.7154	0	3702
715.2807	1.58E-04	0.50632	4	SHD	2M+H	C26 H39 N10 O14	715.2641	-23	1413
SHD with Silicon Nanoparticles (1 scan)									
Gaussian Peak Fit				Peak Assignment				Performance	
Center Max in Da	Max Height in a.u.	FWHM	S/N	Peptide	Ion type	Chemical stochiometry	Theoretical mass in Da	Mass accuracy (ppm)	Mass resolution
105.0664	9.19E-05	0.00467	2	SHD	c1	C3 H9 N2 O2	105.0664	0	22498
110.0716	7.11E-04	0.01198	17	SHD	H	C5 H8 N3	110.0718	-2	9188
122.0716	8.03E-05	0.00467	2	SHD	H	C6 H8 N3	122.0718	-1	26140
197.1040	1.80E-04	0.00613	4	SHD	a2	C8 H13 N4 O2	197.1039	0	32154
242.1259	6.68E-05	0.00508	2	SHD	c2	C9 H16 N5 O3	242.1253	2	47663
255.0856	1.03E-04	0.01101	2	SHD	z2	C10 H13 N3 O5	255.0855	0	23169
271.1044	7.19E-05	0.01055	2	SHD	y2	C10 H15 N4 O5	271.1042	1	25697
297.0840	1.59E-04	0.00747	4	SHD	x2	C11 H13 N4 O6	297.0835	2	39770
314.1467	1.17E-04	0.00854	3	SHD	MH-CO2	C12 H20 N5 O5	314.1464	1	36785
340.1265	3.54E-04	0.01223	8	SHD	MH-H2O	C13 H18 N5 O6	340.1257	2	27811
358.1368	4.92E-03	0.01779	117	SHD	MH	C13 H20 N5 O7	358.1363	2	20131
359.1400	6.70E-04	0.02073	16	SHD	MH	Isotopplogue	359.1391828	2	17325
360.1420	1.13E-04	0.01794	3	SHD	MH	Isotopplogue	360.1405477	4	20075
380.1187	8.43E-05	0.015	2	SHD	M+Na	C13 H19 N5 O7 Na	380.1182235	1	25341
396.0927	9.26E-05	0.01036	2	SHD	M+K	C13 H19 N5 O7 K	396.09213	1	38233
715.2679	6.74E-05	0.02554	2	SHD	2M+H	C26 H39 N10 O14	715.2641235	5	28006

Table C1-continue

GDAE without Silicon Nanoparticles (1 scans)									
Gaussian Peak Fit				Peak Assignment				Performance	
Center Max in Da	Max Height in a.u.	FWHM	S/N	Peptide	Ion type	Chemical stochiometry	Theoretical mass in Da	Mass accuracy (ppm)	Mass resolution
22.9896	1.96E-02	0.00053	501	GDAE	Na	Na	22.9898	6	43038
38.9636	7.51E-03	0.00064	193	GDAE	K	K	38.9637	4	60430
44.0500	1.18E-03	0.00073	30	GDAE	C2 H6 N	C2 H6 N	44.0500	0	60578
132.9054	1.39E-02	0.00468	356	GDAE	Cs	Cs	132.9055	0	28399
413.1427	1.88E-04	0.10348	5	GDAE	M+Na	C14 H22 N4 O9 Na	413.1284	-35	3992
GDAE with Silicon Nanoparticles (1 scan)									
Gaussian Peak Fit				Peak Assignment				Performance	
Center Max in Da	Max Height in a.u.	FWHM	S/N	Peptide	Ion type	Chemical stochiometry	Theoretical mass in Da	Mass accuracy (ppm)	Mass resolution
130.0517	8.93E-05	0.0050	2	GDAE	y1-H2O	C5 H8 N O3	130.0504	10	25907
132.9055	2.26E-04	0.0048	6	GDAE	Cs	Cs	132.9055	0	27980
148.0611	2.89E-04	0.0054	8	GDAE	y1	C5 H10 N O4	148.0610	1	27217
173.0557	4.16E-04	0.0074	11	GDAE	b2	C6 H9 N2 O4	173.0557	0	23291
219.0986	3.38E-04	0.0074	9	GDAE	y2	C8 H15 N2 O5	219.0981	2	29528
244.0939	8.54E-05	0.0088	2	GDAE	b3	C9 H14 N3 O5	244.0933	3	27865
373.1357	1.24E-04	0.0122	3	GDAE	M+H-H2O	C14 H21 N4 O8	373.1359	0	30610
391.1470	2.16E-03	0.0172	58	GDAE	M+H	C14 H23 N4 O9	391.1465	1	22768
392.1502	3.56E-04	0.0167	10	GDAE	M+H	isotopologue	392.1493	2	23468
393.1519	6.73E-05	0.0120	2	GDAE	M+H	isotopologue	393.1509	3	32735
413.1352	1.60E-04	0.0175	4	GDAE	M+Na	C14 H22 N4 O9 Na	413.1284	16	23554
417.1073	2.18E-04	0.0167	6	GDAE	M+Al	C14 H22 N4 O9 Al	417.1196	-30	24962
418.1145	4.87E-05	0.0110	1	GDAE	M+Si	C14 H22 N4 O9 Si	418.1150	-1	37976
429.1089	7.76E-05	0.0158	2	GDAE	M+K	C14 H22 N4 O9 K	429.1024	15	27228
781.2905	5.05E-04	0.0468	14	GDAE	2M+H	C28 H45 N8 O18	781.2852	7	16698

