FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

MOLECULAR LEVEL CHARACTERIZATION OF DISSOLVED ORGANIC MATTER INTEGRATING TRAPPED ION MOBILITY SPECTROMETRY AND FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETRY

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Dennys Leyva Bombuse

2022

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Dennys Leyva Bombuse, and entitled Molecular Level Characterization of Dissolved Organic Matter Integrating Trapped Ion Mobility Spectrometry and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Piero Gardinali

_____
John Berry

_____
Rudolf Jaffé

_____
Joong Ho Moon

_____
John Kominoski

_____
Francisco Fernandez Lima, Major Professor

Date of Defense: July 1, 2022

The dissertation of Dennys Leyva Bombuse is approved.

_____
Dean Michael R. Heithaus
College of Arts, Sciences and Education

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2022

DEDICATION

I dedicate this dissertation to my wife Lissette Monzon Paz, my parents Elena and Manuel, and my parents in law Luisa Maria and Raul.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor Francisco Fernandez Lima for supporting one of my most important career dreams. I would also thank you for pushing me into this challenging field that definitely made me a better researcher. I would also like to thank my committee members Rudolf Jaffé, Piero Gardinali, John Berry, Joongho Moon, and John Kominoski for their support and academic guidance throughout all these years. I would like to thank Dr. Mario Gomez from the FIU Advanced Mass Spectrometry Facility for teaching me so much FT-ICR MS. I would like to thank Christopher J. Thompson, Jeremy Wolff, Mark E. Ridgeway, and Melvin A. Park from Bruker Daltonics for their collaboration. I also thank Dr. Jaroslava Mikšovská and Dr. Fahad Saeed for your great collaborations. I would like to acknowledge Clement, Yarixa, Lilian, Kevin, Meiby, Miguel, Casandra, and Samuel for sharing so much hard and good times with me. Usman Tariq for your hard work on such complicated computational algorithms. Paolo Benigni, Jacob Porter, Anthony Castellanos, and Alan McKenzie for sharing your knowledge and valuable tips. Jessica Courson for your hard work during the 2021 summer research experience for undergraduates. I would also like to acknowledge Todd Crowl, Rita Teutonico, and Bradley Schonhoff from the CREST CAChE center at FIU for the fellowship I received and for all your support. I would like to acknowledge the faculty and staff from the FIU Department of Chemistry and Biochemistry, especially Maggie Autie, Watson Lees, Kelly Rein, Yong Cai, and Jeffrey Joens for sharing their knowledge and supporting me as a graduate student. I would like to thank Edward Castañeda, Ryan Bremen, and Kenny Anderson from the Ecosystem Ecology Laboratory at FIU for their field and laboratory support.

ABSTRACT OF THE DISSERTATION

MOLECULAR LEVEL CHARACTERIZATION OF DISSOLVED ORGANIC MATTER INTEGRATING TRAPPED ION MOBILITY SPECTROMETRY AND FOURIER TRANSFORM ION CYCLOTRON RESONANCE MASS SPECTROMETRY

by

Dennys Leyva Bombuse

Florida International University, 2022

Miami, Florida

Professor Francisco Fernandez-Lima, Major Professor

Dissolved organic matter (DOM) is an extremely complex mixture of organic molecules ubiquitous in aquatic systems and a critical component of the global carbon cycle. Little is known about DOM structural composition at the molecular level. The work presented in this dissertation summarizes the development of a novel analytical toolbox based on trapped ion mobility spectrometry and Fourier transform ion cyclotron resonance mass spectrometry (TIMS-FT-ICR MS) that has significantly contributed to expand our knowledge of DOM molecular complexity and diversity. The TIMS-FT-ICR MS/MS analysis provided for the first-time lower and upper estimation of the molecular isomeric diversity. The TIMS-FT-ICR MS/MS methodology was further developed to allow for chemical formula-based isomeric and neutral loss fragmentation structural description and database validation. This novel procedure enabled the unambiguous assignment of candidate isomeric structures based on accurate mass, database MS/MS matching scores, and ion mobility. A fast and routine structural characterization DOM workflow method

was developed: GraphDOM. The method utilizes neutral loss fragmentation patterns acquired using continuous accumulation of selected ions (CASI)-collision induced dissociation (CID) FT-ICR MS/MS. The neutral mass loss patterns are used to define structural families leading to the identification and visualization of the DOM transformational processes. The GraphDOM methodology was successfully applied to the characterization of DOM along a salinity transect of the Harney River, Florida Everglades. The GraphDOM method was further implemented with isomeric content description at the molecular level and applied to four common aquatic estuaries. The application of the GraphDOM methodology allowed for the first time identification of common and unique DOM transformational networks across aquatic ecosystems.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE PAGE

ABBREVIATIONS AND ACRONYMS

Atmospheric pressure photo ionization	APPI
Carbon-13 nuclear magnetic resonance	13C-NMR
Carboxylic functional group	COOH
Collisional cross-section	CCS
Collision induced dissociation	CID
Continuous accumulation of selected ions	CASI
Continuous wavelet transform	CWT
Correlated harmonic excitation field	CHEF
Differential Ion mobility Spectrometry	DMS
Direct current	DC
Dissolved organic carbon	DOC
Dissolved organic matter	DOM
Double bond equivalents	DBE
Double bond equivalents minus number of oxygens	DBE-O
Drift tube ion mobility spectrometry	DTIMS
Electrospray ionization	ESI
Excitation emission matrix	EEM
Fast Fourier transform	FFT
Field Asymmetric waveform ion mobility spectrometry	FAIMS
Fourier transform ion cyclotron resonance	FT-ICR
Fulvic acids	FA
High-performance liquid chromatography	HPLC

Harney River	HR
Humic acids	HA
Hydrogen-deuterium exchange	HDX
Infrared multiphoton dissociation	IRMPD
International humic substance society	IHSS
Ion cyclotron resonance	ICR
Ion mobility spectrometry	IMS
Kendrik mass defect	KMD
Liquid chromatography	LC
Mass spectrometry	MS
Mega words	MW
Modified aromatic index	AImod
Natural organic matter	NOM
National environmental laboratory accreditation	NELAC
Nuclear magnetic resonance	NMR
Parallel factor analysis	PARAFAC
Particulate organic matter	POC
Proton nuclear magnetic resonance	^{1}H-NMR
Quadrupole	Q
Quadrupole -Time of flight	Q-TOF
Radiofrequency	RF
Reverse phase	RP
River continuum concept	RCC

Size exclusion chromatography	SEC
Software Assisted Molecular Elucidation	SAME
Solid phase extraction	SPE
Styrene-divinylbenzene polymer	PPL
Sustained off resonance irradiation	SORI
Suwannee river fulvic acids	SRFA
Tandem mass spectrometry	MS/MS
Time of flight	TOF
Trapped ion mobility spectrometry	TIMS
Traveling wave ion mobility spectrometry	TWIMS
Two-dimensional	2D
Ultrahigh resolution mass spectrometry	UHRMS
Ultraviolet–visible spectroscopy	UV-visible

CHAPTER I

I. INTRODUCTION

1.1 Definition of DOM and its role in the environment

Carbon, the first chemical element of the group 6 in the Periodic Table, is considered the building block of life on Earth[1,2] This element can be found in nature (chemical rocks, sediments, soils, the ocean, atmosphere, and all living organisms) in various chemical forms. Carbon compounds are fundamental components of the Earth's biogeochemical equilibrium defined as the global carbon cycle. This cycle ultimately determines the climate on our planet by regulating carbon dioxide concentrations in the atmosphere[3].

Dissolved organic carbon (DOC), which represents a significant fraction of dissolved organic matter (DOM), is one of the most important forms of carbon, stored in water bodies in amounts comparable to the ones found in the atmosphere as CO_2 (Figure 1.1)[4-6]. Thus, variation of DOC levels due to natural and anthropogenic events can result in significant disruptions of the global carbon cycle, ultimately impacting the climate on our planet[4,7].

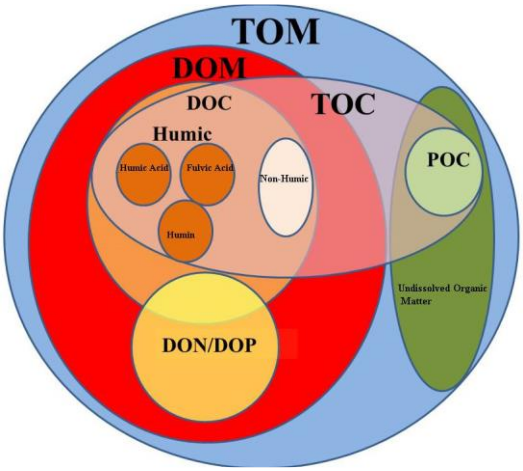


Figure 1. 1. Schematic representation of the most common forms of organic matter in natural waters Total Organic Matter (TOM), Total Organic Carbon (TOC), Dissolved Organic Matter (DOM), Dissolved Organic Carbon (DOC), Particulate Organic Carbon (POC), Dissolved Organic Nitrogen (DON), and Dissolved Organic Phosphorus (DOP) are indicated. DOC is also classified on humic (humic acid (HAs), fulvic acid (FAs), and humin) and non-humic substances based on their acid-base properties. (Reprinted from Fig. 1, Pagano T., et al., Water, 2014, 6, 2862-2897).

DOM is a complex mixture of organic molecules containing carbon, oxygen, nitrogen, sulfur, and phosphorus elements, resulting from the degradation of bacterial, algal, and higher plant organic materials[8-10]. Due to its high chemical heterogeneity, DOM is often operationally defined based on its physical and chemical properties rather than its chemical structure[11]. For example, a practical definition relies on the dissolved material collected after filtration[12]. Although there is no universal agreement on the specific pore size filter to use, the most widely accepted are in the range of 0.2 and 0.7 μm[13]. Fulvic acids (FAs), humic acids (HAs), and humin are fractions of DOM, resulting from a different operational classification based on the acid-base solubility (Figure 1.1). HAs, FAs and humin are part of humic substances, for which HAs are insoluble in water at pH 1, FAs are soluble at all pH conditions, and humin is insoluble under all pH conditions[14].

DOM is an essential component of the global carbon cycle and is considered a critical sink of atmospheric carbon[12,13,15,16] (Figure 1.2). This set of complex organic molecules constitutes a fundamental source of carbon transported from terrestrial environments to aquatic systems[17]. This plays a crucial ecological role as a source of nutrients for aquatic microorganisms[18], and functions as protection for aquatic organisms by controlling light penetration[19,20]. DOM can even alter the bioavailability[21], mobility[22], and ultimate fate of organic contaminants and trace metals[23]. Furthermore, some chemical components of DOM in freshwater systems have been shown to interact with chlorine and ozone, yielding harmful disinfection by-products in drinking water[24]. Because of all these previous reasons, a clear understanding of DOM's impact in climate change, ecology, and toxicology, requires a comprehensive knowledge of its constituents at a molecular level.[25]

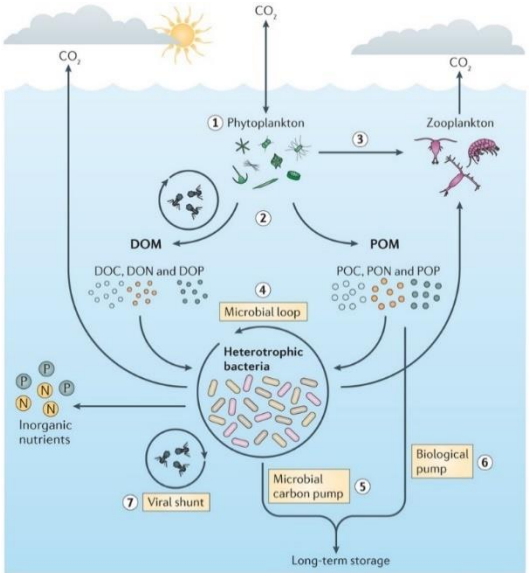


Figure 1. 2. Example highlighting the role of marine-derived DOM in the oceanic carbon cycle. (Adapted from Buchan, A., et.al, Nature Reviews Microbiology 2014, 12, 686-698)

1.2 Analytical methods for DOM characterization

Despite the fundamental role of DOM in aquatic ecosystems, and while many thousands of molecular species have been reported for DOM[26,20,27], the molecular structure of most components in this complex mixture remains largely unknown[12]. This knowledge gap is primarily because DOM compounds are highly variable in size, volatility, polarity, molecular structure, functionality, and elemental composition, leading to significant challenges in their separation and identification[27].

The characterization of DOM has been usually divided into two main approaches depending on how chemical constituents are analyzed. Bulk characterization is the most widely used method to obtain chemical information from DOM. In this approach, dissolved organic carbon concentration, usually combined with UV-visible and fluorescence, is used as a proxy to understand the chemical nature of a photosensitive fraction of DOM and study

its variability in diverse aquatic ecosystems[28-33]. Overall, a relatively high fraction of the DOM pool has been characterized using traditional bulk characterization methods. However, the degree of molecular information obtained has been somewhat limited. Therefore, advanced analytical approaches are needed to address the challenges associated with the complexity of the molecular level analysis of DOM[13,25,34,35]. Over the past 20 years, nuclear magnetic resonance (NMR) and mass spectrometry (MS) have been complementary analytical techniques for obtaining detailed information on the molecular fingerprint of DOM[25,26,36]. Non-targeted ^{1}H-NMR and ^{13}C-NMR, in particular, have been able to provide bulk compositional information (O/C and H/C ratios) and associate spectroscopic signals to structural features (aliphatic, carboxyl-rich alicyclic, carbohydrate-like, olefinic, and aromatic moieties) in marine and wetland (Figure 1.3) SPE-DOM samples[20,26].

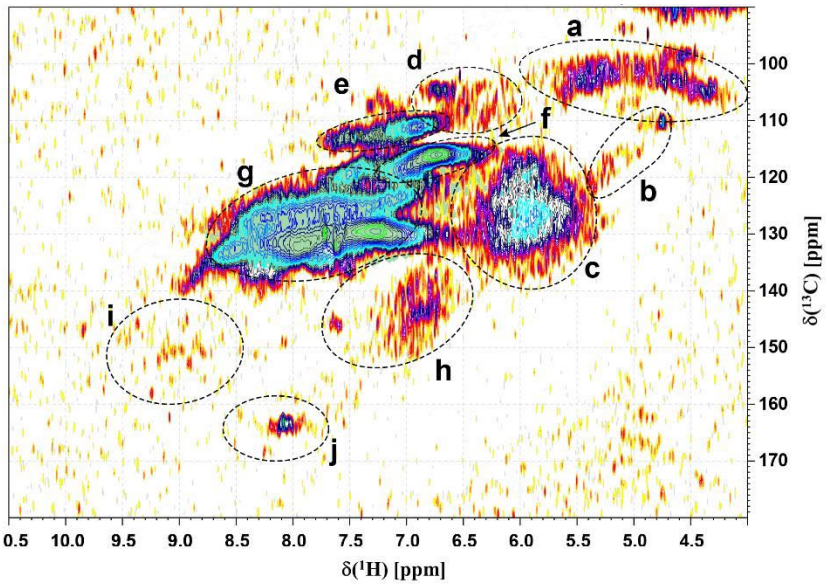


Figure 1. 3. Typical 2D ^{1}H-^{13}C NMR plot obtained from the analysis of a wetland SPE-DOM sample by high field NMR. Encircled areas describe signatures of different functionalities. For example: **a**: CH in carbohydrates; **b**: isolated olefins; c: oxygen heterocycles; **d**: C-conjugated olefins, certain five membered N-, O- and S-heterocycles;

e: phenols, oxygenated aromatics; **f**: likely COOH in NOM aromatics, **g**: double bonds adjacent to aromatics, **h**: nitrogen heterocycles, heteroatom polycyclic aromatics; and **i**: specific nitrogen heterocycles. (Adapted from Hertkorn, N., et.al., Biogeosciences, 2013, 10, 1583-1624.)

Although the identification of DOM structural motifs has been possible using NMR, the information is still insufficient for detecting, isolating, and quantifying single molecular species. Consequently, researchers have relied on MS-based approaches to improve the understanding of DOM molecular complexity.

1.2.1 Mass spectrometry

The development of mass spectrometry as an analytical technique revert to the early 1900's, when MS was employed to explore fundamental aspects of the atomic nature. By the mid-20th century, mass spectrometry instruments were accessible for academic and industrial applications, and more complex questions associated with chemical structures were tackled[37]. A mass spectrometer (Figure 1.4) comprises three main sections: i) an ion source for ion generation in the gas-phase, ii) an analyzer where ionized components are separated based on their characteristic mass-to-charge ratio (*m/z*), and iii) a detector system for recording the ion abundance[38]. A vacuum system along with ion guides and optic elements are required to ensure a free ion path through the instrument sections.

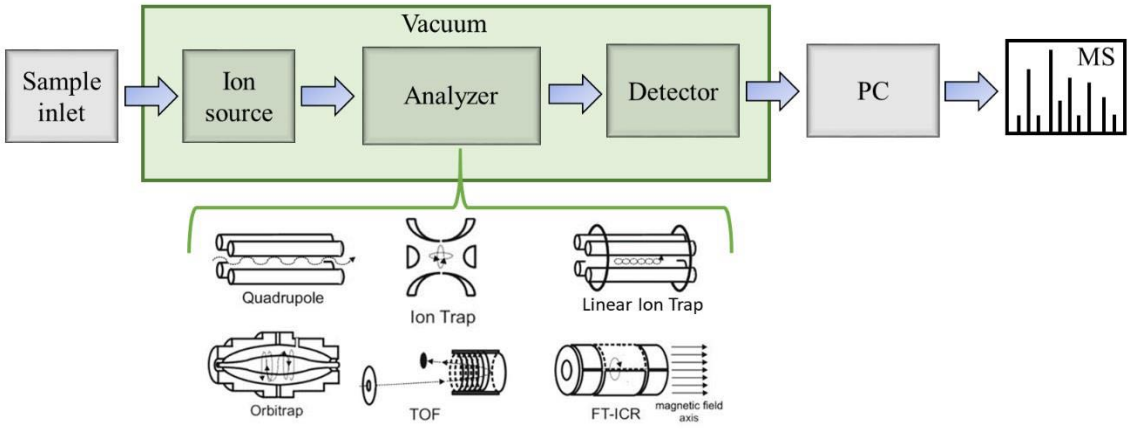


Figure 1. 4. Schematic diagram of a typical mass spectrometer. Different analyzers (quadrupole, ion trap, linear ion trap, time of flight, orbitrap, and ion cyclotron resonance cells) are also described. (Adapted from Schuchardt S., at.al., Protein identification using mass spectrometry: A method overview. In Plant Systems Biology. Birkhäuser Basel: 2007, 97:141-170.)

Owing to its high sensitivity, selectivity, versatility, and high throughput capabilities, MS instruments have found a wide variety of applications in analytical chemistry, such as proteomics, metabolomics, drug discovery, petroleomics, and environmental process monitoring[39-41]. The increasing demand for compound coverage, resolving power, and mass accuracy for the analysis of complex samples, has pushed a swift expansion of ion sources and the development of hybrid instruments integrating multiple analyzers (i.e., quadrupole-TOF[42,43], quadrupole-FT-ICR MS[44,45], among others). The development of different ion sources, especially the electrospray ionization (ESI) source[46,47], high-resolution time-of-flight mass spectrometer (TOF MS)[48,49], and Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS)[50-54] in the early 90's represented a remarkable advancement in the molecular level analysis of complex mixtures (i.e. DOM).

1.2.1.1 Electrospray ionization

Electrospray ionization (ESI) is a soft atmospheric desorption method suitable for DOM analysis due to its ability to generate positive and negative molecular ions from polar species over a wide range of molecular weights (up to 3000 *m/z*)[38,55]. ESI involves the transference of organic molecules in solution to ionized species in the gas phase. In ESI, an electric field generated by a voltage difference, usually below 6 KV, is applied between a capillary and a counter electrode to produce ions (Figure 1.5 left). With the assistance of a nebulizing gas, a spray of charged droplets accumulated at the solution surface, namely

the Taylor cone, is created at the electrospray tip. The remaining neutral solvent molecules are removed using heated nitrogen gas, and a pressure gradient push charged species into the analyzer region of the mass spectrometer. The ESI source can be easily coupled to the front end of different mass spectrometers, which represents a significant advantage. For example, commercial Ion Trap MS, TOF-MS, and FT-ICR MS Bruker Daltonics instruments include the Apollo II ESI (Figure 1.5 right) as one of the ion sources.

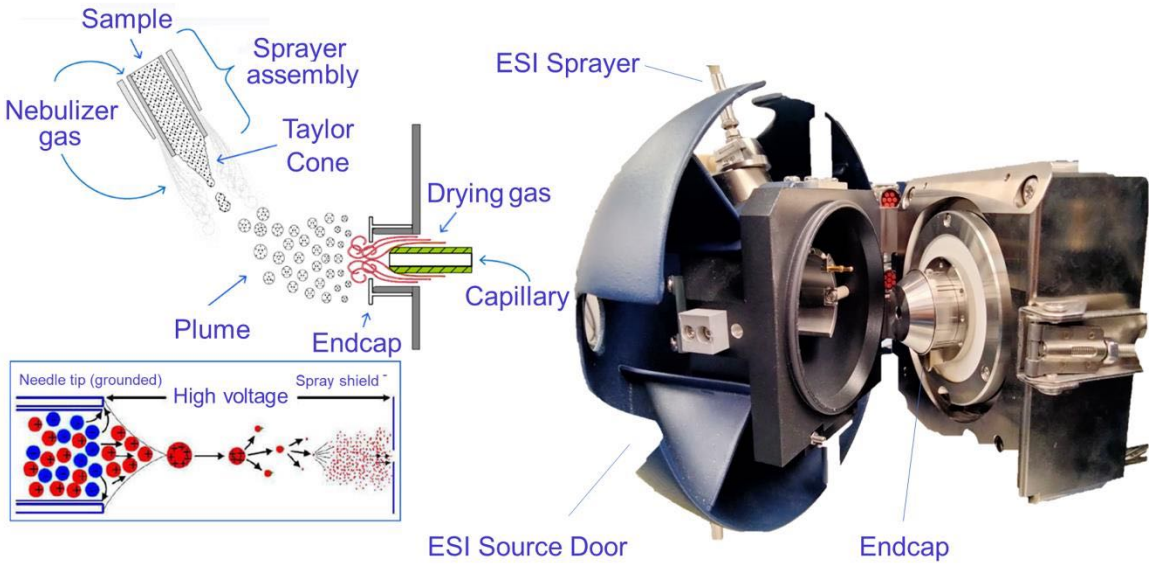


Figure 1. 5. Schematic diagram of a typical atmospheric pressure ESI source and simplified mechanism for the generation of ions during ESI (left). Typical Apollo II ESI source, Bruker Daltonics, Inc., MA (right). (Adapted from Bruker API Solarix manual. Copyright © 2010 Bruker Daltonics)

Other ionization sources such as atmospheric pressure photoionization (APPI) have been also used in combination with ESI to cover the ionization of a non-polar fraction of DOM hardly accessible by electrospray[56-58]. Nevertheless, ESI, coupled to TOF-MS and FT-ICR MS, is still the preferred ionization source for studies involving DOM molecular complexity.

1.2.1.2 Time-of-flight mass spectrometry

Time-of-flight analyzers were described and designed as early as in 1946. The TOF analyzer separates ionized species in a free-field region based on their specific velocities after an initial acceleration by an electric field[59]. Ions generated in the source are guided towards a flight tube by a potential gradient applied between an electrode and an extraction grid (Figure 1.6). As the ions gain the same kinetic energy, they can be separated according to their specific velocities. That is, lower mass ions will travel faster than heavier ions. In the TOF MS instrument, the *m/z* of an ion can be determined by measuring the time that ions take to travel through a field-free region between the source and the detector using the equation (1):

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2eV_s}\right) \qquad (1)$$

Where, t is the time needed to cover the distance L before reaching the detector and Vs is the acceleration potential applied.

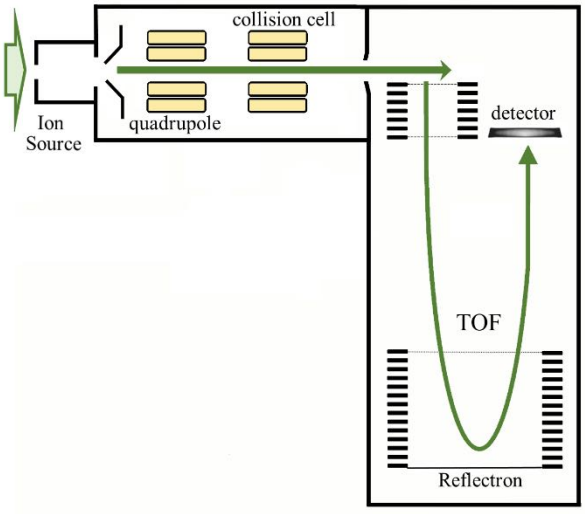


Figure 1. 6. Schematic diagram of a typical quadrupole-time of flight mass spectrometer. The quadrupole and collision cell sections allow for isolation and fragmentation capabilities.

Since 1996, TOF mass spectrometers have been hybridized with quadrupoles and collision cells to perform MS/MS experiments on pre-selected ion species[60]. Further instrument developments have led to significant improvements in mass resolution allowing for enhanced confidence during the molecular assignment. However, the application of TOF-MS in the analysis of complex mixtures, such as DOM, is still limited since ultrahigh-resolution is needed to resolve thousands of molecular species, for which several of them are within the same nominal mass.

1.2.1.3 Fourier transform ion cyclotron mass spectrometry

After being commercially available in the 1990s, FT-ICR MS has become a needed analytical tool for complex mixtures analysis due to its ultrahigh-resolution, high mass accuracy, and versatility in its integration with separation platforms.

In a typical FT-ICR MS instrument, ions generated in the source are first stored in a collision cell before their transference to the analyzer using a set of ion guides and optics. The m/z of an ion is measured from its detected characteristic cyclotron frequency while orbiting inside the ICR cell under the effect of a magnetic field (Figure 1.7).

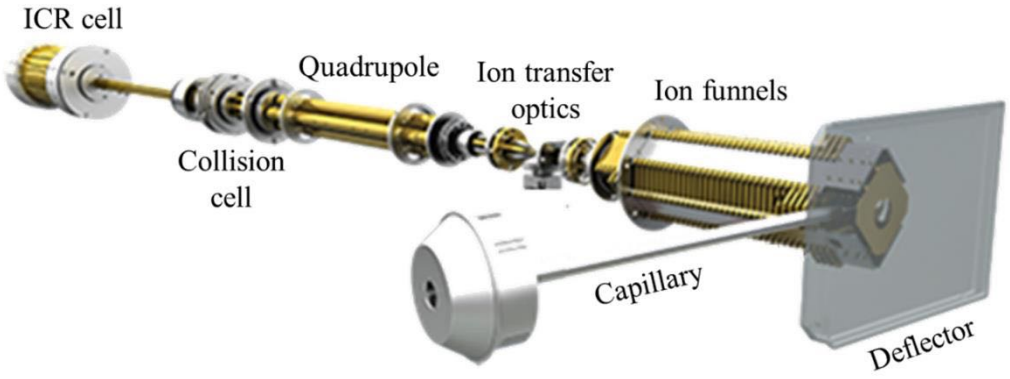


Figure 1. 7. Schematic representation of a typical instrumental components for a Bruker Solarix FT-ICR MS system (Adapted from F. Fernandez-Lima, in Fundamentals and Applications of Fourier Transform Mass Spectrometry, eds. B. Kanawati and P. Schmitt-Kopplin, Elsevier, 2019, 233-251).

The motion of an ion of mass *m* and charge *q* inside the ICR cell, with an applied magnetic field *B*, depends on the Lorentz force (Figure 1.8). The cyclotron angular frequency of the ion can be determined by the equation 2 if no electric field is considered[61-63].

$$\omega_c = \frac{qB}{m} \qquad (2)$$

Where, ω_c is the ion cyclotron orbital frequency of the ion and *B* is the strength of the magnetic field.

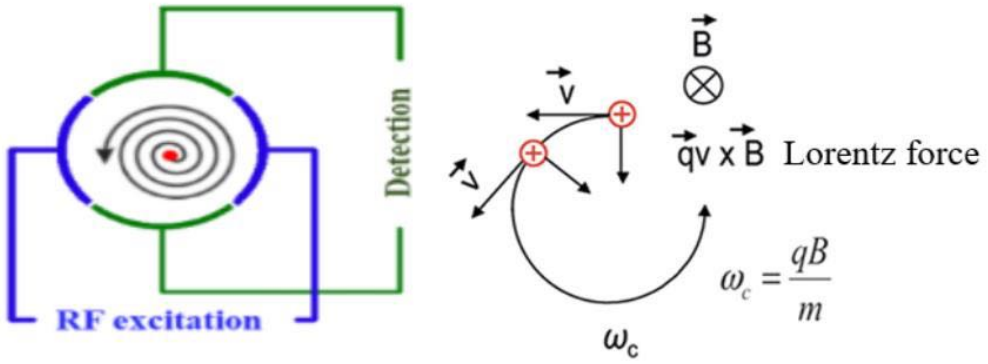


Figure 1. 8. Schematic diagram of the sequential excitation and detection of ions inside the ICR cell (left). Description of ion trajectory inside the ICR cell (right). Note that this representation does not include the influence of the trapping potential on the ion motion.

A quadrupolar potential is typically used in an ICR cell to trap the ions axially. Therefore, the ion motion depends on both a magnetic field and an electric field. The ion motion inside the ICR cell is defined by three components: i) an axial oscillation parallel to the magnetic field B (ω_z), ii) a cyclotron motion with reduced frequency ($\omega+$), and iii) a rotational motion perpendicular to B, described by the magnetron frequency ($\omega-$). In a typical experiment, $\omega+$ of an ion is measured based on the expression described in equation 3[62,64]:

$$\omega_+ = \frac{\omega_c}{2} + \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \frac{\omega_z^2}{2}} \qquad (3)$$

The ion analysis in the ICR cell depends on a sequence of events that starts with a broadband excitation followed by the detection of ion cyclotron frequencies. The recording of ω_+ in the form of induced current over a period of time is known as time-domain signal or transient. This transient is further converted into a reduced cyclotron frequency (ω^+) spectrum, applying a mathematical Fourier transform, and a final *m/z* spectrum is obtained by calibration equations (Figure 1.9)[65].

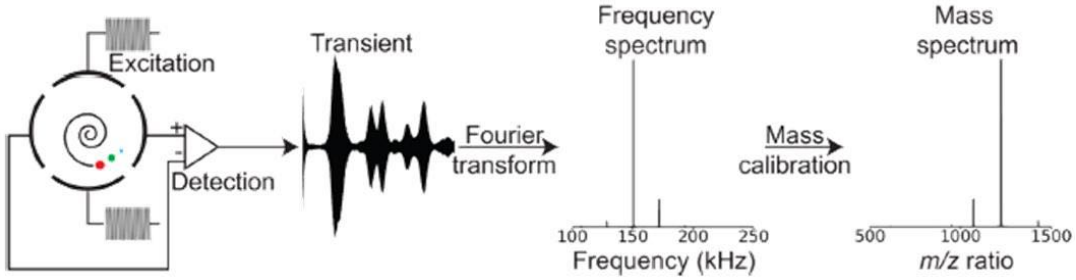


Figure 1. 9. Diagram of the acquisition and data processing of ion signals during a typical FT-ICR MS experiment. (Reprinted from M. A. van Agthoven, et al. *European Biophysics Journal*, 2019, 48, 213-229, with permission from Springer Nature under the terms of the Creative Commons CC BY license.)