Table C1-continue

SKSR and AAG mixture without Silicon Nanoparticles (1 scan)									
Gaussian Peak Fit				Peak Assignment				Performance	
Center Max in Da	Max Height in a.u.	FWHM	S/N	Peptide	Ion type	Chemical stochiometry	Theoretical mass in Da	Mass accuracy (ppm)	Mass resolution
100.0871	1.80E-04	0.005	5	SKSR	R	C4 H10 N3	100.0875	-4	21386
112.0875	3.49E-04	0.005	9	SKSR	R	C5 H10 N3	112.0875	0	21892
115.0873	6.72E-04	0.005	18	AAG	a2	C5 H11 N2 O	115.0871	1	21921
143.0820	9.89E-04	0.008	27	SKSR	y1-NH3-NH	C6 H11 N2 O2	143.0821	0	16993
147.0768	5.17E-04	0.007	14	AAG	y2	C5 H11 N2 O3	147.0770	-1	21471
158.0925	8.15E-05	0.005	2	SKSR	y1-NH3	C6 H12 N3 O2	158.0930	-3	33708
160.1077	2.78E-04	0.006	8	SKSR	z1	C6 H14 N3 O2	160.1086	-6	27557
175.1190	2.59E-04	0.006	7	SKSR	y1	C6 H15 N4 O2	175.1195	-3	30089
201.1002	7.26E-05	0.021	2	SKSR	x1	C7 H13 N4 O3	201.0988	7	9358
218.1139	5.77E-03	0.025	156	AAG	MH	C8 H16 N3 O4	218.1141	-1	8635
233.1607	1.00E-04	0.020	3	SKSR	c2	C9 H21 N4 O3	233.1614	-3	11583
244.0868	6.34E-04	0.027	17	AAG	M+Al	C8 H15 N3 O4 Al	244.0878	-4	9091
256.0684	1.24E-03	0.017	34	AAG	M+K-39	C8 H15 N3 O4 K39	256.0700	-6	15089
258.0650	1.09E-04	0.008	3	AAG	M+K-41	C8 H15 N3 O4 K41	258.0681	-12	34090
320.1937	5.71E-05	0.024	2	SKSR	c3	C12 H26 N5 O5	320.1934	1	13505
390.2498	7.77E-05	0.057	2	SKSR	y3	C15 H32 N7 O5	390.2465	8	6875
431.2733	1.44E-04	0.092	4	SKSR	MH-H2O-C-O	C17 H35 N8 O5	431.2730	1	4711
435.2244	1.63E-04	0.110	4	AAG	2M+H	C16 H31 N6 O8	435.2203	9	3952
447.2642	5.81E-05	0.044	2	SKSR	MH-H2O-C	C17 H35 N8 O6	447.2680	-8	10186
477.2840	1.89E-03	0.114	51	SKSR	MH	C18 H37 N8 O7	477.2785	12	4174
478.2846	4.14E-04	0.115	11	SKSR	MH	Isotopologue	478.2806	8	4147
479.2789	7.34E-05	0.063	2	SKSR	MH	Isotopologue	479.2828	-8	7552
515.2455	5.14E-05	0.104	1	SKSR	M+K-39	C18 H36 N8 O7 K	515.2344	21	4950

Table C1-continue

SKSR and AAG with Silicon Nanoparticles (1 scan)									
Gaussian Peak Fit				Peak Assignment				Performance	
Center Max in Da	Max Height in a.u.	FWHM	S/N	Peptide	Ion type	Chemical stochiometry	Theoretical mass in Da	Mass accuracy (ppm)	Mass resolution
112.0864	5.50E-05	0.0062	1	SKSR	R	C5 H10 N3	112.0875	-10	18020
115.0876	3.39E-04	0.0119	9	AAG	a2	C5 H11 N2 O	115.0871	4	9696
132.9050	7.50E-05	0.0049	2	Cs	Cs	Cs	132.9055	-4	27291
143.0823	4.63E-04	0.0163	12	AAG	b2	C6 H11 N2 O2	143.0815	6	8762
147.0767	2.43E-04	0.0154	6	AAG	y2	C5 H11 N2 O3	147.0770	-2	9557
218.1131	1.05E-03	0.0115	28	AAG	MH	C8 H16 N3 O4	218.1141	-5	18901
256.0763	1.12E-04	0.0182	3	AAG	M+K-39	C8 H15 N3 O4 K	256.0700	24	14070
431.2712	8.57E-05	0.0181	2	SKSR	MH-H2O-C-O	C17 H35 N8 O5	431.2730	-4	23893
435.2200	1.11E-04	0.0303	3	AAG	2M+H	C16 H31 N6 O8	435.2203	-1	14354
477.2820	5.32E-04	0.0775	14	SKSR	MH	C18 H37 N8 O7	477.2785	7	6156

Appendice D: Supplementary material to Chapter 5

Supplementary Material to Proto-Manuscript

Simultaneous measurements of collision cross section and mass to charge ratio via Orbitrap mass spectrometry

Ziqin Ni[1*] and Ricardo Arevalo Jr.[1]

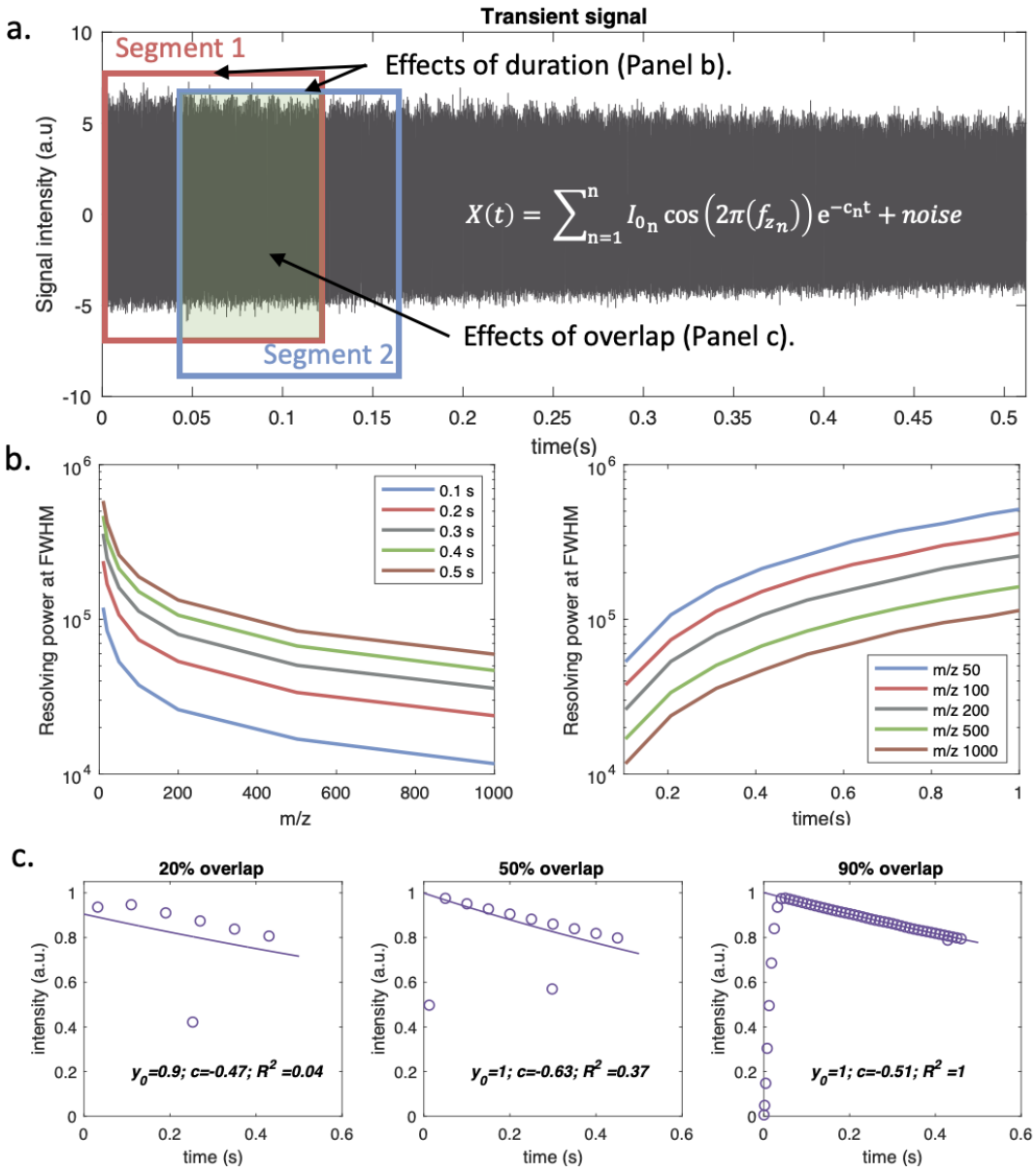
[1]University of Maryland, College Park, MD, 20742.

Email: zni@umd.edu

Content

D1. Implementation of short-time Fast Fourier Transform (StFFT)

In panel a, transient signal was first sliced into multiple segments with a fixed duration in time and a certain overlap to the next segments. Each segment would be treated as a standalone transient signal and processed with standard FT operation to generate mass spectra. The duration of each transient segment will determine the maximum mass resolving power in the mass spectra. Panel b shows the maximum mass resolving power at specific m/z for a given data acquisition time. Given the same peak at a specific m/z, a wider slice window will lead to higher mass resolving power. For the same data acquisition time, mass resolving power decreases with increasing m/z. Panel c shows that a greater overlap between consecutive segments improves the continuity of decay profiles, thus increasing the goodness of the fit as illustrated by R^2, as well as the accuracy of estimated initial intensity I_0 and decay constant c. This trend is observed for peaks at m/z 10-2000 Th. The input I_0 and c for simulated signals were 1 and 0.5. The simulation of transient signals is described in D4.


a.
Transient signal
Segment 1
Effects of duration (Panel b).
Segment 2
Effects of overlap (Panel c).
$X(t) = \sum_{n=1}^{n} I_{0_n} \cos\left(2\pi(f_{z_n})\right) e^{-c_n t} + noise$
Signal intensity (a.u)
time(s)
b.
Resolving power at FWHM
m/z
0.1 s
0.2 s
0.3 s
0.4 s
0.5 s
Resolving power at FWHM
time(s)
m/z 50
m/z 100
m/z 200
m/z 500
m/z 1000
c.
20% overlap
y_0=0.9; c=-0.47; R^2 =0.04
50% overlap
y_0=1; c=-0.63; R^2 =0.37
90% overlap
y_0=1; c=-0.51; R^2 =1
intensity (a.u.)
time (s)

D2. The statistics of identified chemical species in mass spectra.

m/z	identity	S/N	μ_{R^2}*	RSD_R^2	μ_c*	RSD_c	CCS* (Å^2)	RSD_{CCS}	Fig.
333.1765	$[KDA+H]^+$	4243	1.00	0.1%	0.43	5%	-	-	2
355.1581	$[KDA+Na]^+$	94	0.95	2.6%	0.55	13%	-	-	2
665.3435	$[2KDA+H]^+$	190	1.00	0.1%	0.75	19%	227	5%	2
391.1457	$[GDAE+H]^+$	697	1.00	0.1%	0.50	8%	-	-	2
413.1268	$[GDAE+Na]^+$	79	0.99	0.6%	0.73	3%	-	-	2
781.2806	$[2GDAE+H]^+$	101	1.00	0.3%	0.64	19%	260	1%	2
76.0399	$[Glycine+H]^{+\&}$	-	-	-	-	-	-	-	3
90.0555	$[Alanine+H]^{+\&}$	-	-	-	-	-	-	-	3
106.0504	$[Serine+H]^{+\&}$	-	-	-	-	-	-	-	3
116.0803	**$[Proline+H]^+$**	**55**	**0.98**	**1%**	**3.5**	**10%**	**-**	**-**	**3**
118.0960	$[Valine+H]^+$	18	0.99	1%	5.2	12%	99	0.7%	3
120.0907	$[Threonine+H]^+$	16	0.96	3%	4.7	8%	101	0.7%	3
132.1127	$[(Iso)leucine+H]^{+\wedge}$	146	0.94	1%	1.3	4%	111	0.7%	3
134.0453	$[Aspartic\ acid+H]^{+\&}$	-	-	-	-	-	-	-	3
147.1245	$[Lysine+H]^+$	21	0.99	1%	6.5	5%	108	0.6%	3
148.0722	$[Glutamic\ acid+H]^+$	17	1.00	0%	6.9	5%	108	0.4%	3
150.0703	$[Methionine+H]^+$	102	0.99	1%	3.4	6%	115	0.4%	3
156.0891	$[Histidine+H]^+$	61	1.00	0%	6.2	3%	112	0.4%	3
166.0995	$[Phenylalanine+H]^+$	373	0.98	0%	1.3	5%	124	0.6%	3
175.1329	$[Arginine+H]^+$	132	1.00	0%	5.2	4%	121	0.4%	3
182.0957	**$[Tyrosine+H]^+$**	**100**	**1.00**	**0%**	**6.8**	**2%**	**120**	**-**	**3**
241.0497	$[Cystine+H]^+$	97	0.99	0%	8.3	1%	135	0.4%	3
262.6352	$[MRFA+H]^{++}$	23	0.81	19%	0.9	37%	153	0.4%	4
391.1456	**$[GDAE+H]^+$**	**65**	**0.96**	3%	**0.6**	25%	**-**	**-**	**4**
524.2644	$[MRFA+H]^+$	11	0.97	2%	3.2	16%	211	1%	4
781.2849	**$[2GDAE+H]^+$**	**12**	**0.92**	**8%**	**2.6**	36%	-	**-**	**4**