The ability to accurately detect ion frequencies simulatenoulsy[63,66], and perform MS/MS experiments, make FT-ICR MS a powerful technique for the molecular analysis of complex mixtures. Tandem mass spectrometry (MS/MS) for example, is a useful strategy for obtaining structural information from the fragmentation of isolated ions signals. This technique is particularly advantageous when using FT-ICR MS since the isolation and fragmentation of precursor ions can be performed in multiple sections of the instrument and fragment ions are detected with high resolution.

1.2.1.4 FT-ICR MS/MS

In FT-ICR MS instruments, tandem MS/MS experiments can be performed in-space or in-time. A typical MS/MS experiment in-space comprises two sections of the instrument and starts with the selection of precursor molecular ions in the quadrupole after the ionization of the sample in the source. Quadrupoles are mass filters made of two pairs of parallel cylindrical or hyperbolic rods axially aligned[38]. Two opposite rods are subjected to direct currents (DC) of the same polarity and the others two, to radiofrequency (RF) voltages of equal charges (Figure 1.10)[67].

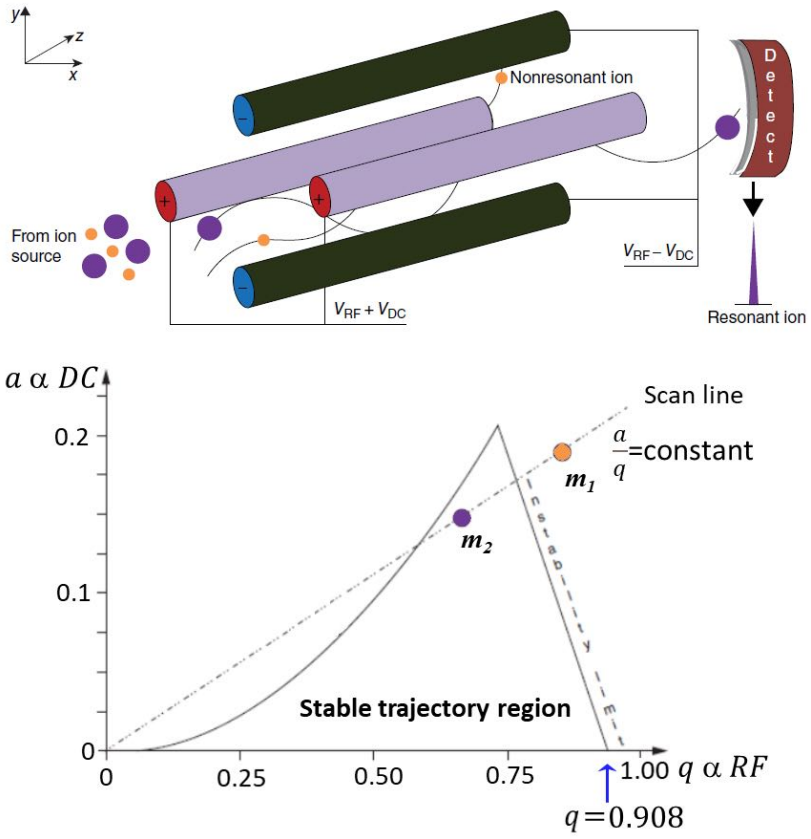


Figure 1. 10. Simplified diagram of a cylindrical quadrupole analyzer highlighting the selected ions that experience free path towards the detector (dark purple) and non-resonant charged species falling out of the stable trajectory (orange). (Adapted from Niessen, W.M.A., et.al., in *Interpretation of MS-MS Mass Spectra of Drugs and Pesticides*, 2017, 1-53)

The ion trajectory in a quadrupole depends on the solutions of the Mathiew's equations, simplified using the variable *a* and *q* which are directly correlated with the DC and RF voltages respectively, and inversely proportional to the ion mass. The plot *a* vs *q* described in Figure 1.10 (bottom) provides an example of a graphical representation for the stability path of two ions with different masses m_1 and m_2. That is, for selecting the ion with mass m_2 in the quadrupole, a stable trajectory needs to be assured by defining particular values for the parameters *a* and *q*. Ions with unstable paths (i.e. m_1 in Figure 1.10) will hit the quadrupole rods and be removed by the vacuum system[67].

The second step of MS/MS in-space involves the activations of preselected ions by applying a potential difference in the region between the quadrupole and the collision cell that increases ions kinetic energy. Activated ions are further dissociated in the collision cell as a result of ions-gas collisions before the detection of product ions in the ICR cell. This fragmentation method is known as collision induced dissociation (CID) and a simplified schematic representation of its principle is described in Figure 1.11. A typical in-space CID experiment the collision of isolated molecular ions with neutral gas molecules such as Argon, yields a cascade of fragment ions and neutral species resulting from covalent dissociation (i.e., $m_p^{n+} + N \rightarrow m_f^{n+} + m_n + N$). Note that other chemical reactions in the gas phase such as charge exchange, partial charge transfer, among others can also occurred[38].

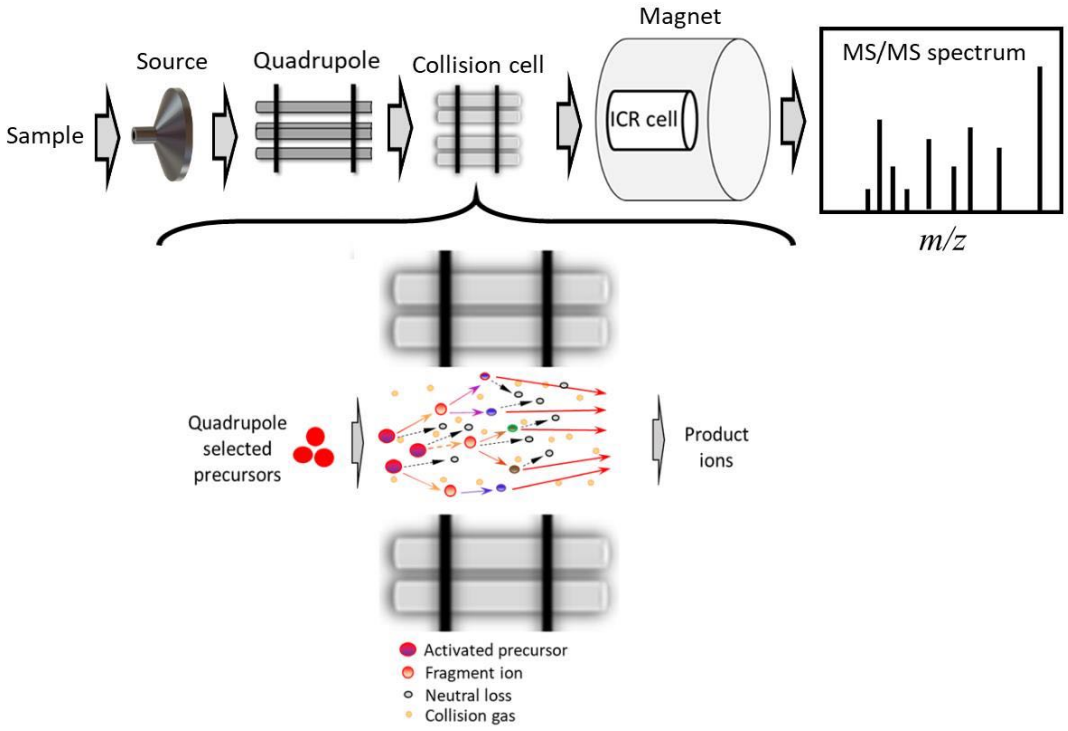


Figure 1. 11. Simplified diagram of a typical quadrupole-collision induced dissociation (CID) experiment performed in an FT-ICR spectrometer. After sample molecules are ionized in the source, the ions of interest (precursors) are selected in the quadruple. Precursors are then activated and subjected to collision with neutral gas molecules in the collision cell, resulting in a cascade of fragments (product ions) and neutral molecules (neutral losses). Product ions along with remaining non-fragmented precursors are detected in the ICR cell and an MS/MS spectrum is generated.

Different than in-space MS/MS experiments, in-time tandem FT-ICR MS is performed in the analyzer section of the instrument by sequentially trapping, selecting, activating, and dissociating the ions of interest. The capacity of manipulating trapped ions in the ICR using RF pulses is especially useful for the selection of ions prior to fragmentation. For example, ions of interest can be selected in the ICR cell by ejecting undesired species using correlated harmonic excitation field (CHEF). The application of a sweep excitation resonant to unwanted ions causes an increase in their cyclotron radii, surpassing the size of the ICR cell[68]. The equation 4 allows for the calculation of the excitation field E_{exc} with

frequency ω_{exc}, required to obtain a radius (r) higher than the effective cyclotron radius with frequency ω_{ion} in the ICR over a time T_{exc} and a magnetic field B.

$$r(ion) = \frac{E_{exc}}{B}\left|\frac{sin\{(\omega_{ion}-\omega_{exc})T_{exc}/2\}}{(\omega_{ion}-\omega_{exc})}\right| \qquad (4)$$

In typical ICR cells it has been demonstrated that ions at nominal mass can be effectively isolated using CHEF[68]. However, in most of the complex samples, several mass signals are detected at nominal mass, thus requiring the use of single-frequency excitation events to eject interferent ions from the ICR cell.

The CHEF technique is usually combined with sustained off resonance irradiation (SORI) for ion activation prior to fragmentation by CID, infrared multiphoton dissociation (IRMPD), among others. Briefly, in SORI activation, selected ions are subjected to excitation by applying a radiofrequency larger than the ion cyclotron frequency as described by equation 4. This allows ions to softly enhance their translational energy due to continuous acceleration-deceleration events prior to collision with a neutral gas[69]. In combination, CHEF, single ion ejection, and SORI-CID activation is a powerful tool that enables mDa mass selection, in-cell fragmentation, and detection at ultrahigh mass resolution. This may be especially advantageous in the molecular analysis of DOM due to its complexity at nominal mass. Nevertheless, the extreme chemical complexity of DOM (i.e., multiple isomers per single chemical formula) demands the exploration of novel analytical approaches that allow for the separation of individual species via liquid or gas phase prior to the analysis by ultra-high resolution mass spectrometry (UHRMS). Note that UHMS refers to mass spectrometers capable to achieve a mass resolving power (R) above 500,000[70], where R is defined as the capacity of an instrument to separate two mass signals

at mass *m* with the same intensity and spaced by Δm ($R = \frac{m}{\Delta m}$)[71]. Currently, Orbitrap MS and FT-ICR MS are the only two mass spectrometers included in the definition of UHRMS.

1.2.2 Analytical approaches for the molecular analysis of DOM

Overall, two analytical methodologies are currently the most common approaches utilized for the molecular characterization of DOM: (1) Traditional direct infusion of SPE DOM samples and analysis by UHRMS[26,72-78] and (2) online/offline high-performance liquid chromatography (HPLC) coupled to UHRMS[79-85].

1.2.2.1 Traditional offline SPE/ESI-FT-ICR MS for DOM analysis

The simple offline SPE method developed by Dittmar et.al[86] in the late 2000s for the isolation and concentration of most DOM components from water samples enabled the identification of thousands of molecular species in a single ESI-FT-ICR MS direct infusion experiment (Figure 1.12). Briefly, prefiltered water samples (0.45 μm) are acidified (pH~2), followed by pre-conditioning an SPE polymer-based cartridge suitable for the retention of hydrophobic polar compounds. Samples are loaded into the cartridge, and the cartridge is rinsed (pH 2 water) to remove salt impurities. In the final step, retained DOM components are eluted from the cartridge using methanol. Further studies on this procedure have led to an optimized and standardized protocol that minimizes deviations across laboratories enhancing the reproducibility of the overall DOM molecular analysis[87].

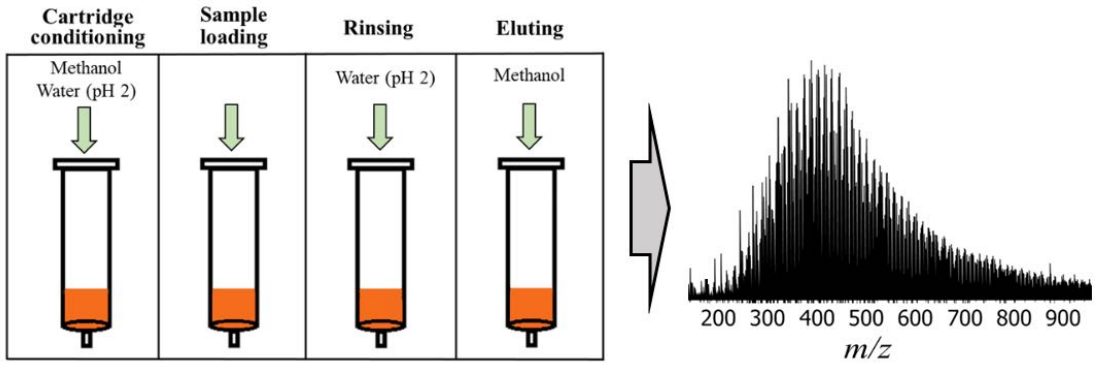


Figure 1. 12. Simplified schematic of the SPE protocol developed by Dittmar et.al, 2008 for the extraction of hydrophobic polar DOM components (left). Typical ESI (-)-FT-ICR MS spectrum of an SPE DOM sample depicting thousands of molecular ion species (right).

Although SPE preferentially concentrates DOM hydrophobic components within a range of polarities, it is the most efficient method to recover up to 90% of DOC in water samples. Moreover, this protocol is cheap and yields samples suitable for mass spectrometry and NMR analysis.

A typical direct infusion broadband ESI-FT-ICR MS of SPE-DOM sample results in a mass spectrum with thousands of singly charged signals uniformly spaced every 2 Da in an m/z range of 200-700. The presence of several isobaric peaks at each nominal mass, and multiple heteroatom compound classes, are commonly observed across DOM samples from different origins. This complex pattern has confirmed the notion of the intricate nature and wide chemo diversity of DOM constituents. Reports from traditional ultrahigh-resolution mass spectrometry experiments have suggested that a significant fraction of DOM is ubiquitous in aquatic ecosystems regardless of water origin and nature[12,26,88-90]. Despite all previous knowledge acquired on DOM molecular composition, recent studies have revealed that a DOM chemical formula may be a combination of multiple isomeric species unresolvable by the traditional ESI-FT-ICR MS approaches[12,27,82,88]. These results

have uncovered the limitations of current analytical capabilities and the necessity to develop novel analytical workflows capable of addressing the challenges associated with the isomeric complexity of DOM.

1.2.2.2 Offline/Online LC-ESI-FT-ICR MS for DOM analysis

Liquid chromatography mass spectrometry has been one of the few hyphenated alternatives utilized to investigate the complex isomeric diversity of DOM[81,82,91]. While exploring offline reverse-phase LC and hydrogen-deuterium exchange (HDX) in Suwanee River fulvic acid (SRFA) DOM standard, Stenson et *al*.[92] found that the same chemical compounds eluting in different LC fractions exhibited different maximum observable HDX, thus proving the existence of distinctive isomers. Indications of potential separation of isomeric species using reverse-phase-high resolution mass spectrometry[81,82,93-95] and even 2D-LC approaches, such as size exclusion chromatography (SEC)-reverse-phase (RP)[79,96] have been recently reported. However, limited resolving power observed in the LC domain has prevented the isolation of single isomers exhibiting very close retention times.

Han et *al*.[97] developed an LC-ESI-FT-ICR MS method using post-LC separation counter gradient and continuous accumulation of selected ions (CASI) to improve ionization efficiency, separation, and detection of highly polar species in DOM. Although more features were detected compared to direct infusion and traditional LC workflows, discrete isomers could not be separated and confirmed.

The results from traditional ultra-high resolution mass spectrometry and hyphenated chromatographic approaches confirmed that understanding the molecular complexity of DOM, separating its isomeric components, and ultimately identifying

structural information are challenging analytical problems that need novel instrumental development and efficient separation methodologies.

One potential alternative to address this analytical problem is the incorporation of separation frameworks alternative to the typical chromatographic approaches. That is the case of ion mobility spectrometry which separates molecules in the gas phase based on their size and shape at a millisecond time scale.

1.2.2.3 Ion mobility spectrometry (IMS)

The first experiments on ions' behavior in gases date back to the late 18th century. By the early 1900s, advances in the theory associated with ion-neutral interactions and instrumentation allowed rapid development of ion mobility spectrometry as an analytical technique with vast potential applications in industry, government, and academia.

In an IMS device (Figure 1.13), ions are injected into a drift region, filled with a buffer gas, under the influence of an electric field (E). While interacting with the bath gas, charged species traverse the drift region towards the detector with a drift velocity v_d[98,99]. An arrival time distribution, which provides information on the homogeneity of ion packages, is further recorded.

The drift velocity of an ion is a function of the time (t_d) ions require to cover a distance *d* between an ion shutter and a detector, using the equation (4).

$$v_d = \frac{d}{t_d} \qquad (4)$$

The ion mobility of an ion (*K*) is defined as the normalized ion drift velocity per applied electric field *E* unit (equation 5).

$$K = \frac{v_d}{E} \qquad (5)$$

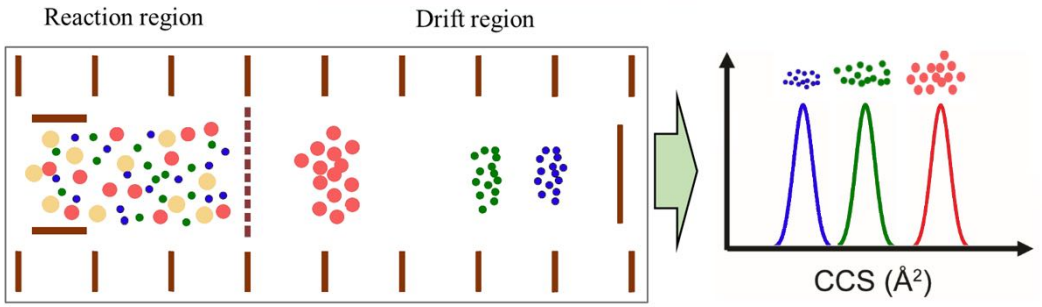


Figure 1. 13. Schematic diagram of a typical drift tube device. Ions (red, green, and blue) are injected in the IMS device and while traversing the drift region they collide with the buffer gas molecules (yellow). They are further separated based on their characteristic ion mobility and detected. (Adapted from Figure 1.2, Eiceman G.A., et.al., Ion Mobility Spectrometry. Third Edition ed.; CRC Press, Taylor & Francis: 2014)

Ion mobility measurements provide valuable information about the ion-neutral gas interactions experienced by an ion when it traverses the drift region. Therefore, *K* is a magnitude that depends on the collision frequency of the ion with gas molecules. This frequency is a function of the gas number-density (*N*, number of gas molecules per unit volume), and it is related to the gas pressure (*p*) and temperature (*T*). Thus, the reduced ion mobility (K_0) is the most common magnitude used in IMS experiments (equation 6)[98].

$$K_0 = K \cdot \frac{N}{N_0} = K \cdot \frac{p}{p_0} \cdot \frac{T_0}{T} \quad (6)$$

One of the advantages of ion mobility spectrometry is that ion-neutral collisions provide details on the conformational space of molecules in the gas phase. The collisional cross-section (CCS) is an intrinsic property of a molecule, and it is indicative of its size and shape under specific experimental conditions[98]. Overall, CCS characterizes the momentum transfer cross-section of ions when they interact with neutral gas molecules. Equation 7 depicts the mathematical expression used to calculate the experimental CCS values from K_0, assuming low-field ion mobility conditions[100].

$$CCS\ (\Omega, \text{Å}^2) = \frac{3}{16}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{ze}{N_0 K_0} \qquad (7)$$

Where, μ is the reduced mass of the ion-gas entity: $\mu = m_i m_g/(m_i + m_g)$, m_i is the mass of the ion and m_g is the mass of the gas, and z is the ion charge.

IMS methods are classified based on their measurement principles in non-linear and linear. In non-linear IMS workflows (i.e., differential IMS, DMS and Field Asymmetric Waveform Ion Mobility Spectrometry, FAIMS), the ion separation in the drift region depends on the change of K at different electric field values, so the dynamics describing ion motion is a non-linear function of the electric field[98]. On the contrary, in linear IMS methods, the ion mobility of ions does not depend on the applied electric field.

Drift tube ion mobility spectrometry (DTIMS), traveling wave ion mobility spectrometry (TWIMS), and trapped ion mobility spectrometry (TIMS) are examples of linear IMS methods. Both DTIMS and TWIMS are also considered time-dispersive devices, and their main difference resides in how the electric field is applied. The electric field in DTIMS is homogeneously distributed across the drift region, while being propagated sequentially in the form of symmetric waves in TWIMS[98,101] (Figure 1.14).

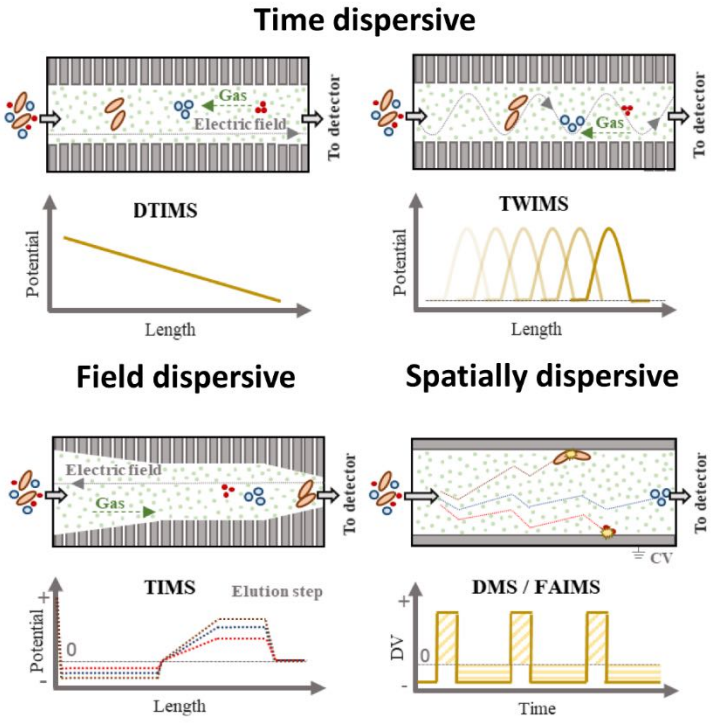


Figure 1. 14. Schematic representation of the different IMS methods and classification based on their measurement principles in time dispersive (DTIMS and TWINS), field dispersive (TIMS), and spatially dispersive (DMS/FAIMS). Note the differences in the way electric field is applied across devices. (Adapted from Hernández-Mesa M., et.al., Molecules, 2019, 24, 2706, 1-28)

1.2.2.3.1 Trapped Ion mobility spectrometry (TIMS)

Trapped ion mobility spectrometry, an IMS method developed by Fernandez Lima and Melvin Park in the late 2010s, introduced a novel separation philosophy different from the traditional IMS approaches. In TIMS, instead of using a stationary gas, ions are rather pushed by a buffer gas, and an electric field is applied to trap the ions so that the drag force exerted by the moving gas is compensated[102,103]. Ions are eluted from the TIMS analyzer region cell depending on their ion mobilities by ramping down the electric field.

This novel configuration (improved effective length and lower electric field than classical IMS devices) results in higher ion mobility resolution. Moreover, the increased

control over the TIMS operational parameters (e.g., scan rate and ion mobility range selection), allows for more flexibility towards coupling with different mass spectrometers. For example, relying on its fast analytical capabilities and gating configuration (Figure 1.15), TIMS has been successfully integrated with TOF[103,104] and FT-ICR[105,106] analyzers respectively.

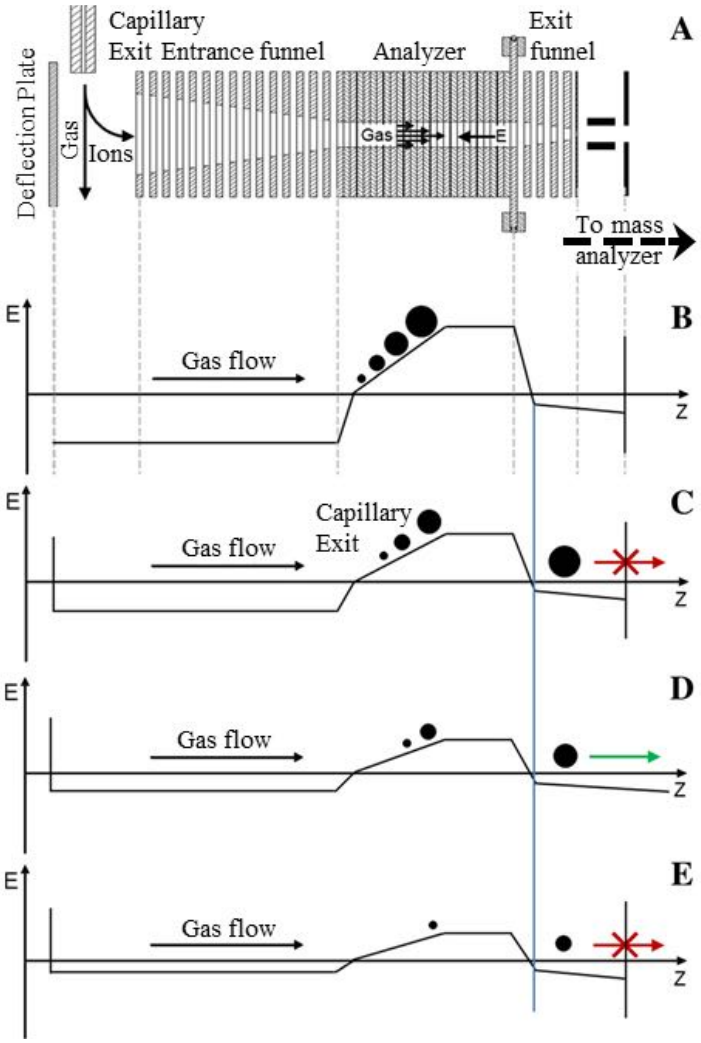


Figure 1. 15. Schematic diagram of the TIMS cell (A) and sequence of events (B-E) in a typical gated TIMS experiment. Ions of different sizes (black circles) traverse the device along the z-axis. Ions are trapped and separated in the TIMS analyzer section (B), and sequentially eluted and accumulated in the collision cell (C-E) based on their ion mobilities. (Adapted from Figure 1, Ridgeway M. E., et.al., *International Journal for Ion Mobility Spectrometry*, 2016, 19, 77-85).

Over the last decades, IMS-MS approaches for complex mixture analysis have been employed[107,108]. The separation of chemical classes by their IMS-MS trend lines and the use of CCS to assign potential structures are few of the described workflows. Overall, shorter analysis time, increased peak capacity, and chemical noise reduction have been some of the reported benefits. The recent development of high-resolution ion mobility analyzers has pushed its integration with ultrahigh-resolution mass analyzers for complex mixtures characterization[109-112]. In 2015, TIMS was integrated with FT-ICR MS[113] by the Fernandez Lima group, and several studies have shown the exceptional advantages of this platform[106,107,114-116]. Taking advantage of the extraordinary capabilities of this novel instrumentation, we hypothesized that integrating trapped ion mobility spectrometry with ESI-FT-ICR MS/MS will allow a better characterization of DOM at the molecular level.

1.3 Structure of the dissertation

Overall, this dissertation describes the development of a novel mass spectrometry toolbox based on TIMS-FT ICR MS for DOM characterization at the molecular level (Chapters 2-4) and environmental case studies, where these methods have been successfully applied (Chapters 5-6). In chapter two, we explored the advantages of using ESI-TIMS-FT ICR MS and nominal mass MS/MS to understand the isomeric complexity of DOM. We provided, for the first time, a lower and upper estimate of the number of isomers in a wetland SPE-DOM sample based on ion mobility information and nominal mass high-resolution fragmentation patterns. We were also able to propose candidate structures based on experimental-theoretical CCS match and in-silico fragmentation scores. Chapter three describes a developed TIMS-FT ICR MS/MS technology capable of ion mobility select (R~100) and fragment single isolated molecular ion signals

(Δm/z ~36mDa) in the ICR cell using single-shot ejections after broadband ejections and sustained off-resonance irradiation collision-induced dissociation (SORI-CID). This novel and challenging approach overcomes traditional challenges associated with the similarity of fragmentation patterns (e.g., common neutral losses of H_2O, CO_2, and CH_4O) by narrowing down the isomeric candidate structures using the ion mobility domain.

A novel unsupervised structural classification of DOM based on fragmentation pathways from comprehensive ESI-FT-ICR CASI-CID MS/MS is described in chapter four. We dissected the fragmentation data obtained from over one hundred nominal masses (261-477 m/z range) using a conceptual model based on neutral mass loss patterns (Pn-1+F1:n+C), implemented in the Python code Graph-DOM. This novel approach enabled the identification of over 1,000 DOM families of structurally related precursor molecules that shared analogous fragmentation pathways. We also found a high degree of isomerism in DOM (numerous identical fragmentation pathways), which is impossible to discriminate with sole precursor ion analysis. A Cytoscape® map of interconnected DOM structural families is also shown for visualizing potential biogeochemical processes of DOM. This CASI-CID workflow is a powerful tool capable of tracking structural changes resulting from biological and geochemical transformations of DOM.

Chapter five describes the first-time application of ESI-TIMS-FT ICR MS to examine DOM molecular complexity across a freshwater-to-estuarine coastal transient of the Florida Everglades. We were able to identify an average estimate of six and up to 12 isomers per chemical formula and characteristic isomeric signals to each section of the freshwater-to-estuarine gradient. We also observed a decrease in DOM chemical complexity and diversity (both in the number of molecular formulas and number of isomers

per chemical formula) with increasing salinity. These trends illustrate the significance of biogeochemical transformations DOM experiences during its transport along a salinity transect and the role of various sources in shaping DOM molecular fingerprint. The isomeric content analysis uncovered a set of species (mainly lignin-type components) that resist transformation and are responsible for the DOM refractory nature. Overall, we demonstrated that TIMS-FT ICR MS is a valuable technique for DOM molecular fingerprinting capable of distinguishing isomeric variabilities along freshwater-to-estuarine gradients.

In chapter six we compare the structural signatures of four SPE-DOM samples (Suwanee River Fulvic Acid standard (SRFA), Pantanal, and Harney River 1 and 5) from different sources, using CASI-CID MS/MS and ESI-TIMS-FT ICR MS. We uncovered clear structural dissimilarities across the DOM samples in the order SRFA>Pantanal>HR-5>HR-1 based on fragmentation pathways and structural families. However, our tool was also able to recognize structural similarities across samples based on 200 shared families. A dissection of neutral loss sequences connecting precursors of shared families evidenced ubiquitous hydration and carboxylation transformational processes across ecosystems. The analysis of the isomeric content across common families showed a new layer of similarities and differences that could be used for further fingerprinting purposes.

1.4 References

1. Schulze-Makuch DirkLouis, N. I. Building Blocks of Life. In *Life in the Universe. Advances in Astrobiology and Biogeophysics*, Springer, Berlin, Heidelberg: 2005; 3, 77-78.

2. Keenan, T. F.; Williams, C. A. The Terrestrial Carbon Sink. *Annual Review of Environment and Resources* **2018,** *43* (1), 219-243.

3. Grace, J. Understanding and managing the global carbon cycle. *Journal of Ecology* **2004,** *92* (2), 189-202.

4. Ridgwell, A.; Arndt, S. Chapter 1 - Why Dissolved Organics Matter: DOC in Ancient Oceans and Past Climate Change. In *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)*, Hansell, D. A.; Carlson, C. A., Eds.; Academic Press: Boston, 2015; 1-20.

5. Amon, R. M. W. Ocean dissolved organics matter. *Nature Geoscience* **2016,** *9* (12), 864-865.

6. Lønborg, C.; Carreira, C.; Jickells, T.; Álvarez-Salgado, X. A. Impacts of Global Change on Ocean Dissolved Organic Carbon (DOC) Cycling. *Frontiers in Marine Science* **2020,** *7*, 2-24.

7. Sexton, P. F.; Norris, R. D.; Wilson, P. A.; Pälike, H.; Westerhold, T.; Röhl, U.; Bolton, C. T.; Gibbs, S. Eocene global warming events driven by ventilation of oceanic dissolved organic carbon. *Nature* **2011,** *471* (7338), 349-352.

8. Kaplan, L. A.; Cory, R. M. Chapter 6 - Dissolved Organic Matter in Stream Ecosystems: Forms, Functions, and Fluxes of Watershed Tea. In *Stream Ecosystems in a Changing Environment*, Jones, J. B.; Stanley, E. H., Eds.; Academic Press: Boston, 2016; pp 241-320.

9. Asmala, E.; Massicotte, P.; Carstensen, J. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnology and Oceanography* **2021,** *66* (4), 1381-1393.

10. Buffam, I.; Turner, M. G.; Desai, A. R.; Hanson, P. C.; Rusak, J. a.; Lottig, N. r.; Stanley, E. H.; Carpenter, S. R. Integrating aquatic and terrestrial components to construct a complete carbon budget for a north temperate lake district. *Global Change Biology* **2011,** *17* (2), 1193-1211.

11. Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Analytical and Bioanalytical Chemistry* **2007,** *389* (5), 1311-1327.

12. Dittmar, T.; Stubbins, A. 12.6 - Dissolved Organic Matter in Aquatic Systems. In *Treatise on Geochemistry (Second Edition)*, Holland, H. D.; Turekian, K. K., Eds.; Elsevier: Oxford, 2014; 125-156.

13. Repeta, D. J. Chapter 2 - Chemical Characterization and Cycling of Dissolved Organic Matter. In *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)*, Hansell, D. A.; Carlson, C. A., Eds.; Academic Press: Boston, 2015; 21-63.

14. Tremblay, L.; Gagné, J.-P. Organic matter distribution and reactivity in the waters of a large estuarine system. *Marine Chemistry* **2009,** *116* (1), 1-12.

15. Hedges, J. I. Global biogeochemical cycles: progress and problems. *Marine Chemistry* **1992,** *39* (1), 67-93.

16. Dittmar, T. Chapter 7 - Reasons Behind the Long-Term Stability of Dissolved Organic Matter. In *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)*, Hansell, D. A.; Carlson, C. A., Eds.; Academic Press: Boston, 2015; 369-388.

17. Tranvik, L. J.; Downing, J. A.; Cotner, J. B.; Loiselle, S. A.; Striegl, R. G.; Ballatore, T. J.; Dillon, P.; Finlay, K.; Fortino, K.; Knoll, L. B.; Kortelainen, P. L.; Kutser, T.; Larsen, S.; Laurion, I.; Leech, D. M.; McCallister, S. L.; McKnight, D. M.; Melack, J. M.; Overholt, E.; Porter, J. A.; Prairie, Y.; Renwick, W. H.; Roland, F.; Sherman, B. S.; Schindler, D. W.; Sobek, S.; Tremblay, A.; Vanni, M. J.; Verschoor, A. M.; von Wachenfeldt, E.; Weyhenmeyer, G. A. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography* **2009,** *54* (6part2), 2298-2314.

18. Fenchel, T. The microbial loop – 25 years later. *Journal of Experimental Marine Biology and Ecology* **2008,** *366* (1), 99-103.

19. Hiriart-Baer, V. P.; Smith, R. E. H. The effect of ultraviolet radiation on freshwater planktonic primary production: The role of recovery and mixing processes. *Limnology and Oceanography* **2005,** *50* (5), 1352-1361.

20. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016,** *13* (8), 2257-2277.

21. Traina, S. J.; McAvoy, D. C.; Versteeg, D. J. Association of Linear Alkylbenzenesulfonates with Dissolved Humic Substances and Its Effect on Bioavailability. *Environmental Science and Technology* **1996,** *30* (4), 1300-1309.

22. Moeckel, C.; Monteith, D. T.; Llewellyn, N. R.; Henrys, P. A.; Pereira, M. G. Relationship between the Concentrations of Dissolved Organic Matter and Polycyclic Aromatic Hydrocarbons in a Typical U.K. Upland Stream. *Environmental Science and Technology* **2014,** *48* (1), 130-138.

23. Abdulla, H. A. N.; Dias, R. F.; Minor, E. C. Understanding the enhanced aqueous solubility of styrene by terrestrial dissolved organic matter using stable isotope mass balance and FTIR. *Organic Geochemistry* **2009,** *40* (5), 547-552.

24. Zeng, T.; Arnold, W. A. Clustering Chlorine Reactivity of Haloacetic Acid Precursors in Inland Lakes. *Environmental Science and Technology* **2014,** *48* (1), 139-148.

25. Minor, E. C.; Swenson, M. M.; Mattson, B. M.; Oyler, A. R. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environmental Science: Processes & Impacts* **2014,** *16* (9), 2064-2079.

26. Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **2013,** *10* (3), 1583-1624.

27. Zark, M.; Christoffers, J.; Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Marine Chemistry* **2017,** *191*, 9-15.

28. Wang, C.; Li, Y.; Li, Y.; Zhou, H.; Stubbins, A.; Dahlgren, R. A.; Wang, Z.; Guo, W. Dissolved Organic Matter Dynamics in the Epipelagic Northwest Pacific Low-Latitude Western Boundary Current System: Insights from optical analyses. *Journal of Geophysical Research: Oceans* **2021,** *126* (9), 2-18.

29. Fellman, J. B.; Hood, E.; Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnology and Oceanography* **2010,** *55* (6), 2452-2462.

30. Yamashita, Y.; Boyer, J. N.; Jaffé, R. Evaluating the distribution of terrestrial dissolved organic matter in a complex coastal ecosystem using fluorescence spectroscopy. *Continental Shelf Research* **2013,** *66*, 136-144.

31. Hansen, A. M.; Kraus, T. E. C.; Pellerin, B. A.; Fleck, J. A.; Downing, B. D.; Bergamaschi, B. A. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. *Limnology and Oceanography* **2016,** *61* (3), 1015-1032.

32. Zhang, Y.; Zhou, Y.; Shi, K.; Qin, B.; Yao, X.; Zhang, Y. Optical properties and composition changes in chromophoric dissolved organic matter along trophic gradients: Implications for monitoring and assessing lake eutrophication. *Water Research* **2018,** *131*, 255-263.

33. Gonçalves-Araujo, R.; Stedmon, C. A.; Heim, B.; Dubinenkov, I.; Kraberg, A.; Moiseev, D.; Bracher, A. From Fresh to Marine Waters: Characterization and Fate of Dissolved Organic Matter in the Lena River Delta Region, Siberia. *Frontiers in Marine Science* **2015,** *2* (108), 1-13.

34. Nebbioso, A.; Piccolo, A. Molecular characterization of dissolved organic matter (DOM): a critical review. *Analytical Bioanaytical Chemistry* **2013,** *405* (1), 109-124.

35. Cooper, W. T.; Chanton, J. C.; D'Andrilli, J.; Hodgkins, S. B.; Podgorski, D. C.; Stenson, A. C.; Tfaily, M. M.; Wilson, R. M. A History of Molecular Level Analysis of Natural Organic Matter by FTICR Mass Spectrometry and The Paradigm Shift in Organic Geochemistry. *Mass Spectrometry Reviews* **2020**, 1-25.

36. Kim, S.; Kim, D.; Jung, M.-J.; Kim, S. Analysis of environmental organic matters by Ultrahigh-Resolution mass spectrometry—A review on the development of analytical methods. *Mass Spectrometry Reviews* **2021**, 1-18.

37. Griffiths, J. A Brief History of Mass Spectrometry. *Analytical Chemistry* **2008,** *80* (15), 5678-5683.

38. de Hoffmann, E.; Stroobant, V. *Mass Spectrometry: Principles and Applications*; John Wiley & Sons Ltd: West Sussex, England, 2013; 512.

39. Awad, H.; Khamis, M. M.; El-Aneed, A. Mass Spectrometry, Review of the Basics: Ionization. *Applied Spectroscopy Reviews* **2015,** *50* (2), 158-175.

40. Uttam Singh Baghel, A. S., Deeksha Singh and Manish. Application of Mass Spectroscopy in Pharmaceutical and Biomedical Analysis. In *Spectroscopic Analyses - Developments and Applications*, University, U. A.-Q., Ed. IntechOpne: 2017.

41. Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the analytical limits: new insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. *Chemical Science* **2019,** *10* (29), 6966-6978.

42. Ens, W.; Standing, K. G. Hybrid Quadrupole/Time-of-Flight Mass Spectrometers for Analysis of Biomolecules. In *Methods in Enzymology*, Academic Press: 2005; 402 pp 49-78.

43. Tucholska, M.; Scozzaro, S.; Williams, D.; Ackloo, S.; Lock, C.; Siu, K. W. M.; Evans, K. R.; Marshall, J. G. Endogenous peptides from biophysical and biochemical fractionation of serum analyzed by matrix-assisted laser desorption/ionization and electrospray ionization hybrid quadrupole time-of-flight. *Analytical Biochemistry* **2007,** *370* (2), 228-245.

44. O'Connor, P. B.; Pittman, J. L.; Thomson, B. A.; Budnik, B. A.; Cournoyer, J. C.; Jebanathirajah, J.; Lin, C.; Moyer, S.; Zhao, C. A new hybrid electrospray Fourier transform mass spectrometer: design and performance characteristics. *Rapid Communications in Mass Spectrometry* **2006,** *20* (2), 259-266.

45. Patrie, S. M.; Charlebois, J. P.; Whipple, D.; Kelleher, N. L.; Hendrickson, C. L.; Quinn, J. P.; Marshall, A. G.; Mukhopadhyay, B. Construction of a hybrid quadrupole/fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (7), 1099-1108.

46. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *246* (4926), 64-71.

47. Banerjee, S.; Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry* **2012,** *2012*, 1-40.

48. Cotter, R. J. Time-of-Flight Mass Spectrometry. In *Time-of-Flight Mass Spectrometry*, American Chemical Society: 1993; 549, 16-48.

49. Kaufmann, A. Chapter 4 - High Mass Resolution Versus MS/MS. In *Comprehensive Analytical Chemistry*, Fernandez-Alba, A. R., Ed. Elsevier: 2012; 58, 169-215.

50. Senko, M. W.; Hendrickson, C. L.; Paša-Tolić, L.; Marto, J. A.; White, F. M.; Guan, S.; Marshall, A. G. Electrospray Ionization Fourier Transform Ion Cyclotron Resonance at 9.4 T. *Rapid Communications in Mass Spectrometry* **1996,** *10* (14), 1824-1828.

51. Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D. H.; Marshall, A. G. External accumulation of ions for enhanced electrospray ionization Fourier transform ion

cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1997,** *8* (9), 970-976.

52. Fievre, A.; Solouki, T.; Marshall, A. G.; Cooper, W. T. High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Humic and Fulvic Acids by Laser Desorption/Ionization and Electrospray Ionization. *Energy & Fuels* **1997,** *11* (3), 554-560.

53. Stenson, A. C.; Landing, W. M.; Marshall, A. G.; Cooper, W. T. Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2002,** *74* (17), 4397-4409.

54. Kujawinski, E. B. Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS): Characterization of Complex Environmental Mixtures. *Environmental Forensics* **2002,** *3* (3), 207-216.

55. Cole, R. B. Some tenets pertaining to electrospray ionization mass spectrometry. *Journal of Mass Spectrometry* **2000,** *35* (7), 763-772.

56. D'Andrilli, J.; Dittmar, T.; Koch, B. P.; Purcell, J. M.; Marshall, A. G.; Cooper, W. T. Comprehensive characterization of marine dissolved organic matter by Fourier transform ion cyclotron resonance mass spectrometry with electrospray and atmospheric pressure photoionization. *Rapid Communications in Mass Spectrometry* **2010,** *24* (5), 643-650.

57. Podgorski, D. C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Cooper, W. T. Selective Ionization of Dissolved Organic Nitrogen by Positive Ion Atmospheric Pressure Photoionization Coupled with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2012,** *84* (11), 5085-5090.

58. He, C.; Fang, Z.; Li, Y.; Jiang, C.; Zhao, S.; Xu, C.; Zhang, Y.; Shi, Q. Ionization selectivity of electrospray and atmospheric pressure photoionization FT-ICR MS for petroleum refinery wastewater dissolved organic matter. *Environmental Science: Processes & Impacts* **2021,** *23* (10), 1466-1475.

59. Kaklamanos, G.; Aprea, E.; Theodoridis, G. 11 - Mass spectrometry: principles and instrumentation. In *Chemical Analysis of Food (Second Edition)*, Pico, Y., Ed. Academic Press: 2020; 525-552.

60. Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry* **2001,** *36* (8), 849-865.

61. Comisarow, M. B.; Marshall, A. G. Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters* **1974,** *25* (2), 282-283.

62. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* **1998,** *17* (1), 1-35.

63. Marshall, A. G.; Hendrickson, C. L. Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *International Journal of Mass Spectrometry* **2002,** *215* (1), 59-75.

64. Cho, E.; Witt, M.; Hur, M.; Jung, M.-J.; Kim, S. Application of FT-ICR MS Equipped with Quadrupole Detection for Analysis of Crude Oil. *Analytical Chemistry* **2017,** *89* (22), 12101-12107.

65. Zhang, L.-K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. Accurate mass measurements by Fourier transform mass spectrometry. *Mass Spectrometry Reviews* **2005,** *24* (2), 286-309.

66. Marshall, A. G.; Grosshans, P. B. Fourier transform ion cyclotron resonance mass spectrometry: the teenage years. *Analytical Chemistry* **1991,** *63* (4), 215A-229A.

67. Niessen, W. M. A. a. C. C., R.A. Introduction To LC–MS Technology. In *Interpretation of MS-MS Mass Spectra of Drugs and Pesticides*, 2017; pp 1-53.

68. de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass selection of ions in a Fourier transform ion cyclotron resonance trap using correlated harmonic excitation fields (CHEF). *International Journal of Mass Spectrometry and Ion Processes* **1997,** *165-166*, 209-219.

69. Sleno, L.; Volmer, D. A. Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry* **2004,** *39* (10), 1091-1112.

70. D'Andrilli, J.; Cooper, W. T.; Foreman, C. M.; Marshall, A. G. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Communications in Mass Spectrometry* **2015,** *29* (24), 2385-2401.

71. Murray, K. K.; Boyd, R. K.; Eberlin, M. N.; Langley, G. J.; Li, L.; Naito, Y. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry* **2013,** *85* (7), 1515-1609.

72. Mopper, K.; Stubbins, A.; Ritchie, J. D.; Bialk, H. M.; Hatcher, P. G. Advanced Instrumental Approaches for Characterization of Marine Dissolved Organic Matter: Extraction Techniques, Mass Spectrometry, and Nuclear Magnetic Resonance Spectroscopy. *Chemical Reviews* **2007,** *107* (2), 419-442.

73. Sleighter, R. L.; Hatcher, P. G. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *Journal of Mass Spectrometry* **2007,** *42* (5), 559-574.

74. Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E. M. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Analytical Chemistry* **2008,** *80* (23), 8908-8919.

75. Witt, M.; Fuchser, J.; Koch, B. P. Fragmentation Studies of Fulvic Acids Using Collision Induced Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2009,** *81* (7), 2688-2694.

76. D'Andrilli, J.; Foreman, C. M.; Marshall, A. G.; McKnight, D. M. Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. *Organic Geochemistry* **2013,** *65*, 19-28.

77. Lu, Y.; Li, X.; Mesfioui, R.; Bauer, J. E.; Chambers, R. M.; Canuel, E. A.; Hatcher, P. G. Use of ESI-FTICR-MS to Characterize Dissolved Organic Matter in Headwater Streams Draining Forest-Dominated and Pasture-Dominated Watersheds. *PLOS ONE* **2016,** *10* (12), 1-21.

78. Hawkes, J. A.; Dittmar, T.; Patriarca, C.; Tranvik, L.; Bergquist, J. Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Analytical Chemistry* **2016,** *88* (15), 7698-7704.

79. Brown, T. A.; Jackson, B. A.; Bythell, B. J.; Stenson, A. C. Benefits of multidimensional fractionation for the study and characterization of natural organic matter. *Journal of Chromatography A* **2016,** *1470*, 84-96.

80. Koch, B. P.; Ludwichowski, K.-U.; Kattner, G.; Dittmar, T.; Witt, M. Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Marine Chemistry* **2008,** *111* (3), 233-241.

81. Patriarca, C.; Bergquist, J.; Sjöberg, P. J. R.; Tranvik, L.; Hawkes, J. A. Online HPLC-ESI-HRMS Method for the Analysis and Comparison of Different Dissolved Organic Matter Samples. *Environmental Science and Technology* **2018,** *52* (4), 2091-2099.

82. Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J. R.; Tranvik, L. J.; Bergquist, J. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnology and Oceanography Letters* **2018,** *3* (2), 21-30.

83. Liu, Z.; Sleighter, R. L.; Zhong, J.; Hatcher, P. G. The chemical changes of DOM from black waters to coastal marine waters by HPLC combined with ultrahigh resolution mass spectrometry. *Estuarine, Coastal and Shelf Science* **2011,** *92* (2), 205-216.

84. Li, Y.; Harir, M.; Lucio, M.; Gonsior, M.; Koch, B. P.; Schmitt-Kopplin, P.; Hertkorn, N. Comprehensive structure-selective characterization of dissolved organic matter by reducing molecular complexity and increasing analytical dimensions. *Water Research* **2016,** *106*, 477-487.

85. Chon, K.; Chon, K.; Cho, J. Characterization of size fractionated dissolved organic matter from river water and wastewater effluent using preparative high performance size exclusion chromatography. *Organic Geochemistry* **2017,** *103*, 105-112.

86. Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* **2008,** *6* (6), 230-235.

87. Li, Y.; Harir, M.; Lucio, M.; Kanawati, B.; Smirnov, K.; Flerus, R.; Koch, B. P.; Schmitt-Kopplin, P.; Hertkorn, N. Proposed Guidelines for Solid Phase Extraction of Suwannee River Dissolved Organic Matter. *Analytical Chemistry* **2016,** *88* (13), 6680-6688.

88. Zark, M.; Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nature Communications* **2018,** *9* (1), 1-8.

89. Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a major refractory component of marine dissolved organic matter. *Geochimica et Cosmochimica Acta* **2006,** *70* (12), 2990-3010.

90. Lam, B.; Baer, A.; Alaee, M.; Lefebvre, B.; Moser, A.; Williams, A.; Simpson, A. J. Major Structural Components in Freshwater Dissolved Organic Matter. *Environmental Science and Technology* **2007,** *41* (24), 8240-8247.

91. Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science* **2017,** *4* (405), 1-14.

92. Stenson, A. C.; Ruddy, B. M.; Bythell, B. J. Ion molecule reaction H/D exchange as a probe for isomeric fractionation in chromatographically separated natural organic matter. *International Journal of Mass Spectrometry* **2014,** *360*, 45-53.

93. Sandron, S.; Davies, N. W.; Wilson, R.; Cardona, A. R.; Haddad, P. R.; Nesterenko, P. N.; Paull, B. Fractionation of Dissolved Organic Matter on Coupled Reversed-Phase Monolithic Columns and Characterisation Using Reversed-Phase Liquid Chromatography-High Resolution Mass Spectrometry. *Chromatographia* **2018,** *81* (2), 203-213.

94. Capley, E. N.; Tipton, J. D.; Marshall, A. G.; Stenson, A. C. Chromatographic Reduction of Isobaric and Isomeric Complexity of Fulvic Acids to Enable Multistage Tandem Mass Spectral Characterization. *Analytical Chemistry* **2010,** *82* (19), 8194-8202.

95. Li, L.; Zhuo, X.; Fang, Z.; Wang, X. Investigation of the molecular structure complexity of dissolved organic matter by UPLC-orbitrap MS/MS. *Talanta* **2021,** *230*, 1-7.

96. Spranger, T.; Pinxteren, D. v.; Reemtsma, T.; Lechtenfeld, O. J.; Herrmann, H. 2D Liquid Chromatographic Fractionation with Ultra-high Resolution MS Analysis Resolves a Vast Molecular Diversity of Tropospheric Particle Organics. *Environmental Science and Technology*. **2019,** *53* (19), 11353-11363.

97. Han, L.; Kaesler, J.; Peng, C.; Reemtsma, T.; Lechtenfeld, O. J. Online Counter Gradient LC-FT-ICR-MS Enables Detection of Highly Polar Natural Organic Matter Fractions. *Analytical Chemistry* **2021,** *93* (3), 1740-1748.

98. Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; De Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan Jr, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T. Recommendations for reporting ion mobility Mass Spectrometry measurements. *Mass Spectrometry Reviews* **2019,** *38* (3), 291-320.

99. G.A. Eiceman • Z. Karpas • H.H. Hill, J. *Ion Mobility Spectrometry*; Third Edition ed.; CRC Press, Taylor & Francis: 2014.

100. Revercomb, H. E.; Mason, E. A. Theory of plasma chromatography/gaseous electrophoresis. Review. *Analytical Chemistry* **1975,** *47* (7), 970-983.

101. Hernández-Mesa, M.; Ropartz, D.; García-Campaña, A. M.; Rogniaux, H.; Dervilly-Pinel, G.; Le Bizec, B. Ion Mobility Spectrometry in Food Analysis: Principles, Current Applications and Future Trends. *Molecules* **2019,** *24* (15), 2706.

102. Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A. Gas-phase separation using a trapped ion mobility spectrometer. *International Journal for Ion Mobility Spectrometry* **2011,** *14* (2), 93-98.

103. Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Review of Scientific Instruments* **2011,** *82* (12), 126106.

104. Ridgeway, M. E.; Bleiholder, C.; Mann, M.; Park, M. A. Trends in trapped ion mobility – Mass spectrometry instrumentation. *TrAC Trends in Analytical Chemistry* **2019,** *116*, 324-331.

105. Ridgeway, M. E.; Wolff, J. J.; Silveira, J. A.; Lin, C.; Costello, C. E.; Park, M. A. Gated trapped ion mobility spectrometry coupled to fourier transform ion cyclotron resonance mass spectrometry. *International Journal for Ion Mobility Spectrometry* **2016,** *19* (2), 77-85.

106. Benigni, P.; Porter, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Increasing Analytical Separation and Duty Cycle with Nonlinear Analytical Mobility Scan Functions in TIMS-FT-ICR MS. *Analytical Chemistry* **2018,** *90* (4), 2446-2450.

107. Tose, L. V.; Benigni, P.; Leyva, D.; Sundberg, A.; Ramírez, C. E.; Ridgeway, M. E.; Park, M. A.; Romão, W.; Jaffé, R.; Fernandez-Lima, F. Coupling trapped ion mobility spectrometry to mass spectrometry: trapped ion mobility spectrometry–time-of-flight mass spectrometry versus trapped ion mobility spectrometry–Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communication in Mass Spectrometry* **2018,** *32* (15), 1287-1295.

108. Lu, K.; Gardner, W. S.; Liu, Z. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). *Environmental Science and Technology* **2018,** *52* (13), 7182-7191.

109. Robinson, E. W.; Leib, R. D.; Williams, E. R. The role of conformation on electron capture dissociation of ubiquitin. *Journal of The American Society for Mass Spectrometry* **2006,** *17* (10), 1470-1479.

110. Robinson, E. W.; Sellon, R. E.; Williams, E. R. Peak deconvolution in high-field asymmetric waveform ion mobility spectrometry (FAIMS) to characterize macromolecular conformations. *International Journal of Mass Spectrometry* **2007,** *259* (1), 87-95.

111. Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. Petroleum Crude Oil Characterization by IMS-MS and FTICR MS. *Analytical Chemistry* **2009,** *81* (24), 9941-9947.

112. Fasciotti, M.; Lalli, P. M.; Klitzke, C. F.; Corilo, Y. E.; Pudenzi, M. A.; Pereira, R. C. L.; Bastos, W.; Daroda, R. J.; Eberlin, M. N. Petroleomics by Traveling Wave Ion Mobility–Mass Spectrometry Using CO2 as a Drift Gas. *Energy & Fuels* **2013,** *27* (12), 7277-7286.

113. Benigni, P.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Targeted High-Resolution Ion Mobility Separation Coupled to Ultrahigh-Resolution Mass Spectrometry of Endocrine Disruptors in Complex Mixtures. *Analytical Chemistry* **2015,** *87* (8), 4321-4325.

114. Benigni, P.; Fernandez-Lima, F. Oversampling Selective Accumulation Trapped Ion Mobility Spectrometry Coupled to FT-ICR MS: Fundamentals and Applications. *Analytical Chemistry* **2016,** *88* (14), 7404-7412.

115. Pu, Y.; Ridgeway, M. E.; Glaskin, R. S.; Park, M. A.; Costello, C. E.; Lin, C. Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **2016,** *88* (7), 3440-3443.

116. Benigni, P.; Sandoval, K.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Gardinali, P.; Fernandez-Lima, F. Analysis of Photoirradiated Water Accommodated Fractions of Crude Oils Using Tandem TIMS and FT-ICR MS. *Environmental Science and Technology* **2017,** *51* (11), 5978-5988.

CHAPTER II

II. UNDERSTANDING THE STRUCTURAL COMPLEXITY OF DISSOLVED ORGANIC MATTER: ISOMERIC DIVERSITY

(Reproduced from Leyva *et al*., 2019, Faraday Discussions, with permission of the Royal Society of Chemistry)

2.1 Abstract

In the present work, the advantages of ESI-TIMS-FT-ICR MS to address the isomeric content of dissolved organic matter are studied. While the MS spectra allowed the observation of a high number of peaks (e.g., PAN-L: 5004 and PAN-S: 4660), over 4x features were observed in the IMS-MS domain (e.g., PAN-L: 22 015 and PAN-S: 20 954). Assuming a total general formula of CxHyN0–3O0–19S0–1, 3066 and 2830 chemical assignments were made in a single infusion experiment for PAN-L and PAN-S, respectively. Most of the identified chemical compounds (_80%) corresponded to highly conjugated oxygen compounds (O_1–O_{20}). ESI-TIMS-FT-ICR MS provided a lower estimate of the number of structural and conformational isomers (e.g., an average of 6–10 isomers per chemical formula were observed). Moreover, ESI-q-FT-ICR MS/MS at the level of nominal mass (i.e., 1 Da isolation) allowed for further estimation of the number of isomers based on unique fragmentation patterns and core fragments; the later suggested that multiple structural isomers could have very closely related CCS. These studies demonstrate the need for ultrahigh resolution TIMS mobility scan functions (e.g., R = 200–500) in addition to tandem MS/MS isolation strategies.

2.2 Introduction

Dissolved organic matter (DOM) is a highly complex mixture of organic compounds that is ubiquitous in aquatic ecosystems, resulting mainly from the degradation of aquatic and terrestrial primary producers[1]. It is mainly composed of carbon, hydrogen, and oxygen, with the other elements being at relatively lower abundance. The biogeochemical functions of natural DOM are extremely important because of its influence on many environmental processes, including fate and transport of contaminants, ecological processes and water

treatment.[1] Despite the important role of DOM in global carbon cycling, and while tens of thousands of molecular formulas have been reported in DOM[2,3], and many structural features identified [4], the molecular structure of most components in this complex mixture remains largely unknown.[5] This is primarily due to the fact that DOM components are highly variable in volatility, polarity, molecular structure, functionality and elemental composition, leading to serious challenges in their separation and identification.[3]

However, the combination of multiple analytical approaches2,6 and the utilization of advance analytical techniques have moved this field forward. In particular, Fourier transform ion cyclotron Resonance-Mass Spectrometry (FT-ICR MS) and Quadrupole Time-of-Flight Mass Spectrometry (Q-TOF-MS) have aided much in the characterization of DOM due to their high-resolution capabilities and flexibility toward coupling with separation techniques. While FT-ICR MS has been widely and successfully used to assess the molecular complexity of DOM, limitations with regards to isomer characterization, an important aspect of DOM complexity, still remain. A recent report focused on characterizing DOM complexity and composition in a highly variable set of DOM samples using FT-ICR MS in combination with advanced statistical methods[7], confirmed the notion that a significant component of DOM seems to be molecularly indistinguishable between samples and is thus ubiquitous in the environment.[8] Not only the co-occurrence of thousands of identical molecular formulae, but also, a remarkable similarity of fragment ion intensities among samples, and thus molecular structure commonalities, were reported. Using a modeling approach, the authors estimated the isomers associated with the large number of identified molecular formulas. However, constraining isomerization aspects in DOM characterization continues to be challenging, such information might be most

accurately achieved by Ion Mobility Spectrometry (IMS) in tandem with mass spectrometry.[9]

During the last decades, several attempts have been made to utilize Ion Mobility Spectrometry (IMS) in tandem with mass spectrometry for the analysis of complex mixtures.[9] A common trend is towards the possibility to separate chemical classes by their IMS-MS trend lines, measurement of ion-neutral collision cross sections, shorter analysis time, easy coupling to other separation techniques (e.g., gas and liquid chromatography), increased peak capacity and reduction of the chemical noise. With the advent of high-resolution mobility analyzers (R>80), there is a natural push for their integration to high resolution mass analyzer for the analysis of complex mixtures. [10-19] Our team has pioneered the integration of TIMS with FT-ICR MS since 2015[20], and several reports have shown the unique advantages of TIMS-FT-ICR MS.[9,21-27]

In the present work, we discuss the advantages and current challenges during ESI-TIMS-FT-ICR MS/MS analysis of complex mixtures. The goal is to address the analytical advantages of ESI-TIMS-FT-ICR MS and ESI-q-FT-ICR MS/MS for two freshwater DOM samples in assessing their isomeric diversity and future challenges provided from MS/MS experiments at nominal mass.

2.3 Experimental

2.3.1 Sample preparation

Surface water was collected from Pantanal (PAN) National Park – SE Brazil, one of the largest subtropical and biodiverse freshwater wetlands in the world. The PAN samples were collected from the Paraguay River (PAN-L) and a wetland channel in Pantanal National Park (PAN-S). For further details on sampling and sample preparation, see ref. 2.

The DOM and the individual standards were dissolved in 50:50 v/v methanol/water to a final concentration of 1 ppm. Prior to analysis, all samples were spiked with 5% (v/v) of the Tuning Mix calibration standard. All solvents used were of Optima LC-MS grade or better, obtained from Fisher Scientific (Pittsburgh, PA).