1121.9963	[Ultramark 1]$^+$	12	0.95	2%	2.7	29%	309	1%	4
1221.9889	[Ultramark 2]$^+$	27	0.98	1%	2.3	25%	325	1%	4
1321.9838	[Ultramark 3]$^+$	42	0.99	1%	2.1	11%	338	0.8%	4
1421.9782	[Ultramark 4]$^+$	57	0.98	2%	1.7	6%	352	0.5%	4
1521.9719	[Ultramark 5]$^+$	56	0.99	1%	1.6	8%	365	0.5%	4
1621.9647	[Ultramark 6]$^+$	42	0.99	1%	1.7	6%	376	0.6%	4
1721.9602	[Ultramark 7]$^+$	31	0.97	2%	2.0	12%	386	0.8%	4
1821.9538	[Ultramark 8]$^+$	15	0.95	4%	2.1	13%	397	0.6%	4

& These amino acids were not detected in the mass spectrum due to S/N < 10

^ Peak at m/z 132.1127 was a combination of two ionic species, protonated isoleucine and protonated leucine. We reported the average reference CCS values of the two protonated amino acids.

D3. The prediction of CCS values via MetCCS

The CCS values were predicted via online MetCCS predictor developed by Zhou et al., 2016. The model was developed using the measured CCS values of ~ 400 metabolites in nitrogen buffer gas. The empirical results were used to train and optimize machine learning algorithms that were built to predict CCS values. The final model was externally validated using independent set of metabolites and demonstrated a median relative error of 3%. The model was used to predict over 35,000 metabolites in the Human Metabolome Database (HMDB). The input parameters and predicted CCS values were summarized in the table below.

Table D3-1. The input parameters requested by MetCCS predictors. The structural information are imported from Pubchem and chemical properties are predicted via online predictor using ChemAxon.

Name:	KDA	GDAE
Exact_Mass:	332.1697	390.1386
Formal_Charge:	1	1
Physiological_Charge:	0	0
logP_ALOGPS:	-3.93	-3.24
logP_ChemAxon:	-8.10	-6.86
logS:	-1.06	-0.97
pKa_Strongest_Acidic:	3.17	3.53
pKa_Strongest_Basic:	10.21	7.84
Hydrogen_Acceptor_Count:	10	13
Hydrogen_Donor_Count:	6	7
Polar_Surface_Area:	184.84	225.22
Rotatable_Bond_Count:	11	12
Refractivity:	81.31	88.19
Polarizability:	31.50	34.11

Table D3-2. The predicted CCS values by MetCCS predictor.

	KDA	GDAE
Charge	1	1
M+H	172.9	186.5
M+Na	174.3	187.9
M-H2O+H	171.7	185.3
M-H	189.6	197.6

D4. Simulation and signal processing of transient signals

The transient signal of n-component mixtures was simulated via a generic formula

$$I = \sum_{n=1}^{n} I_{0_n} cos\ (2\pi f_{z_n})\, e^{-c_n t} + noise, \qquad Eqn\ (D4.1)$$

where I_0 was the initial ion intensity, c was the decay constant, t was the data acquisition time, f_z was axial oscillation frequency of ionic species, and subscript n represents multi-component species in the Orbitrap mass analyzer. The f_z was estimated using formula,

$$f_z = k(m/z)^{-1/2}, \qquad Eqn\ (D4.2)$$

where the k is a restoring factor that is affected by the geometry and voltage potential applied to the center electrode. During data acquisition, the sampling frequency f_s determines the discreteness of the transient signal and the data volume.

This study simulated transient signal of singly charged ionic species of a wide mass range from m/z 100 to 1000 with a fixed starting intensity $I_0 = 1$, same decay constant c of 0.5. A white gaussian noise at signal to noise of 10 was added into the transient signal. The k value and f_s in this study was used as the same values as the commercial Q-Exactive instrument. The transient signal was processed in same way as the data

processing procedure for other collected mass spectra in this study. We applied 10x zerofillings and the hanning window prior to the fourier transform operation. The frequency spectrum was then calibrated using peak at *m/z* 100 to produce mass spectrum. All peaks were identified and fitted with gaussian peak fitting to extrapolate center mass, FWHM, peak maxima to derive mass resolving power and mass accuracy.

Bibliography

Anderson, A., Wirth, R., & Thomas, R. (2008). The alteration of metamict zircon and its role in the remobilization of high-field-strength elements in the Georgeville granite, Nova Scotia. *Canadian Mineralogist*, *46*(1), 1–18. https://doi.org/10.3749/canmin.46.1.1

Arevalo, R., & McDonough, W. F. (2010). Chemical variations and regional diversity observed in MORB. *Chemical Geology*, *271*(1–2), 70–85. https://doi.org/10.1016/j.chemgeo.2009.12.013

Bea, F. (1996). Residence of REE, Y, Th and U in granites and crustal protoliths; implications for the chemistry of crustal melts. *Journal of Petrology*, *37*(3), 521–552. https://doi.org/10.1093/petrology/37.3.521

Bell, E. A., Boehnke, P., Barboni, M., & Harrison, T. M. (2019). Tracking chemical alteration in magmatic zircon using rare earth element abundances. *Chemical Geology*, *510*, 56–71. https://doi.org/10.1016/j.chemgeo.2019.02.027

Bell, E. A., Boehnke, P., Hopkins-Wielicki, M. D., & Harrison, T. M. (2015). Distinguishing primary and secondary inclusion assemblages in Jack Hills zircons. *Lithos*, *234–235*, 15–26. https://doi.org/10.1016/j.lithos.2015.07.014

Belousova, E. A., Griffin, W. L., O'Reilly, S. Y., & Fisher, N. I. (2002). Igneous zircon: Trace element composition as an indicator of source rock type.

Contributions to Mineralogy and Petrology, *143*(5), 602–622. https://doi.org/10.1007/s00410-002-0364-7

Bindeman, I. N., & Valley, J. W. (2001). Low-δ18O Rhyolites from Yellowstone: Magmatic evolution based on analyses of zircons and individual phenocrysts. *Journal of Petrology*, *42*(8), 1491–1517. https://doi.org/10.1093/petrology/42.8.1491

Boehnke, P., Watson, E. B., Trail, D., Harrison, T. M., & Schmitt, A. K. (2013). Zircon saturation re-revisited. *Chemical Geology*, *351*, 324–334. https://doi.org/10.1016/J.CHEMGEO.2013.05.028

Borisov, A., & Aranovich, L. (2019). Zircon solubility in silicate melts: New experiments and probability of zircon crystallization in deeply evolved basic melts. *Chemical Geology*, *510*, 103–112. https://doi.org/10.1016/j.chemgeo.2019.02.019

Burnham, A. D., & Berry, A. J. (2012). An experimental study of trace element partitioning between zircon and melt as a function of oxygen fugacity. *Geochimica et Cosmochimica Acta*, *95*, 196–212. https://doi.org/10.1016/j.gca.2012.07.034

Burnham, A. D., & Berry, A. J. (2017). Formation of Hadean granites by melting of igneous crust. *Nature Geoscience*, *10*(6), 457–461. https://doi.org/10.1038/ngeo2942

Carley, T. L., Miller, C. F., Wooden, J. L., Padilla, A. J., Schmitt, A. K., Economos, R. C., Bindeman, I. N., Jordan, B. T. (2014). Iceland is not a magmatic analog for the Hadean: Evidence from the zircon record. *Earth and Planetary Science Letters*, *405*, 85–97. https://doi.org/10.1016/j.epsl.2014.08.015

Cavosie, A. J., Valley, J. W., & Wilde, S. A. (2006). Correlated microanalysis of zircon: Trace element, δ18O, and U–Th–Pb isotopic constraints on the igneous origin of complex> 3900 Ma detrital grains. Geochimica et Cosmochimica Acta, 70(22), 5601-5616.

Chapman, J. B., Gehrels, G. E., Ducea, M. N., Giesler, N., & Pullen, A. (2016). A new method for estimating parent rock trace element concentrations from zircon. *Chemical Geology*, *439*, 59–70. https://doi.org/10.1016/j.chemgeo.2016.06.014

Chou, M. (1978). Calibration of oxygen buffers at elevated P and T using the hydration fugacity sensor. *American Mineralogist*, *63*, 690–703.

Claiborne, L. L., Miller, C. F., Flanagan, D. M., Clynne, M. A., & Wooden, J. L. (2010). Zircon reveals protracted magma storage and recycling beneath Mount St. Helens. *Geology*, *38*(11), 1011-1014.

Claiborne, L. L., Miller, C. F., Walker, B. A., Wooden, J. L., Mazdab, F. K., & Bea, F. (2006). Tracking magmatic processes through Zr/Hf ratios in rocks and Hf

and Ti zoning in zircons: an example from the Spirit Mountain batholith, Nevada. *Mineralogical Magazine*, *70*(5), 517-543.

Corfu, F., Hanchar, J. M., Hoskin, P. W. O., & Kinny, P. (2003). Atlas of Zircon Textures. *Reviews in Mineralogy and Geochemistry*, *53(1)*, 469–500. https://doi.org/10.2113/0530469

El-Bialy, M. Z., & Ali, K. A. (2013). Zircon trace element geochemical constraints on the evolution of the Ediacaran (600-614Ma) post-collisional Dokhan Volcanics and Younger Granites of SE Sinai, NE Arabian-Nubian Shield. *Chemical Geology*, *360–361*, 54–73. https://doi.org/10.1016/j.chemgeo.2013.10.009

Ferry, J. M., & Watson, E. B. (2007). New thermodynamic models and revised calibrations for the Ti-in-zircon and Zr-in-rutile thermometers. *Contributions to Mineralogy and Petrology*, *154*(4), 429–437. https://doi.org/10.1007/s00410-007-0201-0

Frost, D. J., & McCammon, C. A. (2008). The Redox State of Earth's Mantle. *Annual Review of Earth and Planetary Sciences*, *36*(1), 389–420. https://doi.org/10.1146/annurev.earth.36.031207.124322

Frost, D. J., Mann, U., Asahara, Y., & Rubie, D. C. (2008). The redox state of the mantle during and just after core formation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1883), 4315–4337. https://doi.org/10.1098/rsta.2008.0147