2.3.2 Sample ionization

An electrospray ionization (ESI) source based on the Apollo II ESI design (Bruker Daltonics, Inc., MA) was used in negative ion mode for all experiments. Sample solutions were introduced into the nebulizer at a rate of 360 μLh^{-1} using a syringe pump. Typical operating conditions were 3000–3500 V capillary voltage, 10 $Lmin^{-1}$ dry gas flow rate, 1.0 bar nebulizer gas pressure, and a dry gas temperature 180 °C.

2.3.3 Trapped ion mobility spectrometry analysis

The concept behind TIMS is the use of an electric field to hold ions stationary against a moving gas, so that the drag force is compensated by the electric field and ion packages are separated across the TIMS analyzer axis based on their mobility.[28-30] During mobility separation, a quadrupolar field confines the ions in the radial direction to increase trapping efficiency. The mobility, K, of an ion in a TIMS cell is described by:

$$K = \frac{v_g}{E} \cong \frac{A}{(V_{elution} - V_{out})} \qquad (1)$$

where v_g, E, $V_{elution}$ and V_{out} are the velocity of the gas, applied electric field, elution voltage and tunnel out voltage, respectively. Mobility spectra were calibrated using a Tuning Mix calibration standard (Tunemix, G2421A, Agilent Technologies, Santa Clara, CA) with the following reduced mobility (Ko) values *m/z* 301 K_0=1.909, *m/z* 601 K_0=1.187, *m/z* 1033 K_0=0.776¸ *m/z* 1333 K_0=0.710 cm2 $V^{-1}s^{-1}$. [31,32]

Mobility values (*K*) can be correlated with the ion-neutral collision cross section (Ω, Å^2) using the Mason-Schamp equation:

$$\Omega = \frac{(18\pi)^{1/2}}{16}\frac{z}{(k_B T)^{1/2}}\left(\frac{1}{m_I}+\frac{1}{m_b}\right)^{1/2}\frac{1}{K}\frac{760}{P}\frac{T}{273.15}\frac{1}{N^*} \qquad (2)$$

where *z* is the charge of the ion, k_B is the Boltzmann constant, N^* is the number density and m_I and m_b refer to the masses of the ion and bath gas, respectively.[33]

2.3.4 ESI-TIMS-FT-ICR MS/MS analysis

All experiments were performed on a custom built ESI-TIMS-q-F-ICR MS 7T Solarix spectrometer equipped with an infinity ICR cell (Bruker Daltonics Inc., MA). The TIMS analyzer is controlled using in-house software, written in National Instruments Lab VIEW, and synchronized with the 7T Solarix FT-ICR MS acquisition program. TIMS separation was performed using nitrogen as a bath gas at *ca.* 300 K, P_1 = 2.2 and P_2 = 0.9 mbar, and a constant *rf* (2200 kHz and 140-160 Vpp). A nonlinear stepping scan function was used,[27] with a gate width of 3 ms. The TIMS cell was operated using a fill/trap/elute/quench sequence 9/3/9/3 ms, using an average of 1000 IMS scans per MS spectrum and a voltage difference across the ΔE gradient of 5.0 V. The ramp voltage gradient was stepped by 0.25 V/frame with a ΔV_{ramp} range of -160 to -60, for a total of 400 steps. The deflector (V_{def}), funnel entrance (V_{fun}), analyzer base voltage (V_{out}) and gating lens (V_{gate}) voltages were V_{def} = -180/180V, Vfun = -90V, V_{out} = -50 V and V_{gate} = -80V/80V. TIMS-FT-ICR MS spectra were processed using sine-squared apodization followed by Fast Fourier Transform (FFT), in magnitude mode resulting in an experimental MS resolving power of R ~ 400,000 at *m/z* 400. ESI-q-FT-ICR MS/MS experiments were performed using quadrupole isolation at nominal mass and typical CID energies of 15-20 eV.

2.3.5 Data processing

The ESI-TIMS-FT-ICR MS spectra were externally calibrated for mass and mobility using the Agilent ESI-L mass calibration standard. The formulae calculations from the exact mass domain were performed using Composer software (Version 1.0.6, Sierra Analytics, CA) and confirmed with Data Analysis (Bruker Daltonics v 4.2) using formula limits of $C_xH_yN_{0-3}O_{0-19}S_{0-1}$, and odd and even electron configurations were allowed. The TIMS spectrum for each molecular formula was processed using a custom-built Software Assisted Molecular Elucidation (SAME) package – a specifically designed 2D TIMS-MS data processing script written in Python v2.7.[34] SAME package utilizes noise removal, mean gap filling, "asymmetric least squares smoothing" base line correction, peak detection by continuous wavelet transform (CWT)-based peak detection algorithm (SciPy package), and Gaussian fitting with non-linear least squares functions (Levenberg-Marquardt algorithm). SAME final outcome is [*m/z*; chemical formula; K; CCS] for each TIMS-MS dataset. The 2D TIMS-MS contour plots were generated in Data Analysis (Version v. 5.1, Bruker Daltonics, CA) and all the other plots were generated using matplotlib and OriginPro 2016 (Originlab Co., MA). The MetFrag CL software was used for in silico determination of potential candidate structures using the PubChem database.[35]

2.4 Results and discussion

2.4.1 ESI-TIMS-FT-ICR MS analysis

The analysis of the PAN complex dissolved organic matter using ESI-TIMS-FT-ICR MS resulted in a single, broad trend line in the IMS-MS domain composed of singly charged species (Figure 2.1). Inspection of the MS domain leads to the observation of a

similar profile of a single, broad gaussian distribution centered around m/z 400, regardless of the sample.

Closer inspection of the MS spectra allowed the comparison of the number of MS peaks (e.g., PAN-L: 5,004 and PAN-S: 4,660), with the number of IMS-MS features (e.g., PAN-L: 22,015; PAN-S: 20,954). Assuming a total general formula of $C_xH_yN_{0-3}O_{0-19}S_{0-1}$, we found 3,066 and 2,830 for PAN-L and PAN-S compounds, respectively. Most of the identified chemical compounds (~80%) corresponded to highly conjugated oxygen compounds (O_1-O_{20}), in good agreement with previous reports[36]. This complexity can be visualized at the level of nominal mass (see example in Figure 2.1) for a 391 m/z.

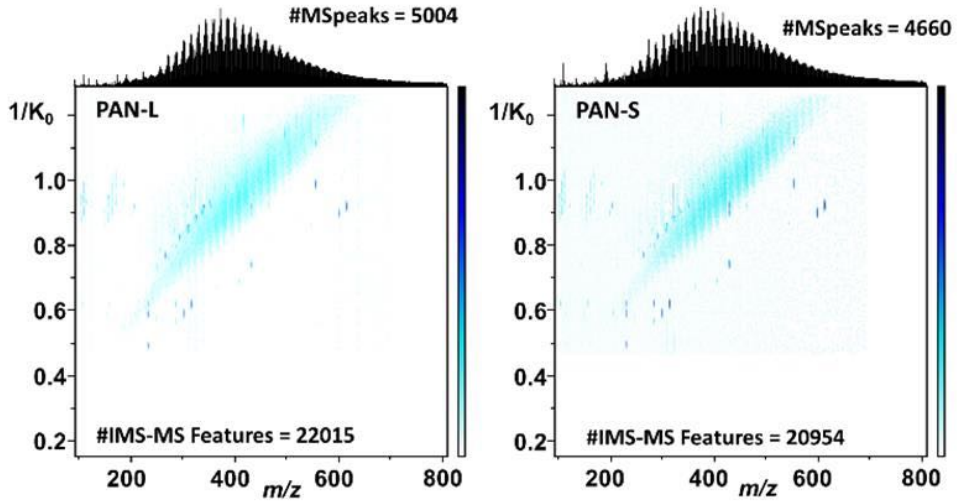


Figure 2. 1. Typical 2D-IMS-MS contour plots for the case of the PAN-L and PAN-S complex dissolved organic matter.

2.4.2 ESI-q-FT-ICR MS analysis

While a large isomeric diversity is observed at the level of nominal mass and per chemical formula, complementary information on the nature of the sample constituents can be obtained by performing tandem MS/MS. At the level of nominal mass, several m/z

signals are observed (e.g., over 7 at 391 m/z). When -subjected to CID, several common neutral losses are observed (see Figure 2.3 and Appendices 2.1-2.3).

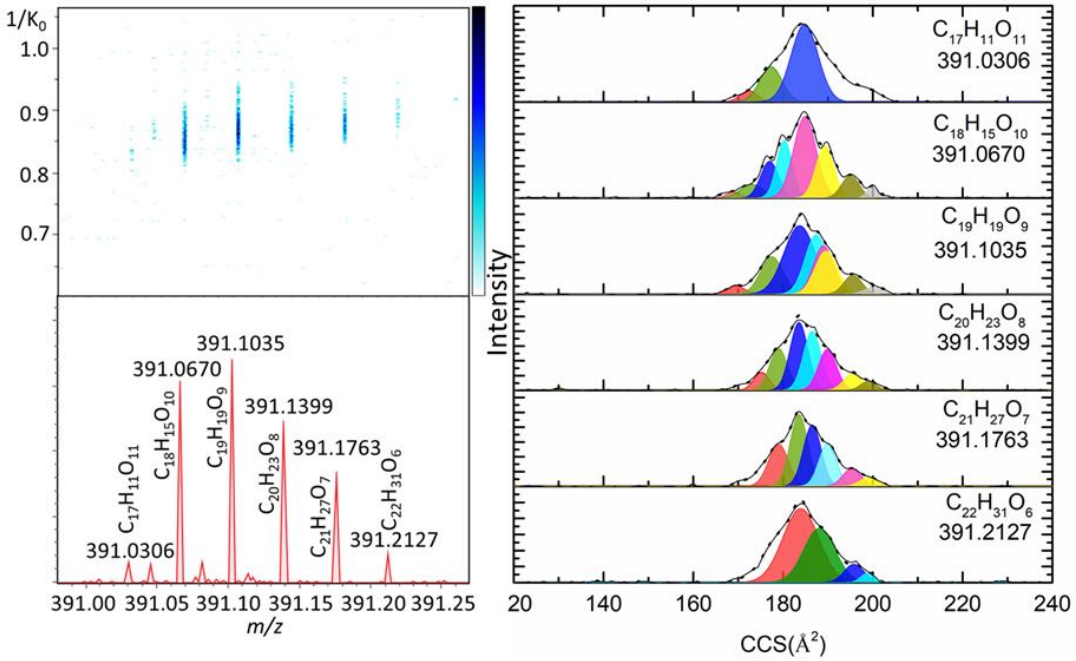


Figure 2. 2. Typical 2D-IMS-MS, as well the MS and IMS projections at nominal mass (i.e., 391*m/z*). Different bands are annotated in the IMS projections based on the SAME algorithm.

If we assume that the neutral losses can be directly associated with functional groups and the overall structure of the parent ion, a number of potential structural isomers can be estimated for a given chemical formula; under this assumption, conformational isomers will present the same fragmentation pathway and are not considered. For example, CO_2 can be associated with carboxyl groups and H_2O loss with the presence of hydroxyl groups. In addition, we observed the CO, CH_2, and CH_4 neutral losses (see Appendix 2.2 for all neutral loss fragments observed), in good agreement with previous FT-ICR MS/MS reports.[37] Taking advantage of the high mass accuracy of the FT-ICR MS measurements, neutral loss assignments can be easily identified. For example, the fragmentation pathways

for the 391.1031 m/z ($C_{19}H_{19}O_9$, Appendix 2.3) were generated utilizing the fragmentation data obtained at nominal mass (Appendix 2.1) and all possible combinations of neutral loss fragments (Appendix 2.2) with a mass tolerance error of 1 mDa.

Duplicate fragmentation pathways with same syntaxes were eliminated (e.g., $2CH_2$-$3CO$ is the same as $3CO$-$2CH_2$), since sequential fragmentation was not performed. Inspection of the fragmentation pathway shows a total of 16 end core fragments, each of them with multiple neutral loss pathways (see Appendices 2.1 and 2.3).

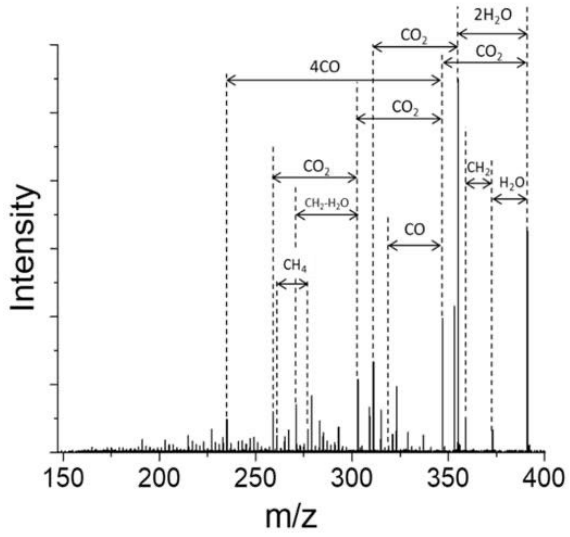


Figure 2. 3. A typical FT-ICR MS/MS spectrum from a 391 m/z precursor ion isolated at nominal mass and subjected to CID prior to injection in the ICR cell.

Since each pathway denotes the number and type of functional groups that were lost during fragmentation, the number of pathways could provide an upper estimate of the number of structural isomers. For instance, $3CO_2$-$2CO$-$2CH_4$ is one of the fragmentation pathways ending in the core formula $C_{12}H_{11}O$ (*m/z* 171.0814). That is, parent ion ($C_{19}H_{19}O_9$, 391.1031) presumably experimented losses of three carboxylic groups, two carbonyl groups and two methane groups, suggesting that one isomer structure contains an

arrangement of these functional groups. Conversely, for the same ending core formula ($C_{12}H_{11}O$), another fragmentation pathway involved consecutive losses of two hydroxyls, one carboxyl, two methylenes and four carbonyls ($2H_2O$-CO_2-$2CH_2$-4CO), indicating the presence of a different structural isomer.

Table 2. 1. Core fragments and number of neutral loss pathways observed for 391.1031 m/z ($C_{19}H_{19}O_9$) during FT-ICR-MS/MS with isolation at nominal mass.

Precursor ion *m/z*	Core Fragment *m/z*	Structural isomers
391.1031 $C_{19}H_{19}O_9$	161.0607 $C_{10}H_9O_2$	13
	163.0763 $C_{10}H_{11}O_2$	7
	165.0192 $C_8H_5O_4$	3
	165.0560 $C_9H_9O_3$	2
	167.0349 $C_8H_7O_4$	1
	171.0814 $C_{12}H_{11}O$	23
	173.0607 $C_{11}H_9O_2$	23
	175.0400 $C_{10}H_7O_3$	15
	183.0450 $C_{12}H_7O_2$	40
	183.0814 $C_{13}H_{11}O$	25
	185.0607 $C_{12}H_9O_2$	29
	187.0400 $C_{11}H_7O_3$	25
	201.0192 $C_{11}H_5O_4$	25
	202.9984 $C_{10}H_3O_5$	15
	205.0140 $C_{10}H_5O_5$	7
	241.0140 $C_{13}H_5O_5$	7

A parallel analysis performed using *in silico* fragmentation of the of 391.1031 *m/z* ($C_{19}H_{19}O_9$) with the MetFrag CL software across PubChem, that included the MS/MS information at nominal mass resulted in 96 hits (see Appendix 2.4). That is, 96 candidate structures were obtained based on accurate mass of the precursor and fragment ions with 1mDa mass tolerance.

While the ESI-q-FT-ICR MS/MS analysis with nominal mass quadrupole isolation is suggested as a rapid way to estimate an upper limit of the structural diversity and complexity of DOM, it is important to consider, that because of the isolation was only performed at the level of nominal mass, potential overestimation of the number of

pathways is possible due to rearrangements of the fragments during CID. That is, interferences from fragments from other isobaric parent ions with similar chemical composition (i.e., $C_cH_hO_o$) may be a limitation in this approach (see Figure 2.2). Nevertheless, the data summarized in Table 2.1 suggest the presence of up to 260 structural isomers. When compared to IMS data and MetFrag output, we can speculate that there are multiple structural isomers that share the same IMS band (only seven band separated by the SAME algorithm).

2.5 Conclusions

In the present work we illustrated the advantages of ESI-TIMS-FT-ICR MS/MS to address the isomeric content of DOM. The MS analysis permitted the identification of chemical components based on mass accuracy. When complemented with IMS measurements, an estimate of structural and conformational isomers can be obtained (e.g., an average of 6-10 isomers were observed). While the MS spectra allowed the observation of a high number of peaks (e.g., PAN-L: 5004 and PAN-S: 4660), over 4x features were observed in IMS-MS domain (e.g., PAN-L: 22 015; PAN-S: 20 954). Assuming a total general formula of $C_xH_yN_{0-3}O_{0-19}S_{0-1}$, 3066 and 2830 for PAN-L and PAN-S chemical assignments were found in a single infusion experiment, respectively. Most of the identified chemical compounds (~80%) corresponded to highly conjugated oxygen compounds (O_1-O_{20}). Moreover, when ESI-q-FT-ICR MS/MS is performed at the level of nominal mass, further estimation of the number of structural isomers is possible based on unique neutral loss fragmentation patterns and core fragments. The data provided shows that multiple structural isomers could have very closely related CCS, which will demand the use of ultrahigh resolution TIMS mobility scan functions in tandem with MS/MS.

Future studies can further push the analytical boundaries of ESI-TIMS-FT-ICR MS by mobility selective ESI-TIMS-FT-ICR MS/MS and applying correlated harmonic excitation field (CHEF)[37] on the quadrupole 1Da isolated parent ions.

2.6 References

1. Kaplan, L. A.; Cory, R. M. Dissolved organic matter in stream ecosystems. In *Stream Ecosystems in a Changing Climate*, Stanley, J. J. a. E., Ed. Academic Press: 2016; pp. 241-320.

2. Hertkorn, N.; Harir, M.; Koch, B.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **2013,** *10*, 1583-1624.

3. Zark, M.; Christoffers, J.; Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: evidence from tandem FT-ICR-MS. *Mar. Chem.* **2017,** *191*, 9-15.

4. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016**, 13, 2257–2277.

5. Dittmar, T.; Stubbins, A. 12.6—Dissolved organic matter in aquatic systems. *Treatise on Geochemistry, 2nd edn. Elsevier: Oxford* **2014**, 125-156.

6. Jaffé, R.; Yamashita, Y.; Maie, N.; Cooper, W.; Dittmar, T.; Dodds, W.; Jones, J.; Myoshi, T.; Ortiz-Zayas, J.; Podgorski, D. Dissolved organic matter in headwater streams: compositional variability across climatic regions of North America. *Geochim. Cosmochim. Acta.* **2012,** *94*, 95-108.

7. Zark, M.; Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **2018,** *9* (1), 3178.

8. Rossel, P. E.; Vähätalo, A. V.; Witt, M.; Dittmar, T. Molecular composition of dissolved organic matter from a wetland plant (Juncus effusus) after photochemical and microbial decomposition (1.25 yr): Common features with deep sea dissolved organic matter. *Org. Geochem.* **2013,** *60*, 62-71.

9. Tose, L. V.; Benigni, P.; Leyva, D.; Sundberg, A.; Ramírez, C. E.; Ridgeway, M. E.; Park, M. A.; Romão, W.; Jaffé, R.; Fernandez-Lima, F. Coupling trapped ion mobility spectrometry to mass spectrometry: trapped ion mobility spectrometry–time-of-flight mass spectrometry versus trapped ion mobility spectrometry–Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2018,** *32* (15), 1287-1295.

10. Robinson, E. W.; Williams, E. R. Multidimensional Separations of Ubiquitin Conformers in the Gas Phase: Relating Ion Cross Sections to H/D Exchange Measurements. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (9), 1427-1437.

11. Robinson, E. W.; Garcia, D. E.; Leib, R. D.; Williams, E. R. Enhanced Mixture Analysis of Poly(ethylene glycol) Using High-Field Asymmetric Waveform Ion Mobility Spectrometry Combined with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2006,** *78* (7), 2190-2198.

12. Robinson, E. W.; Leib, R. D.; Williams, E. R. The Role of Conformation on Electron Capture Dissociation of Ubiquitin. *J. Am. Soc. Mass Spectrom.* **2006,** *17* (10), 1470-1480.

13. Robinson, E. W.; Sellon, R. E.; Williams, E. R. Peak deconvolution in high-field asymmetric waveform ion mobility spectrometry (FAIMS) to characterize macromolecular conformations. *Int. J. Mass Spectrom.* **2007,** *259* (1–3), 87-95.

14. Saba, J.; Bonneil, E.; Pomiès, C.; Eng, K.; Thibault, P. Enhanced Sensitivity in Proteomics Experiments Using FAIMS Coupled with a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer†. *J. Prot. Res.* **2009,** *8* (7), 3355-3366.

15. Xuan, Y.; Creese, A. J.; Horner, J. A.; Cooper, H. J. High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Comm. Mass Spectrom.* **2009,** *23* (13), 1963-1969.

16. Bridon, G.; Bonneil, E.; Muratore-Schroeder, T.; Caron-Lizotte, O.; Thibault, P. Improvement of Phosphoproteome Analyses Using FAIMS and Decision Tree Fragmentation. Application to the Insulin Signaling Pathway in Drosophila melanogaster S2 Cells. *J. Prot. Res.* **2011,** *11* (2), 927-940.

17. Schrader, W.; Xuan, Y.; Gaspar, A. Studying ultra-complex crude oil mixtures by using High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) coupled to an ESI-LTQ-Orbitrap Mass Spectrometer. *Eur. J. Mass Spectrom.* **2014,** *20* (1), 43-49.

18. Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. Petroleum Crude Oil Characterization by IMS-MS and FTICR MS. *Anal. Chem.* **2009,** *81* (24), 9941-9947.

19. Fasciotti, M.; Lalli, P. M.; Klitzke, C. F.; Corilo, Y. E.; Pudenzi, M. A.; Pereira, R. C. L.; Bastos, W.; Daroda, R. J.; Eberlin, M. N. Petroleomics by Traveling Wave Ion Mobility–Mass Spectrometry Using CO2 as a Drift Gas. *Energy & Fuels* **2013,** *27* (12), 7277-7286.

20. Benigni, P.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Targeted high-resolution ion mobility separation coupled to ultrahigh-resolution mass spectrometry of endocrine disruptors in complex mixtures. *Anal. Chem.* **2015,** *87* (8), 4321-5.

21. Benigni, P.; Fernandez-Lima, F. Oversampling Selective Accumulation Trapped Ion Mobility Spectrometry Coupled to FT-ICR MS: Fundamentals and Applications. *Anal. Chem.* **2016,** *88* (14), 7404-12.

22. Pu, Y.; Ridgeway, M. E.; Glaskin, R. S.; Park, M. A.; Costello, C. E.; Lin, C. Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2016,** *88* (7), 3440-3443.

23. Ridgeway, M. E.; Wolff, J. J.; Silveira, J. A.; Lin, C.; Costello, C. E.; Park, M. A. Gated trapped ion mobility spectrometry coupled to fourier transform ion cyclotron resonance mass spectrometry. *Int J Ion Mobil Spectrom* **2016,** *19* (2), 77-85.

24. Benigni, P.; Sandoval, K.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Gardinali, P.; Fernandez-Lima, F. Analysis of Photoirradiated Water Accommodated Fractions of Crude Oils Using Tandem TIMS and FT-ICR MS. *Environ. Sci. Technol.* **2017,** *51* (11), 5978-5988.

25. Benigni, P.; Marin, R.; Sandoval, K.; Gardinali, P.; Fernandez-Lima, F. Chemical Analysis of Water-accommodated Fractions of Crude Oil Spills Using TIMS-FT-ICR MS. . *J. Vis. Exp.* **2017,** *121*, e55352.

26. Benigni, P.; Bravo, C.; Quirke, J. M. E.; DeBord, J. D.; Mebel, A. M.; Fernandez-Lima, F. Analysis of Geologically Relevant Metal Porphyrins Using Trapped Ion Mobility Spectrometry–Mass Spectrometry and Theoretical Calculations. *Energy & Fuels* **2016,** *30* (12), 10341-10347.

27. Benigni, P.; Porter, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Increasing Analytical Separation and Duty Cycle with Nonlinear Analytical Mobility Scan Functions in TIMS-FT-ICR MS. *Anal. Chem.* **2018,** *90* (4), 2446-2450.

28. Hernandez, D. R.; DeBord, J. D.; Ridgeway, M. E.; Kaplan, D. A.; Park, M. A.; Fernandez-Lima, F. Ion dynamics in a trapped ion mobility spectrometer. *Analyst* **2014,** *139* (8), 1913-1921.

29. Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Rev. Sci. Instr.* **2011,** *82* (12), 126106.

30. Fernandez-Lima, F.; Kaplan, D.; Suetering, J.; Park, M. Gas-phase separation using a trapped ion mobility spectrometer. *Int. J. Ion Mobility Spectrom.* **2011,** *14* (2), 93-98.

31. Schenk, E. R.; Ridgeway, M. E.; Park, M. A.; Leng, F.; Fernandez-Lima, F. Isomerization Kinetics of AT Hook Decapeptide Solution Structures. *Anal. Chem.* **2014,** *86* (2), 1210-1214.

32. Schenk, E. R.; Mendez, V.; Landrum, J. T.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Direct observation of differences of carotenoid polyene chain cis/trans isomers resulting from structural topology. *Anal. Chem.* **2014,** *86* (4), 2019-24.

33. McDaniel, E. W.; Mason, E. A. *Mobility and diffusion of ions in gases*; John Wiley and Sons, Inc., New York: New York, 1973; 381.

34. Benigni, P.; Sandoval, K.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Gardinali, P.; Fernandez-Lima, F. Analysis of Photoirradiated Water Accommodated Fractions of Crude Oils Using Tandem TIMS and FT-ICR MS. *Environ. Sci. Technol.* **2017,** *51* (11), 5978-5988.

35. Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminformatics* **2016,** *8* (1), 3.

36. Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol. Oceanogr.* **2010,** *55* (4), 1467-1477.

37. Witt, M.; Fuchser, J.; Koch, B. P. Fragmentation Studies of Fulvic Acids Using Collision Induced Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2009,** *81* (7), 2688-2694.

CHAPTER III

III. STRUCTURAL CHARACTERIZATION OF DISSOLVED ORGANIC MATTER AT THE CHEMICAL FORMULA LEVEL USING TIMS-FT-ICR MS/MS

3.1 Abstract

TIMS-FT-ICR MS is an important alternative to study the isomeric diversity and elemental composition of complex mixtures. While the chemical structure of many compounds in the Dissolved Organic Matter (DOM) remains largely unknown, the high structural diversity has been described at the molecular level using chemical formulas. In this study, we further push the boundaries of TIMS-FT-ICR MS by performing chemical formula-based ion mobility and tandem MS analysis for the structural characterization of DOM. The workflow described is capable to mobility select (R ~100) and isolate molecular ion signals ($\Delta m/z$ =36mDa) in the ICR cell, using single shot ejections after broadband ejections and MS/MS based on sustained off-resonance irradiation collision-induced dissociation (SORI-CID). The workflow results are compared to alternative TIMS-q-FT-ICR MS/MS experiments with quadrupole isolation at nominal mass (~1Da). The technology is demonstrated with isomeric and isobaric mixtures (e.g., 4-methoxy-1-naphthoic acid, 2-methoxy-1-naphthoic acid, decanedioic acid) and applied to the characterization of DOM. The application of this new methodology to the analysis of a DOM is illustrated by the isolation of the molecular ion $[C_{18}H_{18}O_{10}\text{-}H]^-$ in the presence of other isobars at nominal mass 393. Five IMS bands were assigned to the heterogenous ion mobility profile of $[C_{18}H_{18}O_{10}\text{-}H]^-$ and candidate structures from the PubChem database were screened based on their ion mobility and MS/MS matching score. This approach overcomes traditional challenges associated with the similarity of fragmentation patterns of DOM samples (e.g., common neutral losses of H_2O, CO_2, and $CH_2\text{-}H_2O$) by narrowing down the isomeric candidate structures using the mobility domain.

3.2 Introduction

Critical environmental and ecological processes are strongly influenced by Dissolved Organic Matter (DOM)[1,2], one of the most studied natural complex mixtures. Thus, a thorough knowledge of DOM chemical composition and structure at the molecular level is essential for the understanding of its role in the aquatic environments. Although the molecular features of DOM have been the focus of a multitude of studies over the last decades[2-4], the elucidation of its chemical structure and a clear view of DOM isomeric complexity, persist as one of the most challenging analytical problems.[5-8]

Analytical approaches integrating ultra-high resolution mass spectrometry, gas/liquid pre-separation techniques and tandem mass spectrometry strategies have provided much of the existing information on the chemical diversity of DOM.[4,9-16] An online HPLC-Orbitrap MS/MS method developed by Hawkes et al.[5], was used in an attempt to isolate single compounds in DOM samples from different ecosystems. Although the ubiquitous nature of DOM isomeric complexity was demonstrated, the proposed procedure was unable to separate individual compounds and differentiate fragmentation patterns from specific isomers of the same chemical formula. This limitation, commonly perceived in similar studies, was likely due to two main aspects: i) LC traditional approaches are not resolutive enough to separate closely structurally related isomers, regardless of the type of chromatographic column[12], and ii) typical MS/MS experiments do not separate precursor ions within nominal mass leading to ambiguous structural interpretation, even in cases where isobaric interferences are mass resolved and fragments can be assigned with high mass accuracy.[17,18]

Novel workflows that combine both LC and ion mobility spectrometry (IMS) have been explored to assess the DOM complexity at the level of single isomer.[19-21] Lu et.al[20] described the integration of LC-IMS-TOF MS for the analysis of riverine DOM; while the approach allowed molecular components to be separated in both LC and IMS domains, several isomeric species shared close values of retention time and CCS, thus underestimating the isomeric coverage. In addition, the multi-precursor isolation/fragmentation at nominal mass limited the veracity of the molecular structure assignment based on MS/MS data.

Witt et.al[18] reported a negative ion mode electrospray ionization Fourier transform ion cyclotron resonance tandem mass spectrometry (ESI-FT-ICR MS/MS) method for the analysis of DOM at the chemical formula level. Sustained off-resonance irradiation collision-induced dissociation (SORI-CID) of single molecular ions mass peaks, previously isolated by correlated harmonic excitation field (CHEF) and correlated shots of isobars in the ICR cell, permitted the identification of fragmentation pathways at the chemical formula level. Despite the several advantages of this method, the high structural diversity at the isomeric level of DOM limited the candidate structural assignment.

With the advent of high-resolution ion mobility analyzers (R>80), several groups have work on their integration to ultra-high resolution mass analyzers. [22-31] Our team has pioneered the integration of trapped IMS (TIMS) with FT-ICR MS since 2015 for the characterization of isomeric species in complex mixtures[32]. Since them, several reports have shown the unique advantages of TIMS-FT-ICR MS for the characterization of the isomeric content in complex mixtures (e.g., endocrine disruptors, glycans, water accommodated fractions of crude oils, DOM, etc.).[33-38] While TIMS coupling to TOF-MS

and FT-ICR MS showed similar performance and high reproducibility during the analysis of DOM (i.e., both MS platforms were able to capture the major trends and characteristics), as the chemical complexity at the level of nominal mass increases with *m/z* (*m/z* > 300-350), only the TIMS-FT-ICR MS was able to report the lower abundance compositional trends.[21] Recently, a workflow based on TIMS-q-FT-ICR MS/MS at the level of nominal mass (i.e., 1 Da isolation) allowed for further estimation of DOM isomeric content based on ion mobility-selected fragmentation patterns and core fragments.[7] Aside from a novel estimation of the number of isomers based on MS/MS and ion mobility data, the high similarity of neutral losses in DOM MS/MS at nominal mass suggested that better isolation strategies prior to MS/MS were necessary.

In this study, we further push the boundaries of TIMS-FT-ICR MS by performing chemical formula-based ion mobility and tandem MS analysis for the structural characterization of DOM. The workflow described is capable to mobility select (R ~100) and isolate molecular ion signals (<36mDa) in the ICR cell, using single shot ejections after broadband ejections and MS/MS based on sustained off-resonance irradiation collision-induced dissociation (SORI-CID). Taking advantage of the high ion mobility resolution (R~100) and mass resolution (R~400,000), chemical formula-based ion mobility and tandem mass spectrometry profiles were generated. The technology is shown for the case of an isomeric and isobaric standard mixture and a DOM standard.

3.3 Experimental section

3.3.1 Sample preparation

Two isomeric standards (4-methoxy-1-naphthoic acid and 2-methoxy-1-naphthoic acid) and an isobaric compound at nominal mass 201 Da (decanedioic acid) were purchased

from Sigma-Aldrich (St Louis, MO) and Thermo Fisher Scientific (Ward Hill, MA), respectively. Surface water was collected from Pantanal (PAN) National Park-SE Brazil, one of the most important and biodiverse freshwater wetlands around the world. Details on sampling and sample treatment procedure can be found in reference 2. The DOM extracted sample was dissolved in denatured ethanol to a final concentration of 1 ppm. Prior to ESI-TIMS-FT-ICR MS analysis, both the sample and standards were spiked with 5% (v/v) of low-concentration Tuning Mix (G1969-85000) from Agilent Technologies (Santa Clara, CA). All solvents used were of Optima LC-MS grade or better, obtained from Fisher Scientific (Pittsburgh, PA).

3.3.2 ESI source

An electrospray ionization source (Apollo II ESI design, Bruker Daltonics, Inc., MA) was utilized in negative ion mode. Sample solutions were introduced into the nebulizer at 150 μL/h using a syringe pump. Typical operating conditions were 3300-3600 V capillary voltage, 4 L/min dry gas flow rate, 1.0 bar nebulizer gas pressure, and a dry gas temperature 180 °C.

3.3.3 TIMS-FT-ICR MS/MS Experiments

A custom built TIMS-FT-ICR MS Solarix 7T spectrometer equipped with an infinity ICR cell (Bruker Daltonics Inc., MA) was used for all the experiments. The principle on TIMS separation when coupled to the FT-ICR MS can be found in several publications of our group.[21,38,39] Briefly TIMS relies on the utilization of an electric field to hold ions stationary against a moving gas, so that the drag force is balanced with the electric field and ions are spatially separated across the TIMS analyzer axis based on their ion mobility.[40-42] During ion mobility separation, a quadrupolar field confines the ions in the

radial direction to increase trapping efficiency. The ion mobility, K, of an ion in a TIMS cell is described by the equation

$$K_0 = \frac{v_g}{E} = \frac{A}{V_{elution} - V_{out}} \qquad (1)$$

where v_g, E, A, $V_{elution}$, and V_{out} are the velocity of the gas, applied electric field, a calibration constant, elution voltage, and tunnel out voltage, respectively.

Values of K_0 can be correlated with the ion–neutral Collision Cross Section (CCS, Ω, Å^2) using the Mason–Schamp equation

$$\Omega = \frac{(18\pi)^{1/2}\, z}{16(k_B T)^{1/2}} \left(\frac{1}{m_i} + \frac{1}{m_b}\right)^{1/2} \frac{1760\, T}{Ko\ P\ 273.15\ N^*} \qquad (2)$$

where z is the charge of the ion, k_B is the Boltzmann constant, N^* is the number density, and m_i and m_b refer to the masses of the ion and bath gas, respectively.[43]

The TIMS analyzer was controlled using an in-house software, written in National Instruments LabVIEW, and synchronized with the FTMS control acquisition program. TIMS separation was performed in the Oversampling Mode[33] using nitrogen as a bath gas at ca 300 K, P1 = 2.6 and P2 = 0.8 mbar, and a constant RF (2200 kHz and 140–160 Vpp). The TIMS cell was operated using a fill/trap/elute/quench sequence of 9/3/9/3 ms, a maximum of 500 IMS scans per mass spectrum and a voltage difference across the ΔE gradient of 0.05 V (R~100). The ramp voltage gradient was stepped by 0.1 V/frame with a ΔV_{ramp} of 30 to 100 V. The deflector (V_{def}), funnel entrance (V_{fun}) and funnel out (V_{out}) voltages were V_{def} = −120/50 V, V_{fun} = −70 V and V_{out} = −6V/30 V. For the case of the DOM sample, the TIMS cell was operated in a multiple accumulation mode, with an average of 20 FT scans per IMS frame (ramp voltage gradient stepped by 0.0125 V/frame) resulting in a stepped E gradient of 0.25V/frame. A maximum of 400 IMS scans per mass spectrum

was collected and a 5V voltage difference across the ΔE gradient was established. The remaining voltage settings were kept constant.

3.3.4 CHEF-SORI-CID and Q-CID experiments

The precursor ions at nominal mass were pre-isolated in the quadrupole for increasing the ion transmission (10 Da width), then transferred to the ICR cell and isolated with a 1 Da notch applying Correlated Harmonic Excitation Field (CHEF)[18,44,45] with typical 8-10 % isolation power. Each single precursor ion was preserved for further SORI activation and the remaining isobaric peaks were ejected out of the ICR cell by single shots. Argon was pulsed into the ICR cell (10 mbar pressure) and a pump delay of 3s was used to reestablish high vacuum conditions before mass analysis. The ions were then sweep excited and finally detected with 2MW transient. A SORI power ranged from 0.5-1% with a pulse length of 0.1-0.2 s and a frequency offset of −500 Hz were set up for SORI-CID experiments. For comparison purposes, nominal mass q-CID isolation tandem MS experiments were performed using 15-20 eV CID energies prior to injection to the FT-ICR MS.

3.3.5 Data processing

The TIMS-FT-ICR MS spectra were externally calibrated for ion mobility using the Agilent Tuning Mix calibration standard and the reported nitrogen mobility (K_0) values by Stow et al.[46,47] The MS/MS spectra were internally calibrated using the exact masses of known neutral losses in DOM.[13] The assignment of chemical formula was conducted using Data Analysis (Bruker Daltonics v 5.2) based on formula constrains of $C_xH_yO_{0–19}$, and odd and even electron configurations were allowed. TIMS spectra for each molecular formula was processed using the custom-built Software Assisted Molecular Elucidation (SAME)

package – a specifically designed TIMS-MS data processing script written in Python v3.7.3. SAME package utilizes noise removal, mean gap filling, "asymmetric least squares smoothing" base line correction, peak detection by continuous wavelet transform (CWT)-based peak detection algorithm (SciPy package), and Gaussian fitting with non-linear least squares functions (Levenberg–Marquardt algorithm).[39] Data was processed using Data Analysis (v. 5.2, Bruker Daltonics, CA) and all other plots were created using OriginPro 2016 (Originlab Co., MA). Candidate structures were obtained by *in-silico* fragmentation using the MetFrag CL tool and the PubChem database[48,49] and theoretical CCS using the trajectory method (TM) in IMoS software version 1.09c for nitrogen as a bath gas at ca. 300 K.[50,51]

3.4 Results and discussion

The experimental sequence for TIMS in tandem with CHEF-shots isolation and SORI-CID is shown in Figure 3.1. Briefly, ions within a mobility range are scanned in the TIMS cell using a nonlinear scan function, then prefiltered in the quadrupole with 10 Da window and stored in the collision cell prior to the injection in the ICR cell for CHEFT+SHOTS and SORI-CID.

The target precursor is isolated in the ICR cell with a 0.036 *m/z* window by first applying a characteristic CHEF wave function to broadband eject ions out with a 1 *m/z* notch and by further ejecting isobars using correlated shots to 0.036 *m/z* window. Ions are fragmented using SORI-CID, and a final mass spectrum is generated at ultrahigh resolution (R ~ 150 000−400 000). The coupling of the TIMS cell to the FTICR in this setting is particularly advantageous since ions can be identified by their mobility and chemical formula. The mobility-selected high-resolution isolation and fragmentation of precursor

ions can provide unambiguous structural information at the level of the chemical formula. This capability is very challenging, if not impossible, to achieve by traditional q-based MS/MS workflows on complex mixtures.

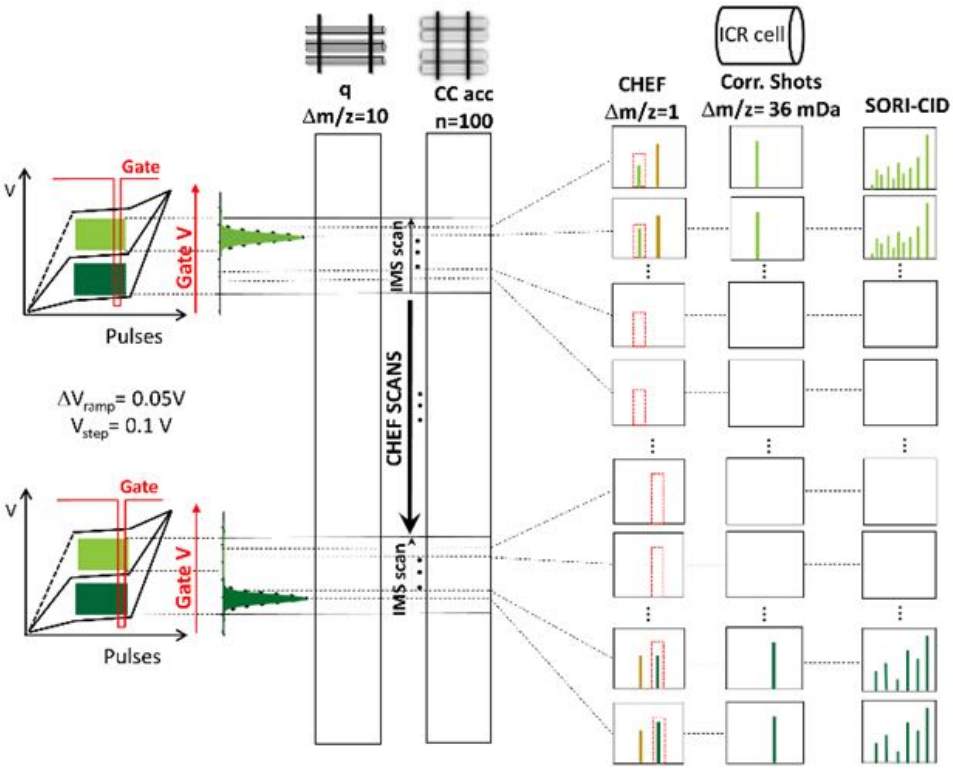


Figure 3. 1. TIMS-q-CHEF-SORI-CID MS/MS schematics. Ion mobility ranges are isolated with a 10 *m/z* window in the quadrupole and accumulated in the collision cell. Ions within the same mobility range are injected in the ICR cell, followed by CHEF-correlated shots isolation at 0.036 *m/z* and SORI-CID fragmentation. After a full ion mobility scan at a single target precursor (0.036 *m/z* window), the next full ion mobility scan at the next precursor is generated. The example is shown for isobars with different ion mobility values (light and dark green signals).

The capabilities of the proposed TIMS-q-CHEF-SORI CID MS/MS method is illustrated for the case of a mixture containing two isomeric standards (4-methoxy-1-naphthoic acid and 2-methoxy-1-naphthoic acid) and one isobaric standard (decanedioic acid). Typical TIMS-q-CID MS/MS and TIMS-q-CHEF-SORI CID MS/MS spectra of the

individual isomeric and isobaric standards are shown in Figure 3.2. Notice that the isobaric standard (C) cannot be separated by ion mobility from one of the isomeric standards (B and C).

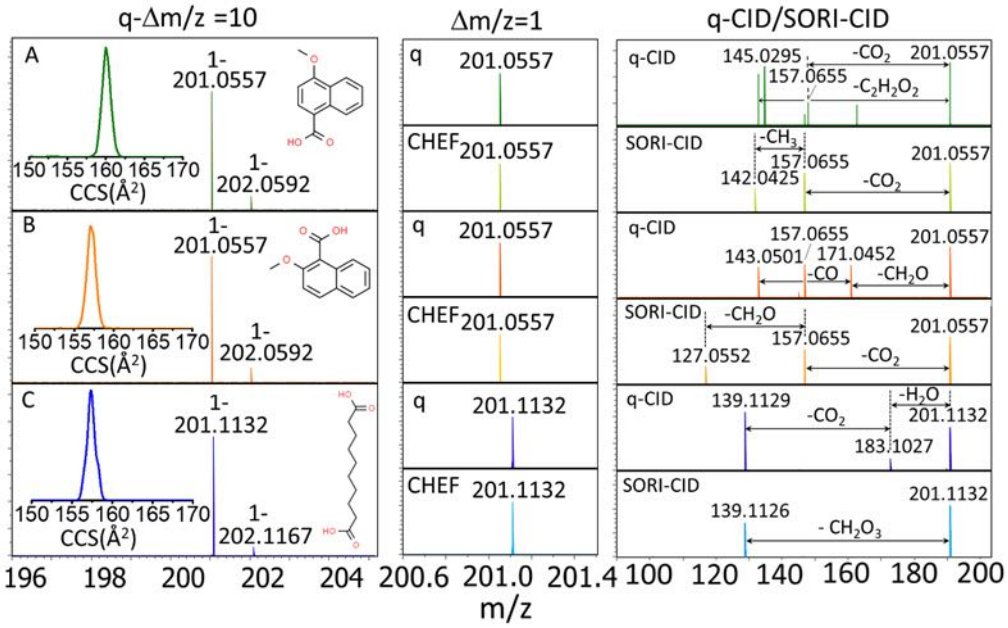


Figure 3. 2. Comparison of ESI (-)-TIMS-q-CID MS/MS and ESI (-)-TIMS-q-CHEF-SORI CID MS/MS of individual isomeric and isobaric standards: 4-methoxy-1-naphthoic acid (A), 2-methoxy-1-naphthoic acid (B) and decanedioic acid(C).

The 10 *m/z* isolation window profiles show monoisotopic mass peaks at nominal mass 201 for $[M-H]^-$ species (A/B: $[C_{12}H_{10}O_3-H]^-$ and C: $[C_{10}H_{18}O_4-H]^-$). The ion mobility projections depict single IMS bands for each standard with closely related CCS values in the range of 155-160 $Å^2$. Although baseline separation of the isomeric species A and B can be achieved in the IMS domain, the isobaric compound C shares the same ion mobility as standard B. The MS/MS spectra of the standards obtained by q-CID or SORI-CID show very similar fragmentation patterns. The MS/MS are both characterized by typical neutral losses of H_2O, CO, CH_2O, and CO_2. A closer view of the MS/MS profiles indicates that the generation of some characteristic fragments is favored when SORI-CID is used. For

example, a CO_2 neutral loss generates a common fragment (*m/z* 157.0657) for the isomeric standards A and B either by CID or SORI-CID. However, unique fragments were found for standard A (*m/z* 142.0425) and B (*m/z* 127.0552) when in-cell fragmentation (SORI-CID) was utilized. That is, the CHEF-SORI-CID approach can be useful to identify structural isomers with distinctive fragmentation profiles. For the case of the isobar (C), apart from a water loss (*m/z* 183.1027) in q-CID, no clear difference in the MS/MS spectra from both fragmentation approaches was observed.

The effectiveness of the TIMS-q-CHEF-SORI CID MS/MS method to unambiguously identify each compound was assessed through the study of the isomeric and isobaric standard mixture (Figure 3.3). The 10 Da and nominal mass isolations of the mixture (A+B+C) resulted in two mass peaks in the MS domain corresponding to the precursor species $[A/B-H]^-$ and $[C-H]^-$ when both strategies (Q vs CHEF) were implemented (first column, Figure 3.3).

A further application of correlated shots permitted a single isolation ($\Delta m/z$ of 0.036) of both A/B and C in different runs by ejecting C and A/B out of the ICR cell respectively (Colum 2, Figure 3.3). From the MS/MS spectra (Top column 3, Figure 3.3) and the IMS projections the standard A (green) can be unambiguously identified based on both ion mobility and fragmentation fingerprint when the q-CID approach is used. That is, standards B and C signals overlap in both MS/MS and IMS domains, which could lead to the misassignment of the fragments, consequently biasing any further structural interpretation. Conversely, the isomeric standards A and B are identified by both IMS and MS/MS (see characteristic fragments *m/z* 142.042 and *m/z* 127.0554) after in-cell isolation/fragmentation of the ion *m/z* 201.0557 in the ICR. Similarly, the standard C is

singly isolated by CHEF and shots ejection of A+B and the correlation of the ion mobility projection (blue profile) with the fragmentation spectrum permitted the unambiguous identification of standard C (bottom column 3, Figure 3.3). The above results clearly represent a proof of concept for the structural analysis of an isobaric and isomeric complex mixture, at the level of chemical formula, combining high-resolution mobility separations with single precursor ion MS/MS in the ICR cell.

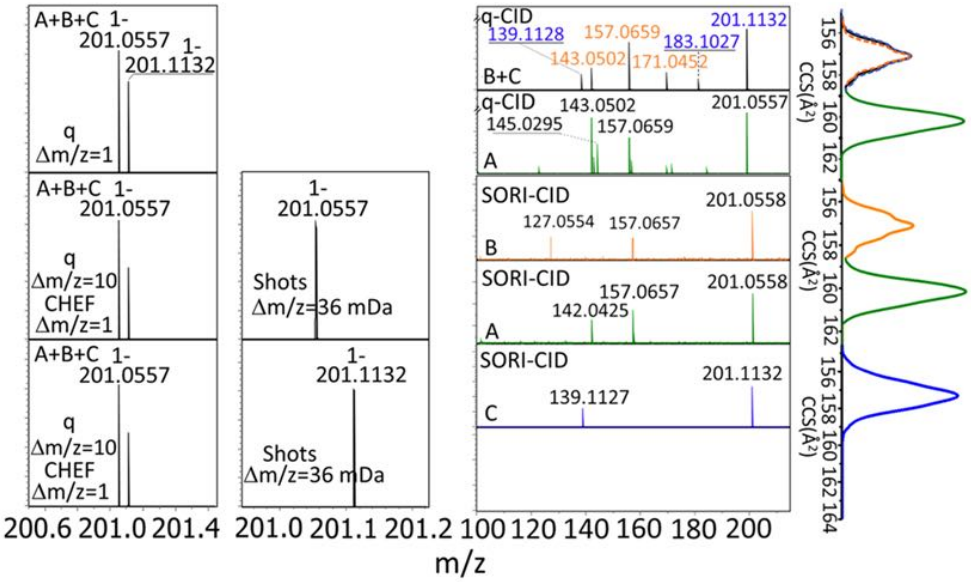


Figure 3. 3. Comparison of ESI (-) TIMS-q-CID MS/MS (Δ*m/z*=1) and TIMS-q-CHEF-SORI CID MS/MS (Δ*m/z*= 0.036) for an isobaric and isomeric standards mixture (4-methoxy-1-naphthoic acid A, 2-methoxy-1-naphthoic acid, B, and decanedioic acid, C). IMS projections shown in the right corner are correlated with each corresponding MS/MS profile. Notice the overlap of both ion mobility profiles and fragment spectra of B and C when ion mobility is combined with q-CID at nominal mass (orange and blue).

Previous analysis of DOM using non-targeted ESI-TIMS-FT-ICR MS illustrated the typical complexity observed in these samples with several thousands of *m/z* signals with a 2 Da space regularity, as the one depicted in the10 Da-isolation profile shown in the bottom of Figure 3.4, and around 3000 assigned chemical formulae based on accurate mass measurements[7,21]. The integration of the gas-phase separation by TIMS into the FT-ICR

MS workflow and the use of a methodology based on the computation of core fragments, neutral losses and fragmentation patterns from nominal mass CID experiments, also revealed a new complexity of DOM in the isomeric dimension. As previously discussed, the nominal mass isolation and CID fragmentation of a DOM sample could lead to an erroneous structural identification due to a false positive assignment of core and intermediate fragments arising from multiple precursors.

The TIMS-q-CHEF-SORI CID MS/MS method was applied to the structural analysis of the Pantanal DOM sample at the level of a chemical formula. The analysis of the Pantanal DOM sample combining TIMS with quadrupole isolation at nominal mass 393 Da and CID fragmentation resulted in five MS/MS profiles that share similar fragmentation patterns with typical neutral losses of H_2O, CO, CO_2, and CH_4 (Figure 3.4, top row). The analysis of the ion mobility domain revealed a heterogenous profile (CCS range 170-200 $Å^2$) that can be better described with five IMS bands (See the color profile in Figure 3.4). Notice that additional isomers/conformers could share the same IMS band; that is, the five assigned IMS bands may contain more than five isomeric species (more details below).

Inspection of the fragmentation data generated per each IMS band using TIMS-q-CID MS/MS (See Appendix S3.1) confirmed the co-occurrence of multiple intermediate fragments (e.g., isobaric fragments at *m/z* 349 resulting from CO_2 losses of four different precursor ions) regardless of the IMS band analyzed. This result, once again, evidences the necessity of a clean isolation of single precursor ions for a better structural analysis of DOM.

The CHEF broadband ejection applied to a 1 Da notch at nominal mass 393 in the Pantanal DOM sample showed several *m/z* signals corresponding to typical $[M-H]^-$ highly oxygenated species (Figure 3.4, bottom).

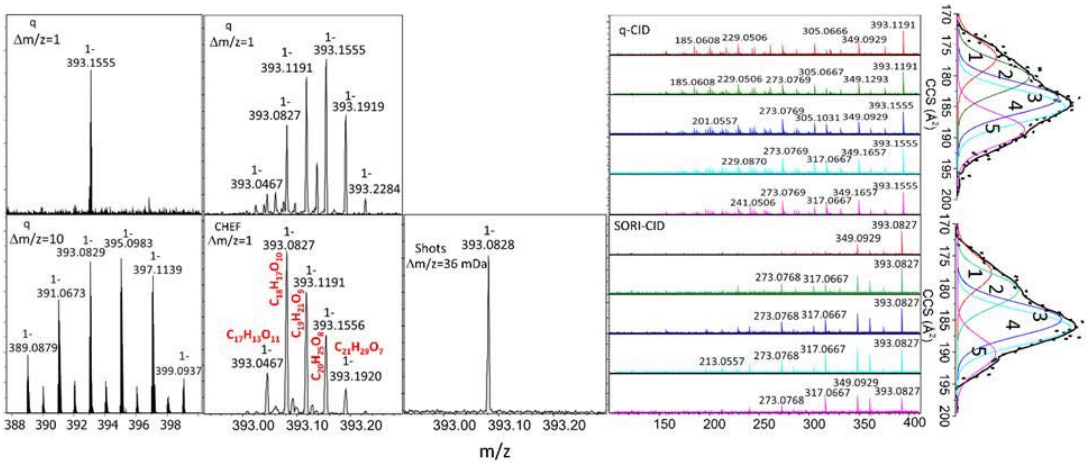


Figure 3. 4. Comparison of ESI (-) TIMS-q-CID MS/MS (Δ*m/z*=1) and TIMS-q-CHEF-SORI CID MS/MS (Δ*m/z* = 0.036) of the Pantanal DOM at nominal mass 393 (top row). The molecular ion $[C_{18}H_{18}O_{10}-H]^-$ isolation and SORI-CID is shown as a function of the ion mobility scans (bottom row). Ion mobility projections of the precursor ions with color annotated IMS bands are shown on the right (black dots represent the experimental data and black solid lines the best smooth fit from the SAME algorithm).

A further application of correlated shots ejections of the interferent isobars permitted a clean isolation of the $[C_{18}H_{18}O_{10}-H]^-$ precursor ion (Δ*m/z* = 0.036). Inspection of the ion mobility domain depicted an heterogenous profile with five bands annotated using the SAME code. The collected FT-ICR MS/MS scans associated with each IMS band of the isolated ion were averaged out resulting in five fragment spectra. Dissection of the MS/MS spectra showed common fragments for all IMS bands associated with neutral losses of H_2O and CH_3OH, multiple decarboxylations (nCO_2 losses), and combinations among them (Appendix 3.2), in good agreement with previous observations from Witt et.al.[18] Moreover,

characteristic fragments correlated with the IMS bands; this suggest that structural isomeric species could be identified by their ion mobility and fragmentation patterns.

Common fragments of the $[C_{18}H_{18}O_{10}\text{-}H]^-$ precursor ion found across the IMS bands were filtered out from the MS/MS data and search over PubChem database and in-silico fragmentation using Metfrag CL. Fifty candidate structures were retrieved solely based on the accurate mass of the precursor ion. Further analysis of each fragmentation data associated to the five experimental IMS bands, generated a list of 34 possible structures sorted by a score that reflects the best MS/MS match.

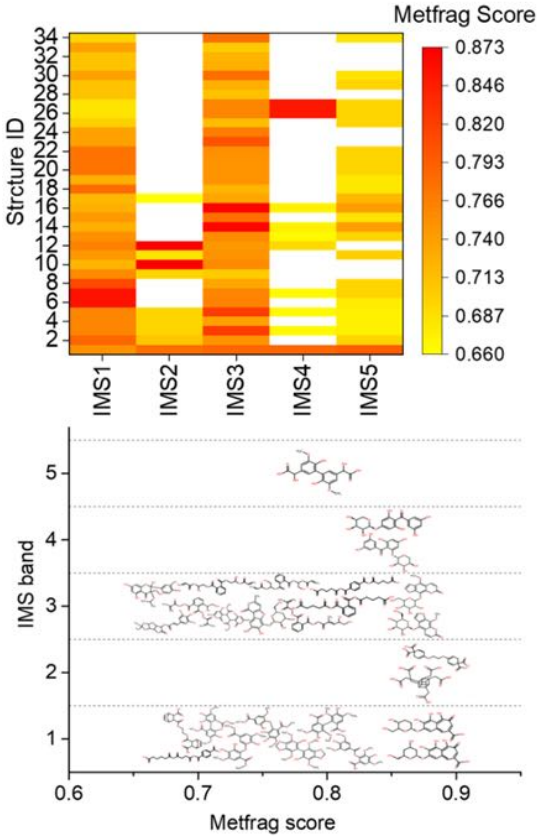


Figure 3. 5. Candidate assignment based on ion mobility and MS/MS matching score (Metfrag CL) from PubChem database for $[C_{18}H_{18}O_{10}\text{-}H]^-$. Notice that only structures with scores higher than 0.7 were plotted. More candidate information is contained in the supporting information.

The refined and sorted list of candidate structures per ion mobility and MS/MS scoring permitted the experimental assignments. Inspection of the heatmap in Figure 3.5 (top) revealed the highest Metfrag scores for the structures associated to the IMS bands 1,3 and 5. However, the more confident assignments, based on the difference between the highest and the second-to-highest scores among IMS bands, were found for the structures with ID 14 and 16 (IMS3), 10 and 12 (IMS2), 6 and 7 (IMS1), and 26 and 27 (IMS4). A distribution of the potential structures per IMS band constrained by the Metfrag score is shown in Figure 3.5 (bottom). In general, as mentioned above, the IMS1 and IMS3 bands group the majority of the assigned structures. While this number of potential isomeric species associated with a single chemical formula may appear high, it provides a short list considering the DOM complexity for secondary confirmation using individual standards

3.5 Conclusions

A novel TIMS-FT-ICR MS/MS workflow for the structural analysis of complex DOM samples at the level of chemical formula was developed. The integration of high-resolution ion mobility with single precursor ion isolation in the ICR with a *m/z* 0.036 isolation window and fragment assignment with both high resolution and mass accuracy resulted in a versatile approach that permits the identification and structural assignment of single species from a complex mixture (i.e., mixtures containing isomeric interferences). The analysis of a model mixture containing 4-methoxy-1-naphthoic acid, 2-methoxy-1-naphthoic acid, and decanedioic acid (isomers and isobars, respectively) demonstrated that every single compound can be unambiguously identified, either by ion mobility or their characteristic fragmentation spectra, thus representing a powerful approach compared to traditional nominal mass MS/MS schemes (co-isolation and fragmentation of several

precursors). The characterization of a chemical formula from a DOM sample at the nominal mass *m/z* 393 using the developed procedure evidenced that a single mass peak (*m/z* 393.0828) can be effectively isolated from several isobaric species in the ICR cell. In the ion mobility domain, five IMS bands were assigned to the heterogenous profile of the precursor ion $[C_{18}H_{18}O_{10}\text{-}H]^-$ and correlated to the fragmentation data obtained by CHEF+SHOTS-SORI-CID MS/MS. Candidate structures from the PubChem database were screened based on their ion mobility and MS/MS matching score. This study provides a proof of concept for the structural analysis of complex mixtures, at the level of chemical formula, combining high-resolution ion mobility separations with single precursor ion MS/MS in the ICR cell, without prior chromatographic separations.

3.6 References

1. Minor, E. C.; Swenson, M. M.; Mattson, B. M.; Oyler, A. R. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci.: Processes Impacts* **2014,** *16* (9), 2064-2079.

2. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016,** *13* (8), 2257-2277.

3. Dittmar, T. Chapter 7 - Reasons Behind the Long-Term Stability of Dissolved Organic Matter. In *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)*, Hansell, D. A.; Carlson, C. A., Eds.; Academic Press: Boston, 2015; pp 369-388.

4. Lu, K.; Liu, Z. Molecular Level Analysis Reveals Changes in Chemical Composition of Dissolved Organic Matter From South Texas Rivers After High Flow Events. *Front. Mar. Sci.* **2019,** *6* (673).

5. Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J. R.; Tranvik, L. J.; Bergquist, J. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnol. Oceanogr. Lett.* **2018,** *3* (2), 21-30.

6. Zark, M.; Christoffers, J.; Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar. Chem.* **2017,** *191*, 9-15.

7. Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the structural complexity of dissolved organic matter: isomeric diversity. *Faraday Discuss.* **2019,** *218* (0), 431-440.

8. Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Front. Mar. Sci.* **2017,** *4* (405).

9. Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the analytical limits: new insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. *Chem. Sci.* **2019,** *10* (29), 6966-6978.

10. Smith, D. F.; Podgorski, D. C.; Rodgers, R. P.; Blakney, G. T.; Hendrickson, C. L. 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures. *Anal. Chem.* **2018,** *90* (3), 2041-2047.

11. Pan, Q.; Zhuo, X.; He, C.; Zhang, Y.; Shi, Q. Validation and Evaluation of High-Resolution Orbitrap Mass Spectrometry on Molecular Characterization of Dissolved Organic Matter. *ACS Omega* **2020,** *5* (10), 5372-5379.

12. Patriarca, C.; Bergquist, J.; Sjöberg, P. J. R.; Tranvik, L.; Hawkes, J. A. Online HPLC-ESI-HRMS Method for the Analysis and Comparison of Different Dissolved Organic Matter Samples. *Environ. Sci. Technol.* **2018,** *52* (4), 2091-2099.