Fu, B., Page, F. Z., Cavosie, A. J., Fournelle, J., Kita, N. T., Lackey, J. S., Wilde, S. A., Valley, J. W. (2008).Ti-in-zircon thermometry: Applications and limitations. *Contributions to Mineralogy and Petrology*, *156*(2), 197–215. https://doi.org/10.1007/s00410-008-0281-5

Gale, A., Dalton, C. A., Langmuir, C. H., Su, Y., & Schilling, J.-G. (2013). The mean composition of ocean ridge basalts. *Geochemistry, Geophysics, Geosystems*, *14*(3), 489–518. https://doi.org/10.1029/2012GC004334

Grimes, C. B., John, B. E., Kelemen, P. B., Mazdab, F. K., Wooden, J. L., Cheadle, M. J., Hanghøj, K., Schwartz, J. J. (2007). Trace element chemistry of zircons from oceanic crust: A method for distinguishing detrital zircon provenance. *Geology*, *35*(7), 643. https://doi.org/10.1130/G23603A.1

Grimes, C. B., Wooden, J. L., Cheadle, M. J., & John, B. E. (2015). "Fingerprinting" tectono-magmatic provenance using trace elements in igneous zircon. *Contributions to Mineralogy and Petrology*, *170*(5–6), 46. https://doi.org/10.1007/s00410-015-1199-3

Hanchar, J. M., & Watson, E. B. (2003). Zircon saturation thermometry. *Reviews in Mineralogy and Geochemistry*, *53*(1), 89–112. https://doi.org/10.2113/0530089

Harrison, T. M., Watson, E. B., & Aikman, A. B. (2007). Temperature spectra of zircon crystallization in plutonic rocks. *Geology*, *35*(7), 635–638. https://doi.org/10.1130/G23505A.1

Hawkesworth, C. J., & Kemp, A. I. S. (2006). Using hafnium and oxygen isotopes in zircons to unravel the record of crustal evolution. *Chemical Geology*, *226*(3–4), 144–162. https://doi.org/10.1016/j.chemgeo.2005.09.018

Hofmann, A. E., Baker, M. B., & Eiler, J. M. (2014). Sub-micron-scale trace-element distributions in natural zircons of known provenance: implications for Ti-in-zircon thermometry. *Contrib Mineral Petrol*, *168*, 1057. https://doi.org/10.1007/s00410-014-1057-8

Hoskin, P. W. O. (2005). Trace-element composition of hydrothermal zircon and the alteration of Hadean zircon from the Jack Hills, Australia. *Geochimica et Cosmochimica Acta*, *69*(3), 637–648. https://doi.org/10.1016/j.gca.2004.07.006

Hoskin, P. W. O., & Black, L. P. (2002). Metamorphic zircon formation by solid-state recrystallization of protolith igneous zircon. *Journal of Metamorphic Geology*, *18*(4), 423–439. https://doi.org/10.1046/j.1525-1314.2000.00266.x

Hoskin, P. W. O., & Ireland, T. R. (2000). Rare earth element chemistry of zircon and its use as a provenance indicator. *Geology*, *28*(7), 627–630. https://doi.org/10.1130/0091-7613(2000)028<0627:REECOZ>2.3.CO;2

Hoskin, P. W. O., & Schaltegger, U. (2003). The composition of zircon and igneous and metamorphic petrogenesis. *Reviews in Mineralogy and Geochemistry*, *53*(1), 27-62. https://doi.org/10.1046/j.1525-1314.2000.00266.x

Kelly, N. M., & Harley, S. L. (2005). An integrated microtextural and chemical approach to zircon geochronology: refining the Archaean history of the Napier Complex, east Antarctica. *Contributions to Mineralogy and Petrology*, *149*(1), 57–84. https://doi.org/10.1007/s00410-004-0635-6

Kirkland, C. L., Smithies, R. H., Taylor, R. J. M., Evans, N., & McDonald, B. (2015). Zircon Th/U ratios in magmatic environs. *Lithos*, *212–215*, 397–414. https://doi.org/10.1016/j.lithos.2014.11.021

Kosler, J., & Sylvester, P. (2005). Present trends and the future of zircon in geochronology: laser ablation ICPMS. *Reviews in Mineralogy and Geochemistry*, *53*(1), 243–275. https://doi.org/10.2113/0530243

Kump, L. R., Kasting, J. F., & Barley, M. E. (2001). Rise of atmospheric oxygen and the "upside-down" Archean mantle. *Geochemistry, Geophysics, Geosystems*, *2*(1), 1525-2027. https://doi.org/10.1029/2000gc000114

Lee, C. T. A., & Bachmann, O. (2014). How important is the role of crystal fractionation in making intermediate magmas? Insights from Zr and P systematics. *Earth and Planetary Science Letters*, *393*, 266-274. https://doi.org/10.1016/j.epsl.2014.02.044

Lee, C.-T. A., Luffi, P., Chin, E. J., Bouchet, R., Dasgupta, R., Morton, D. M., Roux, V. L., Yin, Q., Jin, D. (2012). Copper systematics in arc magmas and implications for crust-mantle differentiation. *Science*, *336*(6077), 64–68. https://doi.org/10.1126/science.1217313

Linnen, R. L., & Keppler, H. (2002). Melt composition control of Zr/Hf fractionation in magmatic processes. *Geochimica et Cosmochimica Acta*, *66*(18), 3293–3301. https://doi.org/10.1016/S0016-7037(02)00924-9

Loucks, R. R., Fiorentini, M. L., & Henríquez, G. J. (2020). New magmatic oxybarometer using trace elements in zircon. *Journal of Petrology*.