13. Zark, M.; Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **2018,** *9* (1), 3178.

14. Nebbioso, A.; Piccolo, A. Molecular characterization of dissolved organic matter (DOM): a critical review. *Anal. Bioanal. Chem.* **2013,** *405* (1), 109-124.

15. Smith, D. F.; Blakney, G. T.; Beu, S. C.; Anderson, L. C.; Weisbrod, C. R.; Hendrickson, C. L. Ultrahigh Resolution Ion Isolation by Stored Waveform Inverse Fourier Transform 21 T Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2020,** *92* (4), 3213-3219.

16. Le Maître, J.; Hubert-Roux, M.; Paupy, B.; Marceau, S.; Rüger, C. P.; Afonso, C.; Giusti, P. Structural analysis of heavy oil fractions after hydrodenitrogenation by high-resolution tandem mass spectrometry and ion mobility spectrometry. *Faraday Discuss.* **2019,** *218* (0), 417-430.

17. Cortés-Francisco, N.; Caixach, J. Fragmentation studies for the structural characterization of marine dissolved organic matter. *Anal. Bioanal. Chem.* **2015,** *407* (9), 2455-2462.

18. Witt, M.; Fuchser, J.; Koch, B. P. Fragmentation Studies of Fulvic Acids Using Collision Induced Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2009,** *81* (7), 2688-2694.

19. Gao, Y.; Wang, W.; He, C.; Fang, Z.; Zhang, Y.; Shi, Q. Fractionation and molecular characterization of natural organic matter (NOM) by solid-phase extraction followed by FT-ICR MS and ion mobility MS. *Anal. Bioanal. Chem.* **2019**.

20. Lu, K.; Gardner, W. S.; Liu, Z. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). *Environ. Sci. Technol.* **2018,** *52* (13), 7182-7191.

21. Tose, L. V.; Benigni, P.; Leyva, D.; Sundberg, A.; Ramírez, C. E.; Ridgeway, M. E.; Park, M. A.; Romão, W.; Jaffé, R.; Fernandez-Lima, F. Coupling trapped ion mobility spectrometry to mass spectrometry: trapped ion mobility spectrometry–time-of-flight mass spectrometry versus trapped ion mobility spectrometry–Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2018,** *32* (15), 1287-1295.

22. Robinson, E. W.; Williams, E. R. Multidimensional Separations of Ubiquitin Conformers in the Gas Phase: Relating Ion Cross Sections to H/D Exchange Measurements. *J. Am. Soc. Mass Spectrom.* **2005,** *16* (9), 1427-1437.

23. Robinson, E. W.; Garcia, D. E.; Leib, R. D.; Williams, E. R. Enhanced Mixture Analysis of Poly(ethylene glycol) Using High-Field Asymmetric Waveform Ion Mobility Spectrometry Combined with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2006,** *78* (7), 2190-2198.

24. Robinson, E. W.; Leib, R. D.; Williams, E. R. The Role of Conformation on Electron Capture Dissociation of Ubiquitin. *J. Am. Soc. Mass Spectrom.* **2006,** *17* (10), 1470-1480.

25. Saba, J.; Bonneil, E.; Pomiès, C.; Eng, K.; Thibault, P. Enhanced Sensitivity in Proteomics Experiments Using FAIMS Coupled with a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer†. *J. Prot. Res.* **2009,** *8* (7), 3355-3366.

26. Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion Mobility–Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angew. Chem. Int. Ed.* **2007,** *46* (42), 8001-8004.

27. Xuan, Y.; Creese, A. J.; Horner, J. A.; Cooper, H. J. High-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled with high-resolution electron transfer dissociation mass spectrometry for the analysis of isobaric phosphopeptides. *Rapid Comm. Mass Spectrom.* **2009,** *23* (13), 1963-1969.

28. Bridon, G.; Bonneil, E.; Muratore-Schroeder, T.; Caron-Lizotte, O.; Thibault, P. Improvement of Phosphoproteome Analyses Using FAIMS and Decision Tree Fragmentation. Application to the Insulin Signaling Pathway in Drosophila melanogaster S2 Cells. *J. Prot. Res.* **2011,** *11* (2), 927-940.

29. Schrader, W.; Xuan, Y.; Gaspar, A. Studying ultra-complex crude oil mixtures by using High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) coupled to an ESI-LTQ-Orbitrap Mass Spectrometer. *Eur. J. Mass Spectrom.* **2014,** *20* (1), 43-49.

30. Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. Petroleum Crude Oil Characterization by IMS-MS and FTICR MS. *Anal. Chem.* **2009,** *81* (24), 9941-9947.

31. Fasciotti, M.; Lalli, P. M.; Klitzke, C. F.; Corilo, Y. E.; Pudenzi, M. A.; Pereira, R. C. L.; Bastos, W.; Daroda, R. J.; Eberlin, M. N. Petroleomics by Traveling Wave Ion Mobility–Mass Spectrometry Using CO2 as a Drift Gas. *Energy and Fuels* **2013,** *27* (12), 7277-7286.

32. Benigni, P.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Targeted High-Resolution Ion Mobility Separation Coupled to Ultrahigh-Resolution Mass Spectrometry of Endocrine Disruptors in Complex Mixtures. *Anal. Chem.* **2015,** *87* (8), 4321-4325.

33. Benigni, P.; Fernandez-Lima, F. Oversampling Selective Accumulation Trapped Ion Mobility Spectrometry Coupled to FT-ICR MS: Fundamentals and Applications. *Anal. Chem.* **2016,** *88* (14), 7404-7412.

34. Pu, Y.; Ridgeway, M. E.; Glaskin, R. S.; Park, M. A.; Costello, C. E.; Lin, C. Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2016,** *88* (7), 3440-3443.

35. Ridgeway, M. E.; Wolff, J. J.; Silveira, J. A.; Lin, C.; Costello, C. E.; Park, M. A. Gated trapped ion mobility spectrometry coupled to fourier transform ion cyclotron resonance mass spectrometry. *Int. J. Ion Mobility Spectrom.* **2016,** *19* (2), 77-85.

36. Benigni, P.; Marin, R.; Sandoval, K.; Gardinali, P.; Fernandez-Lima, F. Chemical Analysis of Water-accommodated Fractions of Crude Oil Spills Using TIMS-FT-ICR MS. *J. Vis. Exp.* **2017,** *121*, e55352.

37. Benigni, P.; Bravo, C.; Quirke, J. M. E.; DeBord, J. D.; Mebel, A. M.; Fernandez-Lima, F. Analysis of Geologically Relevant Metal Porphyrins Using Trapped Ion Mobility Spectrometry–Mass Spectrometry and Theoretical Calculations. *Energy & Fuels* **2016,** *30* (12), 10341-10347.

38. Benigni, P.; Porter, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Increasing Analytical Separation and Duty Cycle with Nonlinear Analytical Mobility Scan Functions in TIMS-FT-ICR MS. *Anal. Chem.* **2018,** *90* (4), 2446-2450.

39. Benigni, P.; Sandoval, K.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Gardinali, P.; Fernandez-Lima, F. Analysis of Photoirradiated Water Accommodated Fractions of Crude Oils Using Tandem TIMS and FT-ICR MS. *Environ. Sci. Technol.* **2017,** *51* (11), 5978-5988.

40. Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Rev. Sci. Instrum.* **2011,** *82* (12), 126106.

41. Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A. Gas-phase separation using a trapped ion mobility spectrometer. *Int. J. Ion Mobil. Spectrom.* **2011,** *14* (2), 93-98.

42. Hernandez, D. R.; DeBord, J. D.; Ridgeway, M. E.; Kaplan, D. A.; Park, M. A.; Fernandez-Lima, F. Ion dynamics in a trapped ion mobility spectrometer. *Analyst* **2014,** *139* (8), 1913-1921.

43. McDaniel, E. W.; Mason, E. A. *Mobility and diffusion of ions in gases*; John Wiley and Sons, Inc: New York, 1973.

44. Heck, A. J. R.; de Koning, L. J.; Pinkse, F. A.; Nibbering, N. M. M. Mass-specific selection of ions in Fourier-transform ion cyclotron resonance mass spectrometry. Unintentional off-resonance cyclotron excitation of selected ions. *Rapid Comm. Mass Spectrom.* **1991,** *5* (9), 406-414.

45. de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass selection of ions in a Fourier transform ion cyclotron resonance trap using correlated harmonic excitation fields (CHEF). *Int. J. Mass Spectrom. Ion Processes* **1997,** *165-166*, 209-219.

46. Schenk, E. R.; Ridgeway, M. E.; Park, M. A.; Leng, F.; Fernandez-Lima, F. Isomerization Kinetics of AT Hook Decapeptide Solution Structures. *Anal. Chem.* **2014,** *86* (2), 1210-1214.

47. Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Anal. Chem.* **2017,** *89* (17), 9048-9055.

48. Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminf.* **2016,** *8* (1), 3.

49. Ruttkies, C.; Neumann, S.; Posch, S. Improving MetFrag with statistical learning of fragment annotations. *BMC Bioinf.* **2019,** *20* (1), 376.

50. Larriba, C.; Hogan, C. J. Ion Mobilities in Diatomic Gases: Measurement versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A* **2013,** *117* (19), 3887-3901.

51. Larriba-Andaluz, C.; Fernández-García, J.; Ewing, M. A.; Hogan, C. J.; Clemmer, D. E. Gas molecule scattering & ion mobility measurements for organic macro-ions in He versus N2 environments. *Phys. Chem. Chem. Phys.* **2015,** *17* (22), 15019-15029.

CHAPTER IV

IV. UNSUPERVISED STRUCTURAL CLASSIFICATION OF DISSOLVED ORGANIC MATTER BASED ON FRAGMENTATION PATHWAYS

(Adapted with permission from Leyva *et al*., 2022, Environmental Science and Technology, copyright 2022 American Chemical Society)

4.1 Abstract

Dissolved organic matter (DOM) is considered an essential component of the Earth's ecological and biogeochemical processes. Structural information of DOM components at the molecular level remains one of the most extraordinary analytical challenges. Advances in chemical formulae determination from molecular studies of DOM have provided limited indications on the structural signatures and potential reaction pathways. In this work, we extend the structural characterization of a wetland DOM sample using precursor and fragment molecular ions obtained by a sequential electrospray ionization – Fourier transform – ion cyclotron resonance tandem mass spectrometry (ESI-FT-ICR CASI-CID MS/MS) approach. The DOM chemical complexity resulted in near 900 precursors (P) and 24,000 fragments (F) molecular ions over a small *m/z* 261-477 range. The DOM structural content was dissected into families of structurally connected precursors based on neutral mass loss patterns ($P_{n-1}+F_{1:n}+C$) across the 2D MS/MS space. This workflow identified over 1,000 structural families of DOM compounds based on precursor and neutral loss (H_2O, CH_4O and CO_2). Inspection of the structural families showed a high degree of isomeric content (numerous identical fragmentation pathways), not discriminable with sole precursor ion analysis. The connectivity map of the structural families allows for the visualization of potential biogeochemical processes that DOM undergoes throughout its lifetime. This study illustrates that integrating effective computational tools on a comprehensive high resolution mass fragmentation strategy further enables the DOM structural characterization.

4.2 Introduction

Decoding the chemical structure of dissolved organic matter remains not only as one of the most interesting but also challenging analytical tasks. Although the molecular features of DOM have been the focus of a multitude of studies over the last decades,[1-7] the elucidation of its compositional structures and a clear view of DOM isomeric complexity persist as one of the most demandingly difficult analytical problems.[8-12]

Nuclear magnetic resonance (NMR)[2,13-18] and hyphenated ultrahigh-resolution mass spectrometry (UHRMS)[9-11,19-22] have been the leading approaches in the structural characterization of DOM. Although most advanced NMR techniques have provided valuable multi-dimensional information on DOM structural characteristics, the extraordinary molecular complexity of this material is still overcoming the NMR capabilities to resolving discrete molecular structures[1,13,23]. On the other hand, analytical approaches integrating ultrahigh-resolution mass spectrometry, gas/liquid separation techniques, and tandem mass spectrometry strategies have provided much of the existing information on the chemical diversity of DOM.[3,9,12,24,25] With the progressive increase in computational power and the high demand in the analysis of complex data, the characterization of DOM at the molecular level has been addressed by molecular dynamics and machine learning approaches as complementary tools to experimental workflows.[26-29]

In general, the study of DOM structural complexity using UHRMS have commonly focused on strategies that analyze regular patterns solely based on molecular ions[6,7,15,30-35]. The van Krevelen-type diagram has been the preferred approach to map UHRMS data from complex samples[2,36,37]. By plotting O/C vs H/C ratios from the molecular composition, it is possible to visualize clusters of compounds that exhibit similar structural characteristics.

Despite that lipid, protein, carbohydrate, tannin, lignin, and carboxylic-rich alicyclic (CRAM) type compound classes have been routinely identified in DOM[38,39], a structural assignment cannot be accurately provided solely on the basis of chemical formulas.[1,40] Kim et.al[36] additionally explored the combination of van Krevelen plots with Kendrick mass defect[37,41,42] to provide structural information based on reaction pathways. For instance, the replacement of 2H by an oxygen atom found along a diagonal of two parallel CH_2 series, was suggested as an oxidation pathway of a primary alcohol to a carboxylic acid.

Several parameters derived from chemical formulas are also commonly utilized to predict structural signatures and compositional trends of DOM molecular species. For instance, double bond equivalents (DBE) and the aromaticity index are used to estimate the degree of structure unsaturation and identify aromatic/condensed species in DOM components respectively.[43,44] Furthermore, the occurrence of several regular patterns in DOM and their potential correlation with families of structurally related compounds has been also reported[1]. However, an explanation on the origin of these regularities and the structural correlation among the compounds belonging to the homologous families has not been yet provided. Reports based on tandem mass spectrometry of selected molecular ions have shown promise for the identification of DOM structural features[4,20,22,45-47].

A FT-ICR MS/MS study[47] of solid-phase-extracted (SPE)-atmospheric organic matter has suggested that structural analogies could exist among members of a CH_2 homologous series since they share identical neutral losses during collision induced dissociation (CID). Similarly, several regular patterns in DOM chemical formulas such as H_2 and CH_2 series and a replacement of a CH_4 by an oxygen atom have been found using molecular level analyses by FT-ICR MS[23,45]. Although the CH_4 vs O pattern has not been clearly explained

from a structural perspective[48], it has been attributed to potential interchanges of functionalities (e.g. C_2H_5 vs CHO). Moreover, in a different contribution, the same authors[49] utilized known degradation pathways observed for lignin (a possible component of DOM) to structurally explain newly found repeating patterns in DOM chemical formulas. Interestingly, the O_2 and CH_4 vs O_2 regularities found in DOM components were correlated with aromatic ring openings ($+O_2$) and a combination of aromatic ring openings after one demethylation ($-CH_2$) and one side-chain oxidation ($-H_2$) respectively.

New structural insights into the H_2 and CH_2 homologous series from low molecular weight compounds of Suwannee River fulvic acid standard using size exclusion chromatography-electrospray ionization-time-of-flight (TOF) tandem mass spectrometry, have been reported by These et.al[48]. The similarity found in the fragmentation patterns of homologous isolated precursors (fragments exhibiting the same H_2 or CH_2 difference as their corresponding precursors), suggested that structural dissimilarities among family members presumably lied on their corresponding core structures.

The structural complexity of marine DOM using ultrahigh-resolution tandem mass analysis based on an orbitrap MS/MS workflow was explored by Cortes-Francisco et.al[20]. Although this study was not oriented to the analysis of structural regularities found in DOM, the potential fragmentation pathways proposed for one of the precursor ions attributed to a lipid-like compound, showed the utility of integrating MS/MS data with van Krevelen information to provide new structural understandings of DOM components.

The advantages of DOM analysis using complementary trapped ion mobility spectrometry-FT-ICR MS/MS with correlated harmonic excitation field have been shown[12]; while this work demonstrated the isolation by mobility and tandem MS/MS at the

level of chemical formula, its routine application is unviable due to the large number of isomers and isobars present in DOM samples. There is a need for simplified strategies capable of establishing structural patterns based on MS1 and tandem MS/MS information using shorter experimental and processing time scales. In this report, we propose a systematic nominal mass UHR MS/MS follow by a computational model capable of correlating structural features (or families) based on the fragmentation pathways of precursor molecules.

Data independent acquisition (DIA) is an acquisition strategy in mass spectrometry based on parallel collection of MS/MS spectra and has recently been utilized to improve the signal-to-noise ratio, reproducibility, and ultimate analyte coverage.[50-52] Recent advances in computing power and electronics have enabled 2D FT-ICR MS as an emerging DIA tool to analyze complex mixtures.[53-55] The application of an RF pulse sequence to manipulate the ion's cyclotron radii in the ICR cell[56,57], along with no ion isolation and ion-neutral collisions (infrared multi-photon dissociation and electron capture dissociation are mostly used), led to the correlation of precursor and fragment ion signals with enhanced resolution and sensitivity. Nevertheless, the presence of abundant scintillation noise[58] and difficulties associated with data processing, are still important limitations that need to be addressed to obtain comprehensive MS/MS data.

The introduction of continuous accumulation of selected ions (CASI) in FT-ICR MS instrument by Senko et.al[59] provided a way to increase the sensitivity and dynamic range, while reducing space charge effects by sequentially transmitting smaller m/z segments. More recently, this strategy has also been implemented in top-down mass spectrometry[60] and protein imaging[61], respectively. In the case of DOM analysis, CASI has allowed for

the detection of a larger number of chemical formulas when compared to traditional broadband acquisitions[23,62]. Despite of the increase on the number of chemical formulas, there is a need for further CASI implementations combined with sequential fragmentation (CASI MS/MS). In the case of complex mixtures analysis, CASI MS/MS workflows can greatly benefit from new computational algorithms for MS/MS data processing and structural correlations.

In this work, we extend the structural characterization of a wetland DOM sample from Pantanal, Brazil, using precursor and fragment molecular ions obtained with electrospray ionization-Fourier transform – ion cyclotron resonance tandem mass spectrometry (ESI-FT-ICR CASI CID MS/MS). Families of structurally related DOM compounds are identified based on characteristic mass loss patterns across heteroatom classes. We propose a novel graphical analysis of interconnected structural families as a potential tool that helps to understand DOM biogeochemical processes.

4.3 Experimental section

4.3.1 Sample preparation

The DOM sample was obtained by SPE of surface water collected from wetlands located at Pantanal National Park, Brazil. Details on sampling, sample treatment and the SPE procedure are described by Hertkorn et.al[2] and Dittmar et. al[63]. Briefly, 2L of surface water were collected using HCl pre-cleaned brown plastic bottles. Samples were kept refrigerated on ice and filtered using GFF pre-combusted glass fiber filters (0.7 μm nominal pore size) within 6 h after collection. Filtered samples acidified to a pH 2 with concentrated HCl, were loaded by gravity onto a 1g-Varian Bond Elut PPL cartridge using Teflon tubing. The PPL cartridge was preconditioned with methanol followed by pH 2 Milli-Q

water. The loaded cartridge was then rinse with pH 2 Milli-Q water and dried in a N2 gas flow for five minutes prior to the elution of DOM molecules with 20 mL of methanol. SPE-DOM extracts were stored in pre-combusted glass vials at -20 °C until further analysis. The choice of the sample comes from its recent UHRMS and IMS-UHRMS characterization (recent papers[11,12] and a 2016 report by Hertkorn et.al[2]). The SPE-DOM sample was diluted ten times by dissolving it in 1 mL of denatured ethanol. All solvents used were of Optima LC-MS grade or better, obtained from Fisher Scientific (Pittsburgh, PA).

4.3.2 ESI-FT-ICR-MS

A SolariX 9T ESI-FT-ICR MS spectrometer (Bruker Daltonics, MA) equipped with an infinity ICR cell was optimized for high transmission in the 100-1200 m/z range. Samples were ionized using an electrospray ionization source (Apollo II ESI design, Bruker Daltonics, Inc., MA) in negative ion mode at 200 μL/h injection. Typical operating conditions were 3700 - 4200 V capillary voltage, 2 L/min dry gas flow rate, 2.0 bar nebulizer gas pressure, and a dry gas temperature 200 °C. Operational parameters were as follows: funnel rf amplitude 160 peak-to-peak voltage (Vpp), capillary exit -150 V, deflector plate -140 V, skimmer1 -20 V, transfer line RF 350 Vpp, octupole RF amplitude 350 Vpp and collision cell RF 1100 Vpp. An Arginine cluster ion series (173-1740 Da) was used during the instrument tuning and control optimization. The broadband MS_1 spectrum (first MS dimension) of 115 co-added scans was collected at 4 MW data acquisition size (mass resolution of 4M at 400 *m/z*).

4.3.3 ESI-FT-ICR CASI CID MS/MS

For the CASI-CID experiments, ions at odd nominal masses were sequentially isolated (1 Da window) in the quadrupole (*m/z* range 261-477), accumulated for 5-7 s in the collision cell, and subject to CID prior to the analysis in the ICR cell. Multiple CID collision voltages (15 V – 27 V) tailored to the precursor nominal *m/z* were utilized for a better coverage across low and high *m/z* fragments. The same ion optics parameters used in broadband analysis were utilized during the MS/MS experiments. Up to 100 scans were co-added for each tandem mass spectrum (MS_2) in the segmental acquisition mode. Eight predefined segments were acquired and stitched for each experiment using the serial run mode.

4.3.4 ESI-FT-ICR CHEF SORI MS/MS

Differences between nominal mass and chemical formula-based MS/MS were evaluated for the case of the 267.087412 *m/z* ion ($C_{13}H_{15}O_6$) using correlated harmonic excitation field (CHEF)[12,46,64,65], shots ejection of isobaric ions (~0.002% power and 0.04 pulse length) and sustained off resonance irradiation (SORI)-CID (1.4% SORI power, 0.1 s pulse length of and -500 Hz frequency offset). A sweep excitation was applied, and six hundred MS/MS scans were collected at 2 MW data size.

4.3.5 Data Processing

Data was processed using Data Analysis (v. 5.2, Bruker Daltonics, CA), and all other plots were created using OriginPro 2016 (Originlab Co., MA). Chemical formulas assignment was conducted using Composer software (version 1.0.6, Sierra Analytics, CA, USA) and confirmed with Data Analysis (version 5.2, Bruker Daltonics). The formulas assignment was based on lowest formula errors, the presence of isotopologue signals and

the removal of isolated assignments (de-assignment of peaks belonging to classes with only a few sparsely scattered members). Theoretical formula constraints of $C_{4\text{-}50}H_{4\text{-}100}N_{0\text{–}3}O_{0\text{–}25}S_{0\text{–}2}$, S/N>3, m/z range 100-900, error <1 ppm and 0<O/C≤2, 0.3≤H/C≤2.5, and DBE-O ≤ 10^{66} were considered. The internal walking calibration performed in Composer using oxygen homologous series (O_4-O_{20}) resulted in an average error < 80 ppb for the mass range 229-890 Da. Both odd and even electron configurations were allowed in Data Analysis software. The MS/MS spectra were internally calibrated using a list of exact masses of fragment ions obtained from commonly occurring neutral losses in DOM and their combinations[8,22]. A four column excel file containing (1) the accurate mass of assigned peaks from both MS2 and MS1 (odd masses *m/z* 261-477), (2) the isolated nominal mass, (3) the intensity, and (4) the chemical formulas was created as input file for further data processing using Graph-DOM, an in-house code written in Python 3.7.3.

Ordered fragmentation pathways were computed based on the following equation:

$$P = [NL_1 + NL_2 + NL_3 + \cdots NL_n] + C \qquad (1),$$

where P corresponds to the chemical formula of the isolated precursor at nominal mass and NL is the neutral specie lost during the fragmentation of precursor and fragment ions. In this study, CH_4, O, H_2O, CO, CH_2O, CH_4O, and CO_2 were considered as potential neutral losses[20,46,67,68]. The sequence [NL_1-NL_2-NL_3-…NL_n] in equation (1) is an ordered array of neutral losses (NL) generated by an approach similar to the one recently described by Simon et.al[67]. Differently from Simon's approach, we sequentially match the exact mass of the theoretical NL with the mass difference of two consecutive assigned peaks with 1 mDa tolerance error. The core fragment (C) was defined as the lowest mass assigned fragment in a given pathway. Note that this approach also considers the search of multiple

NL if the mass difference between two peaks does not match the accurate mass of a single NL. Due to the large amount of fragment data collected in this study, we set the NL multiple at 2. Nevertheless, the Graph-DOM code allows the user to define both the type of NL and its multiples.

Families of structurally related compounds were identified using a conceptual model ($P_{n-1}+F_{1:n}+C$), defined in the Graph-DOM code, based on *de novo* matching of fragmentation pathways. Briefly, a precursor chemical formula along with the full fragmentation pathway is searched across all computed pathways in ascending order of mass. Note that the n-1 subscript in the model indicates the presence of P_{n-1} precursor as first fragment of P_n's fragmentation pathway (See Figure 4.3 panel B). The term $F_{1:n}$ defines the full match condition for all fragments in the pathway to consider a precursor in a family. Cytoscape v.3.8.2[69] was used to visualize the complexity of DOM in the form of structural networks formed by a neutral loss-based interconnection of family members. A list of the precursors found in the structural families was imported into Cytoscape and defined as nodes. Structural functionalities based on neutral loss differences among precursors in a family were imported as edges.

4.4 Results and discussion

The broadband ESI-FT-ICR MS spectrum of the SPE-DOM sample showed a typical distribution of $[M-H]^-$ ion signals with a maximum around 400 *m/z* (Figure 4.1A). A section of the spectrum (4Da mass range) depicts the characteristic DOM pattern of most abundant signals located at every other odd *m/z* and lower intensity peaks at even m/z (See inset, Figure 4.1A). The van Krevelen plot (Figure 4.1B) obtained after assigning near 4,000 molecular formulas, showed a dominance of CHO (green) and CHON (orange)

heteroatoms classes in the region 0.3<O/C<0.8 - 0.4<H/C<1.8 attributed to lignin and tannins type molecules followed by less abundant CHOS (blue) compound classes associated to sulfonated carboxylic-rich alicyclic components[2,36,39,70].

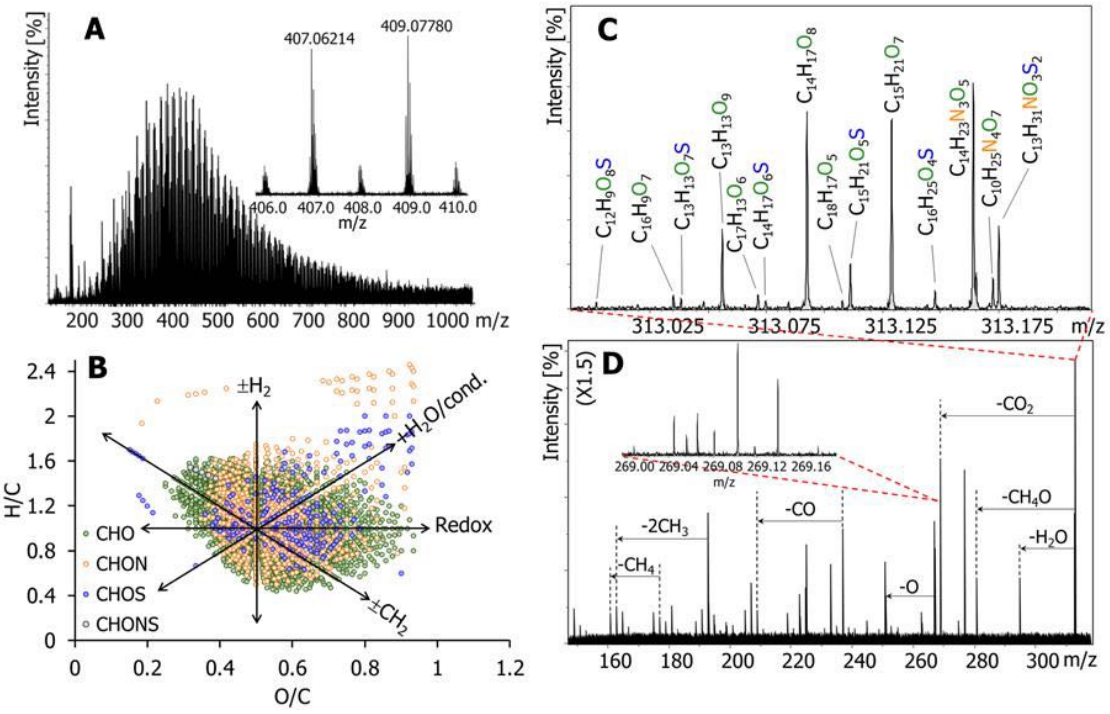


Figure 4. 1. ESI-FT-ICR MS broadband spectrum of the SPE-DOM sample and expanded view of the *m/z* range 406-410 shown in the inset (A). Van Krevelen plot obtained after chemical formula assignment of mass signals with black arrows describing DOM reaction pathways previously suggested by Kim et.al[36]. CHO, CHOS, CHON and CHONS compound classes are represented in green, orange, blue and grey colors respectively (B). Section of a MS/MS spectrum showing $[M-H]^-$ precursor ions isolated at nominal mass 313. Assigned molecular formulas are displayed with heteroatoms indicated with the color code (C). Typical MS/MS spectrum of the precursors isolated at nominal *m/z* 313 with annotated common neutral losses observed in DOM (D). Note that single peaks showed at nominal masses may comprise an envelope of multiple mass signals. For instance, nine peaks resulting from the CO_2 loss of precursors fragmented at m/z 313 are shown at *m/z* 269 (Panel D, inset).

A closer view of the nominal mass 313 (Figure 4.1C), shows the characteristic isobaric complexity of the sample, where up to 14 precursor ions of the CHO, CHON, CHOS and CHONS classes were co-isolated and fragmented. Similar patterns resulting in an average

of 10 precursor ions per MS_2 spectrum (total 110 fragment spectra collected) are found across the studied mass range (*m/z* 261-477). Typical fragmentation patterns showing common DOM neutral losses of CH_4, H_2O, CO, CH_4O, and CO_2 were observed across the fragmentation data set (See the MS_2 profile of precursors isolated at 313 *m/z* in Figure 4.1D). Few other less abundant neutral losses associated with sulfur (SO_3) and nitrogen (NH_2OH and HNO_3) species were also observed.

The analysis of the potential reaction pathways previously reported for DOM[36] and described by black arrows in Figure 4.1B, suggests that compounds found along a pathway (e.g., Redox) in the van Krevelen space are part of a structural family with a potential common backbone. Since structural questions are difficult, if not impossible, to answer solely based on chemical composition obtained from UHRMS, here we explored a fragmentation strategy that will provide new information about the structural complexity of DOM as a complementary tool to the traditional van Krevelen plot.

The application of the ESI-FT-ICR CASI CID MS/MS workflow resulted in more than 24,000 total assigned chemical formulas (~900 precursors). The CHO constituted the most abundant compound class (80% of all the precursors assigned), followed by the CHOS (~17%) and CHON (<3%) classes (Appendix 4.1). 2D MS/MS plots generated using all identified molecular formulas (A) and the filtered *m/z* signals assigned to the CHO, CHON and CHOS compound classes respectively (B-D) are shown in Figure 4.2. A closer view to the panels B-D in Figure 4.2 confirmed the clear dominance of the CHO compounds during fragmentation (>23, 000 chemical formulas) over the less abundant CHON and CHOS compound classes. Consequently, the O-heteroatom class will constitute the main focus of this study.