McCammon, C. (2005). GEOCHEMISTRY: The Paradox of Mantle Redox. *Science*, *308*(5723), 807–808. https://doi.org/10.1126/science.1110532

McDonough, W. F., & Sun, S. S. (1995). The Composition of the Earth. *Chemical Geology*, *120*(3–4), 223–253. https://doi.org/10.1016/0009-2541(94)00140-4

Mysen, B. O., Holtz, F., Pichavant, M., Beny, J. M., & Montel, J. M. (1997). Solution mechanisms of phosphorus in quenched hydrous and anhydrous granitic glass as a function of peraluminosity. *Geochimica et Cosmochimica Acta*, *61*(18), 3913–3926. https://doi.org/10.1016/S0016

Nandedkar, R. H., Ulmer, P., & Müntener, O. (2014). Fractional crystallization of primitive, hydrous arc magmas: An experimental study at 0.7 GPa. *Contributions to Mineralogy and Petrology*, *167*(6), 1–27. https://doi.org/10.1007/s00410-014-1015-5

Page, F. Z., Fu, B., Kita, N. T., Fournelle, J., Spicuzza, M. J., Schulze, D. J., … Valley, J. W. (2007). Zircons from kimberlite: New insights from oxygen isotopes, trace elements, and Ti in zircon thermometry. *Geochimica et*

Cosmochimica Acta, *71*(15), 3887–3903. https://doi.org/10.1016/j.gca.2007.04.031

Pidgeon, R. T., Nemchin, A. A., Roberts, M. P., Whitehouse, M. J., & Bellucci, J. J. (2019). The accumulation of non-formula elements in zircons during weathering: Ancient zircons from the Jack Hills, Western Australia. *Chemical Geology*, *530*. https://doi.org/10.1016/j.chemgeo.2019.119310

Pupin, J.-P. (2000). Granite genesis related to geodynamics from Hf-Y in zircon. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *91*(1–2), 245–256. https://doi.org/10.1017/S0263593300007410

Rasmussen, B., Fletcher, I. R., Muhling, J. R., Gregory, C. J., & Wilde, S. A. (2011). Metamorphic replacement of mineral inclusions in detrital zircon from Jack Hills, Australia: Implications for the Hadean Earth. *Geology*, *39*(12), 1143–1146. https://doi.org/10.1130/G32554.1

Rubatto, D. (2002). Zircon trace element geochemistry: Partitioning with garnet and the link between U-Pb ages and metamorphism. *Chemical Geology*, *184*(1–2), 123–138. https://doi.org/10.1016/S0009-2541(01)00355-2

Rubatto, D. (2017). Zircon: The Metamorphic Mineral. *Reviews in Mineralogy and Geochemistry*, *83*(1), 261–295. https://doi.org/10.2138/rmg.2017.83.9

Rubatto, D., & Hermann, J. (2007). Experimental zircon/melt and zircon/garnet trace element partitioning and implications for the geochronology of crustal rocks.

Chemical Geology, *241*(1–2), 38–61. https://doi.org/10.1016/j.chemgeo.2007.01.027

Rudnick, R. L. (1995). Making continental crust. *Nature*, *378*(6557), 571–578. https://doi.org/10.1038/378571a0

Samperton, K. M., Bell, E. A., Barboni, M., Brenhin Keller, C., & Schoene, B. (2017). Zircon age-temperature-compositional spectra in plutonic rocks. *Geology*, *45*(11), 983–986. https://doi.org/10.1130/G38645.1

Sawaki, Y., Suzuki, K., Asanuma, H., Okabayashi, S., Hattori, K., Saito, T., & Hirata, T. (2017). Geochemical characteristics of zircons in the Ashizuri A-type granitoids: An additional granite topology tool for detrital zircon studies. *Island Arc*, *26*(6), e12216. https://doi.org/10.1111/iar.12216

Shao, T., Xia, Y., Ding, X., Cai, Y., & Song, M. (2019). Zircon saturation in terrestrial basaltic melts and its geological implications. *Solid Earth Sciences*, *4*(1), 27–42. https://doi.org/10.1016/j.sesci.2018.08.001

Siégel, C., Bryan, S. E., Allen, C. M., & Gust, D. A. (2018). Use and abuse of zircon-based thermometers: A critical review and a recommended approach to identify antecrystic zircons. *Earth-Science Reviews*, *176*, 87–116. https://doi.org/10.1016/J.EARSCIREV.2017.08.011

Smythe, D. J., & Brenan, J. M. (2015). Cerium oxidation state in silicate melts: Combined fO2, temperature and compositional effects. *Geochimica et Cosmochimica Acta*, *170*, 173–187. https://doi.org/10.1016/j.gca.2015.07.016

Smythe, D. J., & Brenan, J. M. (2016). Magmatic oxygen fugacity estimated using zircon-melt partitioning of cerium. *Earth and Planetary Science Letters*, *453*, 260–266. https://doi.org/10.1016/j.epsl.2016.08.013

Spandler, C., Hermann, J., & Rubatto, D. (2004). Exsolution of thortveitite, yttrialite and xenotime during low temperature recrystallization of zircon from New Caledonia, and their significance for trace element incorporation in Nanogranitoids and crustal melting View project Magma fertility for Au-Cu, petrogenesis and geodynamic setting of Carboniferous and Permian magmatic complexes (Queensland, Australia) View project. *American Mineralogist*, *89*, 1795–1806. https://doi.org/10.2138/am-2004-11-1226

Spencer, C. J., Cavosie, A. J., Raub, T. D., Rollinson, H., Jeon, H., Searle, M. P., … Evans, N. J. (2017). Evidence for melting mud in Earth's mantle from extreme oxygen isotope signatures in zircon. *Geology*, *45*(11), 975–978. https://doi.org/10.1130/G39402.1

Tang, M., Erdman, M., Eldridge, G., & Lee, C. T. A. (2018). The redox "filter" beneath magmatic orogens and the formation of continental crust. *Science Advances*, *4*(5). https://doi.org/10.1126/sciadv.aar4444

Taylor, R. J. M., Harley, S. L., Hinton, R. W., Elphick, S., Clark, C., & Kelly, N. M. (2015). Experimental determination of REE partition coefficients between zircon, garnet and melt: a key to understanding high-T crustal processes. *Journal of Metamorphic Geology*, *33*(3), 231–248. https://doi.org/10.1111/jmg.12118

Trail, D., Bruce Watson, E., & Tailby, N. D. (2012). Ce and Eu anomalies in zircon as proxies for the oxidation state of magmas. *Geochimica et Cosmochimica Acta*, *97*, 70–87. https://doi.org/10.1016/J.GCA.2012.08.032

Trail, D., Mojzsis, S. J., Harrison, T. M., Schmitt, A. K., Watson, E. B., & Young, E. D. (2007). c. *Geochemistry, Geophysics, Geosystems*, *8*(6). https://doi.org/10.1029/2006GC001449

Trail, D., Tailby, N., Wang, Y., Mark Harrison, T., & Boehnke, P. (2017). Aluminum in zircon as evidence for peraluminous and metaluminous melts from the Hadean to present. *Geochemistry, Geophysics, Geosystems*, *18*(4), 1580–1593. https://doi.org/10.1002/2016GC006794

Trail, D., Watson, E. B., & Tailby, N. D. (2011). The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature*, *480*(7375), 79–82. https://doi.org/10.1038/nature10655

Valley, J. W. (2003). Oxygen Isotopes in Zircon. *Reviews in Mineralogy and Geochemistry*, *53*(1), 343–385. https://doi.org/10.2113/0530343

Valley, J. W., Kinny, P. D., Schulze, D. J., & Spicuzza, M. J. (1998). Zircon megacrysts from kimberlite: oxygen isotope variability among mantle melts. *Contributions to mineralogy and petrology*, *133*(1-2), 1-11. https://doi.org/10.1007/s004100050432

Wang, Q., Zhu, D. C., Zhao, Z. D., Guan, Q., Zhang, X. Q., Sui, Q. L., Hu, Z. C., Mo, X. X. (2012). Magmatic zircons from I-, S- and A-type granitoids in Tibet: Trace element characteristics and their application to detrital zircon provenance study. *Journal of Asian Earth Sciences*, *53*, 59–66. https://doi.org/10.1016/j.jseaes.2011.07.027

Wang, X., Griffin, W. L., & Chen, J. (2010). Hf contents and Zr/Hf ratios in granitic zircons. *Geochemical Journal*, *44*(1), 65–72. https://doi.org/10.2343/geochemj.1.0043

Watson, E. B. (1979). Zircon saturation in felsic liquids: Experimental results and applications to trace element geochemistry. *Contributions to Mineralogy and Petrology*, *70*(4), 407–419. https://doi.org/10.1007/BF00371047

Watson, E. B., & Harrison, T. M. (1983). Zircon saturation revisited: temperature and composition effects in a variety of crustal magma types. *Earth and Planetary Science Letters*, *64*(2), 295–304. https://doi.org/10.1016/0012-821X(83)90211-X

Watson, E. B., & Harrison, T. M. (2005). Zircon Thermometer Reveals Minimum Melting Conditions on Earliest Earth. *Science*, *308*(5723), 841–844. https://doi.org/10.1126/science.1110265

Watson, E. B., Wark, D. A., & Thomas, J. B. (2006). Crystallization thermometers for zircon and rutile. *Contributions to Mineralogy and Petrology*, *151*(4), 413–433. https://doi.org/10.1007/s00410-006-0068-5

Whitehouse, M. J., & Platt, J. P. (2003). Dating high-grade metamorphism - Constraints from rare-earth elements in zircon and garnet. *Contributions to Mineralogy and Petrology*, *145*(1), 61–74. https://doi.org/10.1007/s00410-002-0432-z

Wipperfurth, S. A., Guo, M., Šrámek, O., & McDonough, W. F. (2018). Earth's chondritic Th/U: Negligible fractionation during accretion, core formation, and crust–mantle differentiation. *Earth and Planetary Science Letters*, *498*, 196–202. https://doi.org/10.1016/j.epsl.2018.06.029

Wotzlaw, J. F., Bindeman, I. N., Stern, R. A., D'Abzac, F. X., & Schaltegger, U. (2015). Rapid heterogeneous assembly of multiple magma reservoirs prior to Yellowstone supereruptions. *Scientific Reports*, *5*, 14026. https://doi.org/10.1038/srep14026

Xiang, W., Griffin, W. L., Chen, J., Huang, P., & Li, X. (2011). U and Th Contents and Th/U Ratios of Zircon in Felsic and Mafic Magmatic Rocks: Improved

Zircon-Melt Distribution Coefficients. *Acta Geologica Sinica*, *85*(1), 164–174. https://doi.org/10.1111/j.1755-6724.2011.00387.x

Yakymchuk, C., Kirkland, C. L., & Clark, C. (2018). Th/U ratios in metamorphic zircon. *Journal of Metamorphic Geology*, *36*(6), 715–737. https://doi.org/10.1111/jmg.12307

Yang, X., Gaillard, F., & Scaillet, B. (2014). A relatively reduced Hadean continental crust and implications for the early atmosphere and crustal rheology. *Earth and Planetary Science Letters*, *393*, 210–219. https://doi.org/10.1016/J.EPSL.2014.02.056

Zhong, S., Feng, C., Seltmann, R., Li, D., & Dai, Z. (2018a). Geochemical contrasts between Late Triassic ore-bearing and barren intrusions in the Weibao Cu–Pb–Zn deposit, East Kunlun Mountains, NW China: constraints from accessory minerals (zircon and apatite). *Mineralium Deposita*, *53*(6), 855–870. https://doi.org/10.1007/s00126-017-0787-8

Zhong, S., Feng, C., Seltmann, R., Li, D., & Qu, H. (2018b). Can magmatic zircon be distinguished from hydrothermal zircon by trace element composition? The effect of mineral inclusions on zircon trace element composition. *Lithos*, *314–315*, 646–657. https://doi.org/10.1016/j.lithos.2018.06.029

Zou, X., Qin, K., Han, X., Li, G., Evans, N. J., Li, Z., & Yang, W. (2019). Insight into zircon REE oxy-barometers: A lattice strain model perspective. *Earth and*

Planetary Science Letters, *506*, 87–96.

https://doi.org/10.1016/J.EPSL.2018.10.031