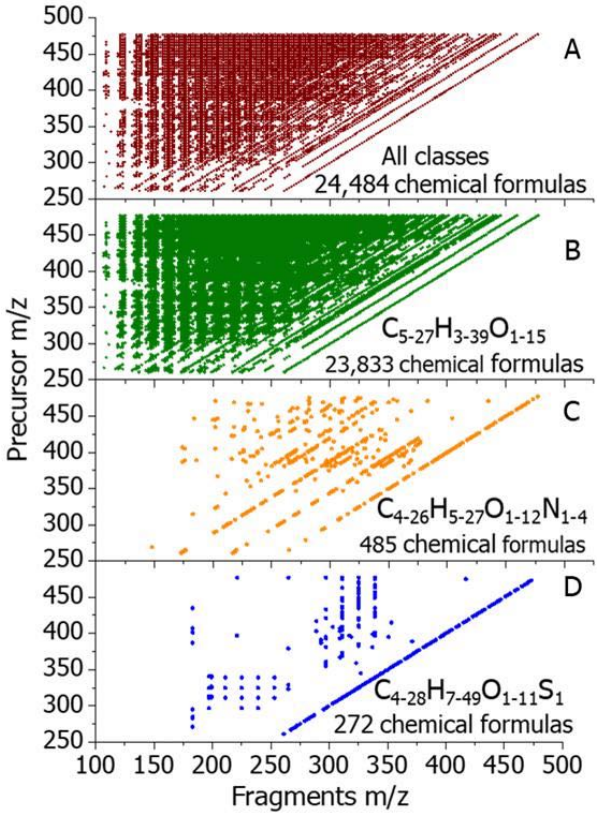


Figure 4. 2. 2D MSMS plots generated after chemical formula assignment of ion signals obtained from the FT-ICR CASI-CID MS/MS experiments.

Similar to the 2D mass spectrum described by van Agthoven[54], in our 2D MS/MS plot of Fragment *m/z* vs Precursor *m/z*, typical straight lines can be observed. Examination of the Figure 4.2A denotes that data points are aligned over diagonal lines described by the equation (2):

$$Precursor\ \left(\frac{m}{z}\right) = Fragment\left(\frac{m}{z}\right) + neutral\ loss \qquad (2)$$

The first diagonal line observed (right towards left) represents the precursor line and it contains the precursor ions. Since the chemical formula assignment was based on accurate mass (error < 1 ppm), precursors and fragments can be directly correlated (data points horizontally aligned in the 2D MS/MS domain)[8,20,22]. Neutral loss lines are parallel to the

precursor line and the NL mass (relative to precursor line) can be determined by the intercept of the equation (2). For instance, the characteristic line of one H_2O loss (first line from precursor line in Figure 4.2B) can be described using the equation $Precursor\left(\frac{m}{z}\right) = Fragment\left(\frac{m}{z}\right) + 18$. Other typical NLs observed (e.g., CO, CH_4O, CO_2, etc.) and their corresponding multiples can be visualized in the form of their characteristic lines.

The alignment of the MS/MS data along unique NL lines observed in Figure 4.2 A-C evidenced the similarity of DOM fragmentation pathways regardless of the precursor chemical composition. These structural patterns are in good agreement with previous findings obtained from fragmentation experiments of few selected nominal masses[8,46,47,71]. The systematic occurrence of NL-line patterns in the 2D MS/MS space, suggests that DOM molecules are clustered by families of compounds that could likely share common backbone structures.

The analysis of the complex fragmentation data generated from the FT-ICR CASI-CID MS/MS experiments was performed by designing an efficient data mining approach implemented in the Graph-DOM (Figure 4.3). The first step consisted of computing all possible ordered fragmentation pathways for the assigned precursors (Figure 4.3, panel A) using equation (1). For instance, a fragmentation pathway for the precursor $C_{16}H_{19}O_9$ is described as $C_{16}H_{19}O_9 = [H_2O+CO_2+CH_4O+CO_2] + C_{13}H_{13}O_3$. Since NLs are directly correlated with structural functionalities, the H_2O, CH_4O and CO_2 chemical units could be interpreted as CID fragments associated with hydroxyl, methoxy and carboxylic moieties, respectively. Since a precursor formula comprises a variety of isomeric species, multiple fragmentation pathways (with the same or different core fragments) can be associated with

the same precursor[11]. Note that the core fragment chemical formula could be interpreted as the backbone of the precursor structure and can also hold isomeric diversity. In the examples shown, the core fragment is limited by the lower m/z experimentally observed (in this instrument and settings, *m/z* below 100 are not detected).

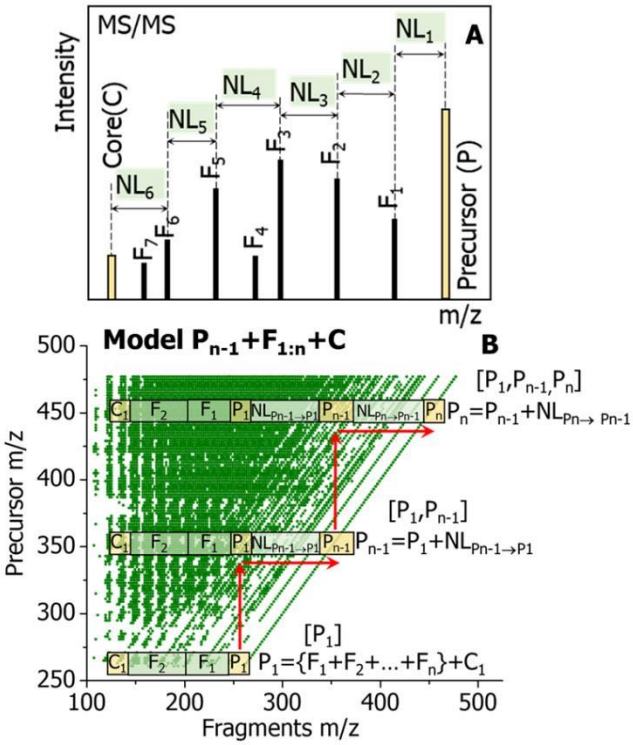


Figure 4. 3. Conceptual models designed to compute ordered fragmentation pathways (panel A) and find structural families in DOM based on sequential matching of fragmentation pathways (Panel B). Note that for the precursor P1 to be considered in a family, its ion mass should match (1 mDa tolerance) the mass of the first fragment in P_2's fragmentation pathway.

Over 10^7 ordered fragmentation pathways were computed for the CHO compound class following the workflow described in Figure 4.3A. Precursor compounds within the mass range 395-477 exhibited the highest number of fragmentation pathways (>100,000), in agreement with the extensive amount of product ions detected (See Appendix 4.1). An

average of 7 million pathways was found for precursor molecules containing 11-13 oxygens in their composition. On the other hand, less oxygenated DOM compounds (8-12 oxygens) generated a larger number of core fragments (Appendix 4.2). The relative high abundance of fragmentation pathways and core fragments for O-rich molecules suggests that the degree of oxygenation plays a key role on DOM structural diversity.

Assuming that the fragmentation pathway of a precursor $C_xH_yO_z$ is fully matched to the pathway of another precursor $C_{x+a}H_{y+b}O_{z+c}$, we could presume that they are structurally related. Consequently, the compositional difference between these two precursors will be the chemical unit $C_aH_bO_c$. Since many of the fragments assigned in the MS/MS spectra are also observed in the MS_1 domain, other precursors will likely show the same behavior as both $C_xH_yO_z$ and $C_{x+a}H_{y+b}O_{z+c}$. Therefore, they can be grouped into families characterized by a NL-based sequence resulting from the difference in chemical units among precursors.

The computation of structural families of DOM was conducted by implementing the conceptual model $P_{n-1}+F_{1:n}+C$ graphically described in Figure 4.3B. An overlapping strategy of the fragmentation pathways in the form of $P=[F_1+F_{2+}\ldots+Fn]+C$ was utilized. The overlap step consisted of matching both the initial lowest mass precursor P_1 and its fragmentation pathway in the database generated from the previous step (Figure 4.3, panel A) in ascending order of mass. The initial precursor P_1 is further grouped into the family $[P_1, P_{n-1}]$ with the newly matched precursor P_{n-1} and the chemical unit difference $NL_{Pn-1\rightarrow P1}$ is stored as the structural difference between P_1 and P_{n-1}. The resulting pathway $P_{n-1}= P_1+[F_1+F_{2+}\ldots+Fn] + C_1$ is searched again for a new match and the loop is repeated until no further match is found. Finally, the family ($[P_1, P_2,\ldots, P_{n-1}, P_n]$) is created as an array of the precursors sharing the same fragmentation pathways. The chemical unit difference

identified as a neutral loss among precursors within a family represents the functionality that is being added /subtracted to/from the family members. This array of neutral loss-based moieties illustrates the potential biogeochemical transformation processes experienced by DOM molecules. Once a family is retrieved, a new precursor higher in mass than P_1 is reset as initial lowest mass precursor and the pathway matching algorithm is repeated until all potential families are computed. Note that since various fragmentation pathways might be common to different precursors, multiple identical structural families will be expected. We define these sequences of analogous precursors as isomeric families, and they are an important indication of the confidence during the computation of the families.

The model performance to retrieve CHO structural families (coverage of precursors, intermediate fragments and core fragments) is described in Figure 4.4. Although the coverage of precursors in the families was 60%, over 2,000 DOM structural families were identified. A higher coverage was found for both the intermediate (~90%) and core fragments (> 80%). Note that since the same core and intermediate fragments might be found at different nominal masses, those fragments were counted in the families every time they were linked with a different precursor. These results suggests that there are potential structural families that remain undetected under the current conceptual model. While our workflow provides higher confidence, in grouping structurally related DOM compounds, than previous approaches, there are still limitations associated with the considerations of the proposed model.

Structural families containing two-to-four precursor members of the CHO class were the most abundant (261-477 *m/z* range). A decrease in the number of families was also

observed as the family size (number of precursors in a family) increased from four to six members (Figure 4.4B). Up to five precursors were found in over 300 structural families and the lowest abundant family (< 100) contained six DOM compounds. The relatively high number of 2-members families (>400) could be attributed to the limited mass range analyzed in the current study, preventing the match of fragmentation pathways from precursors with higher mass (>477).

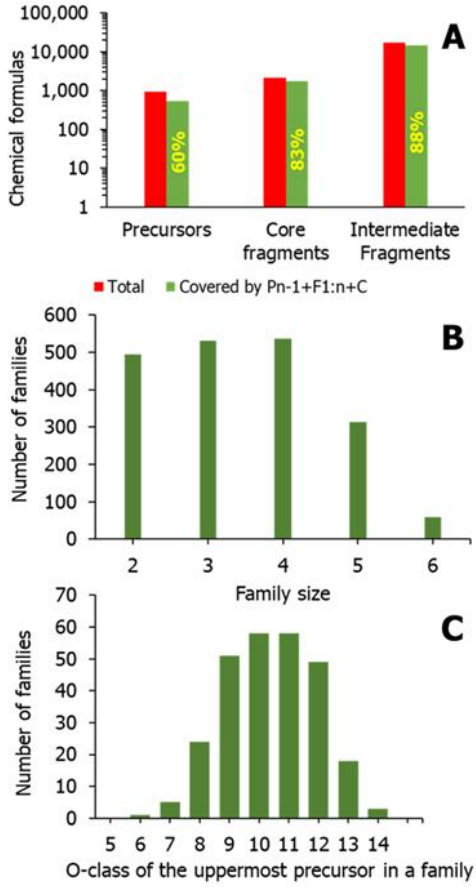


Figure 4. 4. Number of covered precursors, core and intermediate fragments by the model $P_{n-1}+F_{1:n}+C$ (A), distribution of the number of families per family size (B) and families per oxygen class of the uppermost precursor (compound with the highest oxygen content in the family) (C) respectively for the CHO class.

The number of families per oxygen class of the uppermost precursor within a family depicts a gaussian-type distribution centered in the O-class 10 (Figure 4.4C). This pattern is in good agreement with the distribution of pathways and core fragments per O-class found for the CHO compound class (Appendix 4.2). Nevertheless, a closer view of the Figure 4.4C evidenced a shift of the distribution towards less oxygenated family parents and an increase in the number of these uppermost precursors with 8-9 oxygens.

Overall, the families retrieved from the FT-ICR CASI-CID MS/MS data collected in the studied mass range showed that the structural transformation of CHO components in DOM depends on oxygenation/deoxygenation processes driven by both single and mixed addition or subtraction of H_2O, CH_4O and CO_2 chemical units (Appendix 4.3). This finding suggests that the structural alteration of DOM involves complex mechanisms compared to the uniform trends (e.g., hydration and carboxylation) previously observed from broadband FT-MS data[36,45]. Although our findings are constrained to O-compounds negatively ionized, the proposed approach allows the structural analysis of other molecular classes (e.g., CHOS and CHON) upon availability of substantial fragmentation data.

A closer view at the compositional relationship among members within a family revealed that oxygenation (increase in O/C ratio) through the addition of carboxylic (CO_2) moieties, increase de unsaturation degree of the resulting species (+1 DBE). Conversely, hydroxyl (H_2O) and methoxy (CH_4O) additions are accompanied by a decrease in one DBE unit of the subsequent molecule.

A 2D MS/MS representation of the molecular transformations exhibited by the structural family [$C_{14}H_{13}O_5$-$C_{15}H_{13}O_7$-$C_{16}H_{17}O_8$-$C_{17}H_{17}O_{10}$-$C_{17}H_{19}O_{11}$] identified in the SPE-DOM sample is shown in Figure 4.5A. The double arrows placed between family

members indicate that potential biogeochemical transformations of DOM can be viewed from a bidirectional perspective. For instance, a sequential addition (synthesis-like) of carboxylic, hydroxyl and methoxy moieties (-CO_2, -CH_4O, -CO_2, and -OH) starting from $C_{14}H_{13}O_5$ up to the family parent $C_{17}H_{19}O_{111}$ is described in Figure 4.5A.

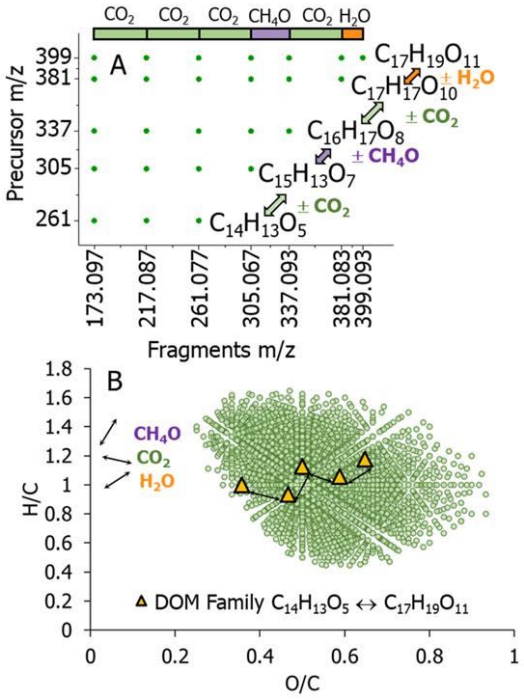


Figure 4. 5. 2D MS/MS visualization of a characteristic DOM family of 6 precursor (Panel A). Chemical unities (H_2O, CH_4O and CO_2) differences among precursors are shown using a color code. Fragmentation pathways described as neutral losses are also shown as colored bars. Van Krevelen plot (B) of the CHO class compounds obtained from the MS_1 experiment highlighting the compositional nature of the structural family.

This successive functionalization of O-depleted low molecular weight compounds resulting in high molecular weight O-rich molecules, could be explained as aging processes. Evidence of an increase in oxidized species observed in relatively old DOM from deep ocean water[19,23] compared to younger freshwater DOM has been previously reported. Similar findings of fresh (^{14}C dating) DOM exhibiting less oxygenated and lighter

molecules compared to older terrestrial DOM species have been also reported by Benk et.al.[72] However, the consistent decrease in unsaturation of oxygenated high molecular weight DOM components observed in the previous study, contrasts with our results of alternating unsaturation patterns along an ascending-order structural family (e.g., DBE change 8-7-8-7-8 from $C_{14}H_{13}O_5$ to $C_{17}H_{19}O_{11}$). A notable increase in O-rich molecules at the expense of the consumption of poor oxygenated species was also reported in biodegradation experiments conducted on DOM from landfill leachate[73] and from the surface of glaciers and ice sheets.[74] Although the impact of biodegradation on DOM structural transformation was not investigated in this study, the increasing oxygenation trend reported in both contributions, is in good agreement with the O-based functionalization found in our structural families. Other abiotic processes such as photo or chemical oxidation have been also indicated as responsible for the presence of highly oxygenated and CRAM species in DOM[19,75,76], yet supporting our findings observed along a structural family in ascending order.

The analysis of the structural families in the reverse direction (top to bottom) suggests that DOM molecules could also undergo mineralization-like transformations resulting in low molecular weight reduced species. For instance, the oxygen-rich family parent $C_{17}H_{19}O_{11}$ (Figure 4.5A) experiences de-functionalization processes characterized by consecutive eliminations of H2O, CO_2, CH_4O, and CO_2, resulting in the poorly oxygenated low molecular weight compound $C_{14}H_{13}O_5$. Interestingly, it has been suggested that highly-oxygenated compounds[77] and aromatic oxidized species[78] from terrestrial DOM, determined by broadband FT-ICR-MS, are preferentially removed by biodegradation, resulting in low molecular weight components. Similarly, a significant decrease in aromatic

content and oxygen functionalities was observed by Ward et. al[76] during photodegradation experiments of soil DOM compared to dark controls. Moreover, Hawkes et.al[79] have found that hydrothermal environments such as the ones observed in ocean deep hydrothermal vents, could induce potential de-functionalization processes (e.g., decarboxylation and dehydration) of O-rich high molecular weight DOM species, resulting in less O-functionalized low molecular weight components. The results described in Figure 4.5 illustrate that our model provides useful information that could help to elucidate the complex DOM transformational mechanisms at the structural level.

A representation of the DOM structural family (Figure 4.5A) superimposed on the van Krevelen space generated for the sample's CHO class is shown in Figure 4.5B. The discontinuous line patterns described by different directional vectors representing neutral loss-based functionalities contrast with the traditional straight lines utilized in the van Krevelen plot to describe chemical transformations and reaction pathways of DOM components deduced from elemental composition obtained from UHRMS data (Figure 4.1B).[1,36] Therefore, our results suggest that DOM biogeochemical transformation mechanisms are more complex than traditionally described, based upon the heterogeneous nature of the structural information obtained from neutral mass loss patterns observed in this study. For example, DOM molecules assigned from an MS_1 analysis describing a regular addition/subtraction of H_2O chemical units are conventionally interpreted as a family characterized by a hydration/condensation process. Similarly, chemical formulas differing in exactly CO_2 have been also placed into a homologous series resulting from carboxylation/decarboxylation pathways[49]. However, our findings indicate that CHO compounds in this DOM sample form more complex families characterized by multiple

heterogenous combinations of neutral loss based structural moieties (e.g., H_2O, CH_4O and CO_2) such as the one described in Figure 4.5. These results illustrate that the integration of efficient computational tools with comprehensive UHRMS fragmentation workflows allows the identification of valuable structural information of DOM components, that cannot be accurately predicted by traditional FT-MS workflows.

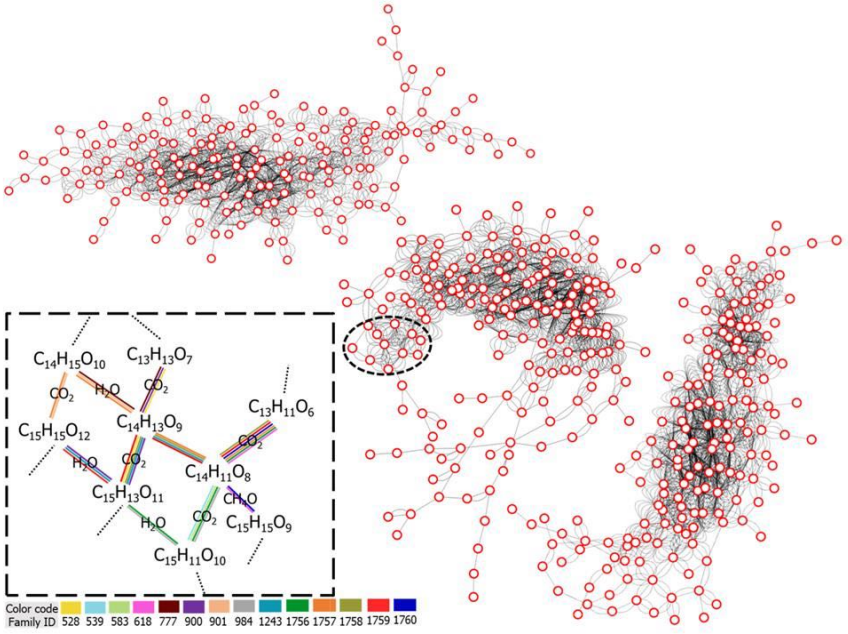


Figure 4. 6. View of the three main clusters observed in the network of neutral-loss based structurally connected DOM precursors for the CHO class. Precursor molecules are described by nodes and the family indexes are shown as edges. An expanded view of fourteen interconnected DOM families is shown as inset. A comprehensive web-based network can be found at https://github.com/Usman095/Graph-DOM.

The visualization of the computed families using Cytoscape confirms the notion that DOM forms a complex assembly of interconnected molecules (Figure 4.6). Similar results using broadband FT-ICR MS data of DOM from both surface and deep sea[26] water samples and from secondary organic aerosols[80] have been reported.

A closer view at the structural network in Figure 4.6 revealed three main clusters of related DOM components (red dots) connected by neutral loss-based structural functionalities (edges). A more detailed analysis of a specific region of the network described in the inset of Figure 4.6, illustrates that several precursors are common to multiple families. This result, not previously observed at the precursor level, shows the crucial role that the structural isomers play in the interconnection of DOM compounds and confirms that isomeric diversity is a fundamental component of DOM molecular complexity. The level of complexity observed in this network suggests that previous elemental-based composition interpretations cannot accurately describe structural patterns in DOM.

In this model the intersection of structural families relates to the isomeric content of DOM. However, it should be note that the model may overestimate the number of fragmentation pathways due to the nominal mass CASI CID data collection. The analysis of the fragmentation pathways determined by nominal mass and chemical formula-based MS/MS for the case of the 267.087412 *m/z* ion ($C_{13}H_{15}O_6$) showed that all nine fragmentation pathways determined by chemical formula-based MS/MS are also observed in the nominal mass analysis (Appendix 4.4). This is an expected result and speaks to the effective processing of the computational code. The nominal mass MS/MS processing resulted in thirteen additional fragmentation channels. While some of the additional fragmentation channels (overestimation) can be derived from differences in the fragmentation mechanism (CASI CID vs SORI CID), the application of the model to nominal mass CASI CID MS/MS will inherently carry potential overestimations.

The analytical power of this workflow is based on the fast acquisition of nominal mass CASI-CID datasets from complex DOM samples. The model applied to nominal mass CASI CID MS/MS effectively reports all the “real” fragmentation pathways. One alternative to reduce the workflow overestimation is to utilize chemical formula-based MS/MS, but this approach is unpractical for routine DOM analysis. A more viable alternative is the implementation of complementary artificial intelligence and machine learning approaches trained with small subsets of chemical formula-based MS/MS data from DOM samples.

4.5 References

1. Dittmar, T.; Stubbins, A. 12.6 - Dissolved Organic Matter in Aquatic Systems. In *Treatise on Geochemistry (Second Edition)*, Holland, H. D.; Turekian, K. K., Eds.; Elsevier: Oxford, 2014; pp 125-156.

2. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016,** *13* (8), 2257-2277.

3. Lu, K.; Liu, Z. Molecular Level Analysis Reveals Changes in Chemical Composition of Dissolved Organic Matter From South Texas Rivers After High Flow Events. *Front. Mar. Sci.* **2019,** *6* (673), 1-18.

4. Kurek, M. R.; Poulin, B. A.; McKenna, A. M.; Spencer, R. G. M. Deciphering Dissolved Organic Matter: Ionization, Dopant, and Fragmentation Insights via Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Environ. Sci. Technol.* **2020,** *54* (24), 16249-16259.

5. Xu, W.; Gao, Q.; He, C.; Shi, Q.; Hou, Z.-Q.; Zhao, H.-Z. Using ESI FT-ICR MS to Characterize Dissolved Organic Matter in Salt Lakes with Different Salinity. *Environ. Sci. Technol.* **2020,** *54* (20), 12929-12937.

6. Pan, Q.; Zhuo, X.; He, C.; Zhang, Y.; Shi, Q. Validation and Evaluation of High-Resolution Orbitrap Mass Spectrometry on Molecular Characterization of Dissolved Organic Matter. *ACS Omega* **2020,** *5* (10), 5372-5379.

7. Cooper, W. T.; Chanton, J. C.; D'Andrilli, J.; Hodgkins, S. B.; Podgorski, D. C.; Stenson, A. C.; Tfaily, M. M.; Wilson, R. M. A History of Molecular Level Analysis of Natural Organic Matter by FTICR Mass Spectrometry and The Paradigm Shift in Organic Geochemistry. *Mass Spectrom. Rev.* **2020**, 1-25.

8. Zark, M.; Christoffers, J.; Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar. Chem.* **2017,** *191*, 9-15.

9. Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Front. Mar. Sci.* **2017,** *4* (405), 1-14.

10. Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J. R.; Tranvik, L. J.; Bergquist, J. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnol. Oceanogr. Lett.* **2018,** *3* (2), 21-30.

11. Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the structural complexity of dissolved organic matter: isomeric diversity. *Faraday Discuss.* **2019,** *218* (0), 431-440.

12. Leyva, D.; Jaffe, R.; Fernandez-Lima, F. Structural Characterization of Dissolved Organic Matter at the Chemical Formula Level Using TIMS-FT-ICR MS/MS. *Anal. Chem.* **2020,** *92* (17), 11960-11966.

13. Lam, B.; Baer, A.; Alaee, M.; Lefebvre, B.; Moser, A.; Williams, A.; Simpson, A. J. Major Structural Components in Freshwater Dissolved Organic Matter. *Environ. Sci. Technol.* **2007,** *41* (24), 8240-8247.

14. Woods, G. C.; Simpson, M. J.; Koerner, P. J.; Napoli, A.; Simpson, A. J. HILIC-NMR: Toward the Identification of Individual Molecular Components in Dissolved Organic Matter. *Environ. Sci. Technol.* **2011,** *45* (9), 3880-3886.

15. Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **2013,** *10* (3), 1583-1624.

16. Zhang, F.; Harir, M.; Moritz, F.; Zhang, J.; Witting, M.; Wu, Y.; Schmitt-Kopplin, P.; Fekete, A.; Gaspar, A.; Hertkorn, N. Molecular and structural characterization of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry. *Water Res.* **2014,** *57*, 280-294.

17. Hertkorn, N.; Harir, M.; Schmitt-Kopplin, P. Nontarget analysis of Murchison soluble organic matter by high-field NMR spectroscopy and FTICR mass spectrometry. *Magn. Reson. Chem.* **2015,** *53* (9), 754-768.

18. Simpson, A. J.; Kingery, W. L.; Hatcher, P. G. The Identification of Plant Derived Structures in Humic Materials Using Three-Dimensional NMR Spectroscopy. *Environ. Sci. Technol.* **2003,** *37* (2), 337-342.

19. Reemtsma, T.; These, A.; Linscheid, M.; Leenheer, J.; Spitzy, A. Molecular and Structural Characterization of Dissolved Organic Matter from the Deep Ocean by FTICR-MS, Including Hydrophilic Nitrogenous Organic Molecules. *Environ. Sci. Technol.* **2008,** *42* (5), 1430-1437.

20. Cortés-Francisco, N.; Caixach, J. Fragmentation studies for the structural characterization of marine dissolved organic matter. *Anal. Bioanal. Chem.* **2015,** *407* (9), 2455-2462.

21. Lu, K.; Gardner, W. S.; Liu, Z. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). *Environ. Sci. Technol.* **2018,** *52* (13), 7182-7191.

22. Zark, M.; Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **2018,** *9* (1), 3178.

23. Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **2006,** *70* (12), 2990-3010.

24. Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the analytical limits: new insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. *Chem. Sci.* **2019,** *10* (29), 6966-6978.

25. Le Maître, J.; Hubert-Roux, M.; Paupy, B.; Marceau, S.; Rüger, C. P.; Afonso, C.; Giusti, P. Structural analysis of heavy oil fractions after hydrodenitrogenation by high-resolution tandem mass spectrometry and ion mobility spectrometry. *Faraday Discuss.* **2019,** *218* (0), 417-430.

26. Longnecker, K.; Kujawinski, E. B. Using network analysis to discern compositional patterns in ultrahigh-resolution mass spectrometry data of dissolved organic matter. *Rapid Commun. Mass Spectrom.* **2016,** *30* (22), 2388-2394.

27. Herzsprung, P.; Wentzky, V.; Kamjunke, N.; von Tümpling, W.; Wilske, C.; Friese, K.; Boehrer, B.; Reemtsma, T.; Rinke, K.; Lechtenfeld, O. J. Improved Understanding of Dissolved Organic Matter Processing in Freshwater Using Complementary Experimental and Machine Learning Approaches. *Environ. Sci. Technol.* **2020,** *54* (21), 13556-13565.

28. Vialykh, E. A.; McKay, G.; Rosario-Ortiz, F. L. Computational Assessment of the Three-Dimensional Configuration of Dissolved Organic Matter Chromophores and Influence on Absorption Spectra. *Environ. Sci. Technol.* **2020,** *54* (24), 15904-15913.

29. Devarajan, D.; Liang, L.; Gu, B.; Brooks, S. C.; Parks, J. M.; Smith, J. C. Molecular Dynamics Simulation of the Structures, Dynamics, and Aggregation of Dissolved Organic Matter. *Environ. Sci. Technol.* **2020,** *54* (21), 13527-13537.

30. Kujawinski, E. B. Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS): Characterization of Complex Environmental Mixtures. *Environ. Forensics* **2002,** *3* (3), 207-216.

31. Kujawinski, E. B.; Freitas, M. A.; Zang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. *Org. Geochem.* **2002,** *33* (3), 171-180.

32. Sleighter, R. L.; Hatcher, P. G. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* **2007,** *42* (5), 559-574.

33. Sleighter, R. L.; Hatcher, P. In Fourier Transform Mass Spectrometry for the Molecular Level Characterization of Natural Organic Matter: Instrument Capabilities, Applications, and Limitations, 2011.

34. Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Anal. Bioanal. Chem.* **2007,** *389* (5), 1311-1327.

35. Smith, D. F.; Podgorski, D. C.; Rodgers, R. P.; Blakney, G. T.; Hendrickson, C. L. 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures. *Anal. Chem.* **2018,** *90* (3), 2041-2047.

36. Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal. Chem.* **2003,** *75* (20), 5336-5344.

37. Wu, Z.; Rodgers, R. P.; Marshall, A. G. Two- and Three-Dimensional van Krevelen Diagrams: A Graphical Analysis Complementary to the Kendrick Mass Plot for Sorting Elemental Compositions of Complex Organic Mixtures Based on Ultrahigh-Resolution Broadband Fourier Transform Ion Cyclotron Resonance Mass Measurements. *Anal. Chem.* **2004,** *76* (9), 2511-2516.

38. Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol. Oceanogr.* **2010,** *55* (4), 1467-1477.

39. D'Andrilli, J.; Cooper, W. T.; Foreman, C. M.; Marshall, A. G. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Commun. Mass Spectrom.* **2015,** *29* (24), 2385-2401.

40. Nebbioso, A.; Piccolo, A. Molecular characterization of dissolved organic matter (DOM): a critical review. *Anal. Bioanal. Chem.* **2013,** *405* (1), 109-124.

41. Kendrick, E. A Mass Scale Based on CH2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **1963,** *35* (13), 2146-2154.

42. Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001,** *73* (19), 4676-4681.

43. Roth, V.-N.; Dittmar, T.; Gaupp, R.; Gleixner, G. The Molecular Composition of Dissolved Organic Matter in Forest Soils as a Function of pH and Temperature. *PLOS ONE* **2015,** *10* (3), e0119188.

44. Koch, B. P.; Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **2006,** *20* (5), 926-932.

45. Stenson, A. C.; Landing, W. M.; Marshall, A. G.; Cooper, W. T. Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2002,** *74* (17), 4397-4409.

46. Witt, M.; Fuchser, J.; Koch, B. P. Fragmentation Studies of Fulvic Acids Using Collision Induced Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2009,** *81* (7), 2688-2694.

47. LeClair, J. P.; Collett, J. L.; Mazzoleni, L. R. Fragmentation Analysis of Water-Soluble Atmospheric Organic Matter Using Ultrahigh-Resolution FT-ICR Mass Spectrometry. *Environ. Sci. Technol.* **2012,** *46* (8), 4312-4322.

48. These, A.; Winkler, M.; Thomas, C.; Reemtsma, T. Determination of molecular formulas and structural regularities of low molecular weight fulvic acids by size-exclusion chromatography with electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004,** *18* (16), 1777-1786.

49. Stenson, A. C.; Marshall, A. G.; Cooper, W. T. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Anal. Chem.* **2003,** *75* (6), 1275-1284.

50. Doerr, A. DIA mass spectrometry. *Nat. Methods* **2015,** *12* (1), 35-35.

51. Pegg, C. L.; Phung, T. K.; Caboche, C. H.; Niamsuphap, S.; Bern, M.; Howell, K.; Schulz, B. L. Quantitative data independent acquisition glycoproteomics of sparkling wine. *Mol. Cell. Proteom.* **2020,** *20*, 1-12.

52. Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020,** *20* (17-18), 1900276.

53. Ross, C. W.; Simonsick, W. J.; Aaserud, D. J. Application of Stored Waveform Ion Modulation 2D-FTICR MS/MS to the Analysis of Complex Mixtures. *Anal. Chem.* **2002,** *74* (18), 4625-4633.

54. van Agthoven, M. A.; Delsuc, M.-A.; Bodenhausen, G.; Rolando, C. Towards analytically useful two-dimensional Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Bioanal. Chem.* **2013,** *405* (1), 51-61.

55. Bray, F.; Bouclon, J.; Chiron, L.; Witt, M.; Delsuc, M.-A.; Rolando, C. Nonuniform Sampling Acquisition of Two-Dimensional Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Increased Mass Resolution of Tandem Mass Spectrometry Precursor Ions. *Anal. Chem.* **2017,** *89* (17), 8589-8593.

56. Pfändler, P.; Bodenhausen, G.; Rapin, J.; Houriet, R.; Gäumann, T. Two-dimensional fourier transform ion cyclotron resonance mass spectrometry. *Chem. Phys. Lett.* **1987,** *138* (2), 195-200.

57. Ross, C. W.; Guan, S.; Grosshans, P. B.; Ricca, T. L.; Marshall, A. G. Two-dimensional Fourier transform ion cyclotron resonance mass spectrometry/mass spectrometry with stored-waveform ion radius modulation. *J. Am. Chem. Soc.* **1993,** *115* (17), 7854-7861.

58. van Agthoven, M. A.; Coutouly, M.-A.; Rolando, C.; Delsuc, M.-A. Two-dimensional Fourier transform ion cyclotron resonance mass spectrometry: reduction of scintillation noise using Cadzow data processing. *Rapid Commun. Mass Spectrom.* **2011,** *25* (11), 1609-1616.

59. Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D. H.; Marshall, A. G. External accumulation of ions for enhanced electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **1997,** *8* (9), 970-976.

60. Lakshmanan, R.; Wolff, J. J.; Alvarado, R.; Loo, J. A. Top-down protein identification of proteasome proteins with nanoLC-FT-ICR-MS employing data-independent fragmentation methods. *Proteomics* **2014,** *14* (10), 1271-1282.

61. Wenke, J. L.; Rose, K. L.; Spraggins, J. M.; Schey, K. L. MALDI Imaging Mass Spectrometry Spatially Maps Age-Related Deamidation and Truncation of Human Lens Aquaporin-0. *Investig. Ophthalmol. Vis. Sci.* **2015,** *56* (12), 7398-7405.

62. Sleighter, R. L.; McKee, G. A.; Hatcher, P. G. Direct Fourier transform mass spectral analysis of natural waters with low dissolved organic matter. *Org. Geochem.* **2009,** *40* (1), 119-125.

63. Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **2008,** *6* (6), 230-235.

64. Heck, A. J. R.; de Koning, L. J.; Pinkse, F. A.; Nibbering, N. M. M. Mass-specific selection of ions in Fourier-transform ion cyclotron resonance mass spectrometry. Unintentional off-resonance cyclotron excitation of selected ions. *Rapid Commun. Mass Spectrom.* **1991,** *5* (9), 406-414.

65. de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass selection of ions in a Fourier transform ion cyclotron resonance trap using correlated harmonic excitation fields (CHEF). *Int. J. Mass Spectrom. Ion Processes* **1997,** *165-166*, 209-219.

66. Herzsprung, P.; Hertkorn, N.; von Tümpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Anal. Bioanal. Chem.* **2014,** *406* (30), 7977-7987.

67. Simon, C. P., D.; Roth, V.-N.; Dührkop, K.; Böcker, S.; Dorrestein, P.; Gleixner, G. Mass difference matching crystallizes hidden molecular structures of dissolved organic matter from ultrahigh-resolution tandem mass spectra. *ChemRxiv* **2021**, 1-26.

68. Osterholz, H.; Niggemann, J.; Giebel, H.-A.; Simon, M.; Dittmar, T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nat. Commun.* **2015,** *6* (1), 7422.

69. Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P.-L.; Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2010,** *27* (3), 431-432.

70. Minor, E. C.; Swenson, M. M.; Mattson, B. M.; Oyler, A. R. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci. Process. Impacts.* **2014,** *16* (9), 2064-2079.

71. Capley, E. N.; Tipton, J. D.; Marshall, A. G.; Stenson, A. C. Chromatographic Reduction of Isobaric and Isomeric Complexity of Fulvic Acids To Enable Multistage Tandem Mass Spectral Characterization. *Anal. Chem.* **2010,** *82* (19), 8194-8202.

72. Benk, S. A.; Li, Y.; Roth, V.-N.; Gleixner, G. Lignin Dimers as Potential Markers for 14C-young Terrestrial Dissolved Organic Matter in the Critical Zone. *Front. Earth Sci.* **2018,** *6* (168), 1-9.

73. Yuan, Z.; He, C.; Shi, Q.; Xu, C.; Li, Z.; Wang, C.; Zhao, H.; Ni, J. Molecular Insights into the Transformation of Dissolved Organic Matter in Landfill Leachate Concentrate during Biodegradation and Coagulation Processes Using ESI FT-ICR MS. *Environ. Sci. Technol.* **2017,** *51* (14), 8110-8118.

74. Antony, R.; Willoughby, A. S.; Grannas, A. M.; Catanzano, V.; Sleighter, R. L.; Thamban, M.; Hatcher, P. G.; Nair, S. Molecular Insights on Dissolved Organic Matter Transformation by Supraglacial Microbial Communities. *Environ. Sci. Technol.* **2017,** *51* (8), 4328-4337.

75. Ward, C. P.; Cory, R. M. Assessing the prevalence, products, and pathways of dissolved organic matter partial photo-oxidation in arctic surface waters. *Environ. Sci. Process. Impacts.* **2020,** *22* (5), 1214-1223.

76. Ward, C. P.; Cory, R. M. Complete and Partial Photo-oxidation of Dissolved Organic Matter Draining Permafrost Soils. *Environ. Sci. Technol.* **2016,** *50* (7), 3545-3553.

77. Kim, S.; Kaplan, L. A.; Hatcher, P. G. Biodegradable dissolved organic matter in a temperate and a tropical stream determined from ultra-high resolution mass spectrometry. *Limnol. Oceanogr.* **2006,** *51* (2), 1054-1063.

78. Mostovaya, A.; Hawkes, J. A.; Dittmar, T.; Tranvik, L. J. Molecular Determinants of Dissolved Organic Matter Reactivity in Lake Water. *Front. Earth Sci.* **2017,** *5* (106), 1-13.

79. Hawkes, J. A.; Hansen, C. T.; Goldhammer, T.; Bach, W.; Dittmar, T. Molecular alteration of marine dissolved organic matter under experimental hydrothermal conditions. *Geochim. Cosmochim. Acta* **2016,** *175*, 68-85.

80. Tziotis, D.; Hertkorn, N.; Schmitt-Kopplin, P. Kendrick-Analogous Network Visualisation of Ion Cyclotron Resonance Fourier Transform Mass Spectra: Improved

Options for the Assignment of Elemental Compositions and the Classification of Organic Molecular Complexity. *Eur. J. Mass Spectrom.* **2011,** *17* (4), 415-421.

CHAPTER V

V. MOLECULAR LEVEL CHARACTERIZATION OF DOM ALONG A FRESHWATER-TO-ESTUARINE COASTAL GRADIENT IN THE FLORIDA EVERGLADES

5.1 Abstract

Understanding dissolved organic matter (DOM) export to the ocean is needed to assess the impact of climate change on the global carbon cycle. The molecular-level characterization of DOM compositional variability and complexity in aquatic ecosystems has been analytically challenging. Advanced analytical studies based on ultra-high resolution mass spectrometry (FT ICR MS) have proven highly successful to better understand the dynamics of DOM in coastal ecosystems. In this work, the molecular signature of DOM along a freshwater-to-estuarine gradient in the Harney River, Florida Everglades was analyzed for the first time using a novel approach based on tandem high resolution ion mobility and ultra-high resolution mass spectrometry (ESI-TIMS-FT ICR MS). This method enhances traditional DOM molecular characterization by including the molecular isomeric complexity. An average of six and up to 12 isomers were observed per chemical formula and characteristic isomers to each section of the freshwater-to-estuarine gradient were successfully identified. We measured a decrease on the chemical complexity and diversity (both in the number of molecular formulas and number of isomers per chemical formula) with increasing salinity; this trend is representative of the biogeochemical transformations of DOM during transport and along source variations, showing both clear degradation products and formation of new components along the salinity transect. The inclusion of the isomeric content at the molecular formula allowed to differentiate isomeric species that are present along the transect (mainly lignin-type components) and responsible for the DOM refractory nature. DOM isomeric fingerprints characteristic of the molecular variability along the Everglades freshwater-to-estuarine gradient are also described.

5.2 Introduction

Dissolved organic matter (DOM) is a complex mixture of chemical compounds and constitutes a large quantity of organic carbon in aquatic ecosystems[1,2]. The characterization of DOM has been widely used as a proxy to understand the biogeochemical dynamics of aquatic environments including complex phenomena such as the impact of land-use and climate change[3-7]. In coastal wetlands, DOM has been the target of a multitude of studies due to the significant amount of organic carbon exported to the ocean[6,8-11].

A wide variety of analytical techniques have been applied to obtain comprehensive information on the spatial and temporal variability of DOM in aquatic systems. Traditional measurements of dissolved organic carbon concentrations and other water quality parameters along with optical properties (UV-VIS and fluorescence) have been the most explored methods for bulk characterization of DOM[4,12-17]. More advanced analytical studies based on nuclear magnetic resonance (NMR) and ultra-high, resolution mass spectrometry (FT-ICR MS) have proved to be critical to understanding chemical transformations of DOM and linking sources with compositional signatures[18-23].

The Florida Everglades is a complex mosaic of wetland ecosystems connected hydrologically through both natural and human-altered processes that are changing biogeochemical patterns across the landscape[24,25]. Although several studies have approached the influence of environmental factors on the variability of DOM in the Everglades[7,25-30], few of them explored detailed molecular level analysis[31,32]. For example, Hertkorn et.al[31] conducted a comparison of DOM samples from various wetlands, including the Everglades, integrating optical properties including EEM-PARAFAC, as well as NMR and ultrahigh-resolution mass spectrometry demonstrating for the first time

the extensive molecular diversity and compositional complexity of DOM in wetlands. Despite similar compositional patterns found across the samples, distinctive signatures could be explained by combinations of environmental drivers such as source variations, wildfires, and anthropogenic influences. Regarding the latter (mainly observed for CHOS species), anthropogenic sources associated with agriculture were primarily attributed to the singular DOM sulfur classes found. Combining optical properties, including PARAFAC fluorescence, and FT-ICR MS, Wagner et. al.[32] studied DOM from Shark River Slough and Taylor Slough in the Everglades. Several thousands of chemical compounds mainly distributed in the CHO, CHON, and CHOS heteroatom classes were detected across the samples and were closely correlated with the UV and fluorescent properties of the DOM, allowing for a good estimate of the origin and variability of DOM in this coastal ecosystem.

Recent efforts have increased the analytical power of ultra-high resolution mass spectrometry with the coupling of complementary separation techniques[33-39]. However, the high DOM chemical complexity and diversity has proven analytically challenging due to the close structural similarity at the molecular level. An alternative to traditional separations is the use of ion mobility spectrometry in tandem with mass spectrometry (IMS-MS) with focus on DOM separations[38,40-51]. For instance, a dissimilarity at the isomeric level of samples from freshwater and marine endmembers was clearly observed using IMS-MS experiments[49]; a limiting factor has been the relatively low IMS and MS resolution of commercial IMS-MS instruments[48,49,52,53]. Our previous research has developed novel analytical workflows that integrate high-resolution mobility analyzers (e.g., trapped ion mobility spectrometry, TIMS) with ultra-high mass resolution (e.g., Fourier transformed ion cyclotron resonance mass spectrometry, FT-ICR MS) for the

molecular characterization of DOM[38,45,46,54]. In TIMS, gas phase ionized species are trapped and separated in space using a DC electric field applied across the funnel and tunnel axis. Packages of ions with a given ion mobility are gated out of the TIMS cell and accumulated in the collision cell before their analysis in the ICR cell[55-57]. The high resolving power of the TIMS cell (R > 400 reported), along with the high mass accuracy and ultra-high mass resolution of the FT-ICR MS, have allowed a better description of the isomeric complexity of DOM[56-58]. Recently, we advanced the TIMS-FT-ICR MS/MS capabilities by generating mobility selected chemical formula fragmentation pathways and providing candidate structures based on their collisional cross sections (CCS) and fragmentation patterns[45,46].

Previous works have demonstrated the applicability of advanced analytical methods to advance our understanding of Everglades DOM biogeochemistry and have shown the necessity to further expand such studies and explore novel approaches to better understand how environmental drivers affect the DOM dynamics at the molecular level in vulnerable coastal ecosystems[31,32]. Here, we conduct the first-ever characterization of the isomeric complexity of DOM along a freshwater-to-estuarine gradient in the Harney River, Florida Everglades, using a novel approach based on tandem high resolution ion mobility and ultra-high resolution mass spectrometry (ESI-TIMS-FT ICR MS). Unique and common DOM molecular signatures have been successfully identified across the transect as a proxy to understand biogeochemical processes, the role of input sources in DOM spatial variability, and DOM involvement in near coastal processes.

5.3 Experimental

5.3.1 Sample collection and treatment

Water samples were collected along a salinity transect of the Harney River, Everglades National Park (Figure 5.1), at the end of the subtropical dry season (April-May 2021). The Harney River exports fresh water from the upstream Everglades marshes, through a mangrove estuary, and out into the Gulf of Mexico. Salinity was measured on-site using a YSI sensor (YSI, Yellow Springs, OH, USA). Surface water was collected 50 cm below the surface and stored in HCl pre-washed 2-L amber plastic bottles (Nalgene®, Whaltham, MA, USA). Samples were kept on ice during transportation to the lab and filtered (0.22-μm, Whatman, Maidstone, UK) within two days after collection. Filtered samples were stored at 4°C before treated by a solid-phase-extraction (SPE) procedure following the protocol developed by Dittmar et.al.[59]. Briefly, one liter of sample previously acidified to pH 2 with Hydrochloric Acid was loaded onto a 1g-Bond Elut PPL cartridge (Agilent, Santa Clara, CA , USA) preconditioned with one-cartridge volume of methanol followed by 2-cartridges-volume of water (pH 2). The loaded cartridge was then rinsed with water (pH 2) and dried with N_2 gas for five minutes prior to the elution of DOM molecules with 20 mL of methanol. SPE methanol extracts were stored in pre-washed glass vials at −20 °C. For comparison purposes, a dilution was used based on the estimated DOC content of the SPE extracts (less that 2x variation across samples). The purpose of the dilution is to have similar ionization and matrix effects across the transient samples. All samples were diluted to a final concentration of ~ 12 mg/l C in a 1:9 methanol:ethanol v:v solution mixture. All solvents used were of Optima LC-MS grade obtained from Fisher Scientific (Pittsburgh, PA, USA). Dissolved Organic Carbon (DOC) concentrations were analyzed at

the NELAC-accredited CAChE Nutrient Analysis Core facility at Florida International University.

5.3.2 ESI-TIMS-FT ICR MS

A custom-built ESI-TIMS-FT-ICR MS platform based on a Solarix 7T FT-ICR MS equipped with an infinity ICR cell (Bruker Daltonics Inc., MA) was used for the experiments. An electrospray ionization source (Apollo II ESI design, Bruker Daltonics, Inc., MA) was utilized in negative ion mode and sample solutions were infused at 130 μL/h. Typical operating conditions were 3200-3500 V capillary voltage, 4 L/min dry gas flow rate, 1.0 bar nebulizer gas pressure, and a dry gas temperature 180°C. The instrument was optimized for high transmission of ions in the 100-1200 m/z range. Typical operational parameters included funnel rf amplitude 220 peak-to-peak voltage (Vpp), capillary exit -100 V, deflector plate -90 V, skimmer1 -60 V, transfer line rf 350 Vpp, octupole rf amplitude 350 Vpp and collision cell rf 1000 Vpp. Mass spectra were collected with a 4 MW data acquisition size for a mass resolution of ~ 300K at 400 m/z. Ions were accumulated for 20 seconds in the instrument collision cell before their transmission to the ICR cell. Experimental run time was about 3 hours using the FT-ICR MS serial mode. Solvent blanks were analyzed under the same experimental conditions as the samples and mass peaks detected in the blanks were removed from the samples mass lists. Agilent Tuning Mix calibration standard was used during the instrument tuning and control optimization.

The fundamentals on the gas phase ion TIMS separation and the coupling the FT-ICR MS can be found elsewhere[38,54,60,61]. Briefly, an electric field is applied in the TIMS cell to hold ions stationary against a flowing gas, so that the drag force is evened with the electric

field and ions are spatially separated across the TIMS analyzer axis based on their mobility[56-58]. A quadrupolar field is also used during the gas phase separation to confine the ions radially, thus enhancing the trapping efficiency. The mobility, K_0, of an ion in a TIMS cell is described by the equation (1):

$$K_0 = \frac{v_g}{E} = \frac{A}{V_{elution} - V_{out}} \qquad (1)$$

where v_g, E, A, $V_{elution}$, and V_{out} are the velocity of the gas, electric field, a calibration constant, elution voltage, and tunnel out voltage, respectively.

Values of K_0 can be converted into ion–neutral CCS (Ω, Å^2) using the Mason–Schamp equation (2):

$$\Omega = \frac{(18\pi)^{1/2} z}{16(k_B T)^{1/2}} \left(\frac{1}{m_i} + \frac{1}{m_b}\right)^{1/2} \frac{1760\, T}{Ko\ P\ 273.15\ N^*} \qquad (2)$$

where z is the charge of the ion, k_B is the Boltzmann constant, N^* is the number density, and m_i and m_b are the masses of the ion and bath gas, respectively[62].

The TIMS analyzer was controlled using an in-house software, written in National Instruments LabVIEW, and synchronized with the FTMS control acquisition program. Ion mobility separation was conducted in the oversampling mode[38,45,46,54] using nitrogen as a bath gas at ca 300 K, P_1 = 2.4 mbar and P_2 = 1.0 mbar, and a 220 V_{pp} rf. Briefly, an ion mobility range that is sampled in a TIMS analyzer is defined by the voltage ramp (V_{ramp}). A voltage step increment (V_{step}) is used to sample the following ion mobility range until the full ion mobility space of the sample is interrogated. In the oversampling mode ($V_{step} << V_{ramp}$), the same ion mobility range is sampled multiple times, thus increasing the number of analytical points across a mobility peak. The TIMS cell was operated using a simultaneous fill/trap sequence synchronized with the accumulation during detect mode of

the FTMS control software. A voltage difference of 4 V across the ΔE gradient and a 0.2 V stepping scan function across the total ΔV_{ramp} 10 V-110 V were utilized. The deflector (V_{def}), funnel entrance (V_{fun}) and funnel out (V_{out}) voltages were as follow: V_{def} = −110/50 V, V_{fun} = 0/0 V and V_{out}= −6V/30 V. A maximum of 1,000 IMS scans were collected per mass spectrum.

5.3.3 Data analysis

ESI-TIMS-FT-ICR MS data was processed using Data Analysis (v. 5.2, Bruker Daltonics, City, CA, USA). Chemical formula assignment was conducted using Composer software (version 1.0.6, Sierra Analytics, City, CA, USA) and confirmed with Data Analysis (version 5.2, Bruker Daltonics). Molecular formulas with the lowest errors, isolated assignments removal (de-assignment of peaks belonging to classes with only a few sparsely scattered members) and the presence of isotopologue signals were used as validation parameters. Theoretical formula constraints of $C_{4\text{-}50}H_{4\text{-}100}N_{0–3}O_{0–25}S_{0–2}$, S/N>3, m/z range 100-650, error <1 ppm and 0<O/C<1, 0.3<H/C<2.5, and DBE-O < 10[63] were considered. Near 50 mass peaks of predefined oxygen homologous series (O_3-O_{10}) compounds spaced approximately 14 mass units across the 200-600 *m/z* range were utilized for internal calibration in Composer. The walking internal recalibration resulted in an average error < 110 ppb for the mass range 200-600 Da. The TIMS spectra were externally calibrated for mobility using the Agilent Tuning Mix calibration standard[64]. The number of isomers per chemical formula are estimated by deconvoluting (fitting gaussian peaks under the IMS profiles) the extracted ion mobiliograms of the assigned chemical formulas in the samples. This process was conducted by using a custom-built Software Assisted Molecular Elucidation (SAME) written in Python v3.7.3. The SAME algorithm utilizes

noise removal, mean gap filling, asymmetric least squares smoothing for base line correction, continuous wavelet transform (CWT)-based peak detection (SciPy package), and Gaussian fitting with non-linear least squares functions (Levenberg–Marquardt algorithm)[60]. Plots were created using Microsoft Excel 365, OriginPro 2016 (Originlab Co., MA), and Python 3.7.3. In order to compare the molecular signatures of DOM along the transect, isomeric weighted averages of specific molar ratios (H/C, O/C, #N), modified aromatic index (AI_{mod}) were calculated for all samples based on the number of isomers assigned to each molecular formula using a similar approach as the ones previously reported[20,65]. Chemical formulas identified across DOM samples were tagged with their underlaying isomeric information ($^{TIMS}CCS_{N2}$ values). A comparison of the chemical formulas- $^{TIMS}CCS_{N2}$ across DOM samples considering an 8% matching error on $^{TIMS}CCS_{N2}$ values was conducted using mathematical sets in Python 3.7.3.

5.4 Results and discussion

The ESI(-)-TIMS- FT ICR MS analysis of the five SPE-DOM samples collected along the salinity transect of the Harney River estuary resulted in the typical single, broad trend line of $[M-H]^-$ molecular species ($1/K_0$ range: 0.6-1.2 Vs/cm^2) in the 2D IMS-MS domain (Figure 5.2A). Overall, a Gaussian-like MS distributions across 190-650 *m/z* with the apex centered at ~ 370 *m/z* were found for all samples (See Figure 5.2A insets). Interestingly, a slightly different profile depleted in mass signals over 400 m/z was found for the sample HR-4, which may suggest a change in the compositional nature of this sample.

Around 70% of the chemical formulas detected at the freshwater marsh station were identified at the marine end-member. This decreasing trend in molecular diversity

resonates with a declining pattern of CDOM (UV absorbance, 254 nm) previously observed in the same system [26]. The change in DOM molecular diversity perceived along the Harney River could be partially explained by the fact that much of the DOM in the upper headwaters of the Harney River is derived from the Everglades freshwater marshes. This freshwater signal is reduced along the salinity transect through biogeochemical processing including bio- and photo- degradation, while a smaller but significant amount of DOM components is added from the mangrove marsh and the marine end member[26]. Other studies[9,66] have reported similar trends for more geographically extensive watersheds (compared to this study) and suggested this pattern to be driven by the river continuum concept (RCC), where low stream orders (headwaters) feature more chemically complex DOM than high stream orders (estuary) since more labile DOM components may degrade downstream[67,68]. In agreement with the literature, the difference in molecular diversity observed at the station HR-5 (1625 formulas) compared to HR-4 (1373 formulas), suggests a potential contribution of mangrove and marine-derived DOM at the marine endmember.

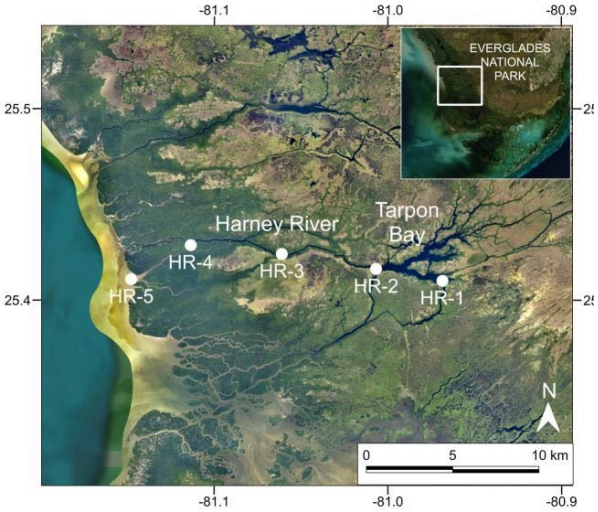


Figure 5. 1. Map of sampling points located at the Harney River, Everglades National Park.

An increase in unsaturation of higher molecular mass and size DOM components was clearly observed (Figure 5.2B). The decrease in the number of compounds with higher m/z and DBE values observed along the salinity gradient suggests that highly unsaturated DOM components are mainly derived from the Everglades freshwater marsh environment and are preferentially transformed during their transport from land to sea. A detailed inspection of Table 5.1 confirmed a decrease of DBE_w and an increase in the H/C_w along the transect. Previous reports have concluded that unsaturated aromatic DOM compounds are favorably bleached during the photodegradation of DOM samples from fresh/brackish waters[66,69-73]. Furthermore, it has been found that molecular species resulting from photodegraded freshwater DOM resembled the composition of higher salinity samples[70,74]. A closer view at the van Krevelen plots in Figure 5.2C evidenced that the removal of unsaturated, tannins-like compounds at the freshwater endmember significantly contributed to the overall decreasing trend observed in aromaticity and degree of unsaturation along the transect. These findings suggest that photodegradation plays a crucial role during DOM transport and mineralization in the Harney River.

Table 5. 1. Summary of the isomeric information obtained for the Harney River DOM samples using TIMS-FT-ICR MS.

Parameter	Sample ID				
	HR-1	HR-2	HR-3	HR-4	HR-5
m/z_w	414.36	406.47	396	379.23	378.23
C_w	18.56	19.47	18.84	18.15	17.89
N_w	0.35	0.42	0.41	0.42	0.39
O_w	10.09	8.57	8.38	7.92	7.97
S_w	0.24	0.22	0.26	0.23	0.3
H/C_w	1.00	1.21	1.22	1.24	1.23
O/C_w	0.55	0.45	0.46	0.45	0.46
N/C_w	0.02	0.03	0.03	0.03	0.03
S/C_w	0.02	0.01	0.02	0.02	0.02
DBE_w	10.51	8.99	8.61	8.16	8.1

$AImod_w$	0.143	0.115	0.107	0.106	0.11
Total isomers	22185	18236	12549	10414	13256
CHO isomers	11783	11027	7591	6817	7475
CHO isomers%	53.1	60.5	60.5	65.5	56.4
CHON isomers	5109	3127	1733	1228	1848
CHON isomers%	23.0	17.1	13.8	11.8	13.9
CHOS isomers	5293	2981	2359	1527	3206
CHOS isomers%	23.9	16.3	18.8	14.7	24.2
CHONS isomers	0	1101	866	842	727
CHONS isomers%	0.0	6.0	6.9	8.1	5.5

The projection of the elemental ratios in the compositional space revealed a set of core molecules in the lignin-type area of the van Krevelen plot that remained unaltered along the salinity transect (See Figure 5.2C). Previous studies using optical properties, in particular EEM PARAFAC, have shown that while some components increased or decreased along the salinity gradient (suggesting additional contributions from the mangrove marshes and seagrass, and DOM removal resulting from degradation processes), one PARAFAC component remained unaltered showing high resistance to degradation and a conservative mixing behavior along the salinity transect[26]. The unaltered and seemingly resistant isomers observed in this study may be related to this PARAFAC component. These DOM components have been previously associated with a constrained set of terrestrial recalcitrant material (e.g. aliphatic and carboxylic-rich alicyclic molecules) that partially resist degradation[74]. However, a visual inspection of the plots in Figure 5.2C indicates that an important set of CHO, CHON, and CHOS molecular species change significantly from the freshwater marshes to the estuary. A detailed analysis of Table 5.1 confirmed an overall decrease in the number of compounds for all the heteroatom classes along the transect. Interestingly, while heteroatom components are clearly bio- or photo-

degraded along the salinity gradient, the enriched molecular complexity observed at the lower estuary (site HR-5) compared with HR-4 is primarily attributed to an increase in CHON (~30%) and CHOS (~50%) molecular formulas. The source for these heteroatomic components at higher salinity is usually associated with DOM inputs from the mangrove swamps and/or marine vegetation, or biogeochemical transformations of DOM components transported along the transect[9,20,66,75]. Although there is no documented ultrahigh-resolution mass spectrometry analysis of the Harney River DOM, a previous report based on optical properties and stable isotopes found significant contributions of seagrass DOM to the pool of DOC in several regions of Florida Bay[30], and a biomarker-based study of sediments in the Harney River estuary reported on seagrass as a potential source of sedimentary organic matter in the estuary[76]. A molecular-level analysis of DOM performed at a terrestrial-salinity transect in Elizabeth River, Virginia, correlated the significant input of CHON and CHONS molecular species at the estuary with a source of primary producers[66]. In contrast, the enrichment of CHOS components to the DOM signature at the mangrove-estuary region has not been clearly elucidated[32,66].

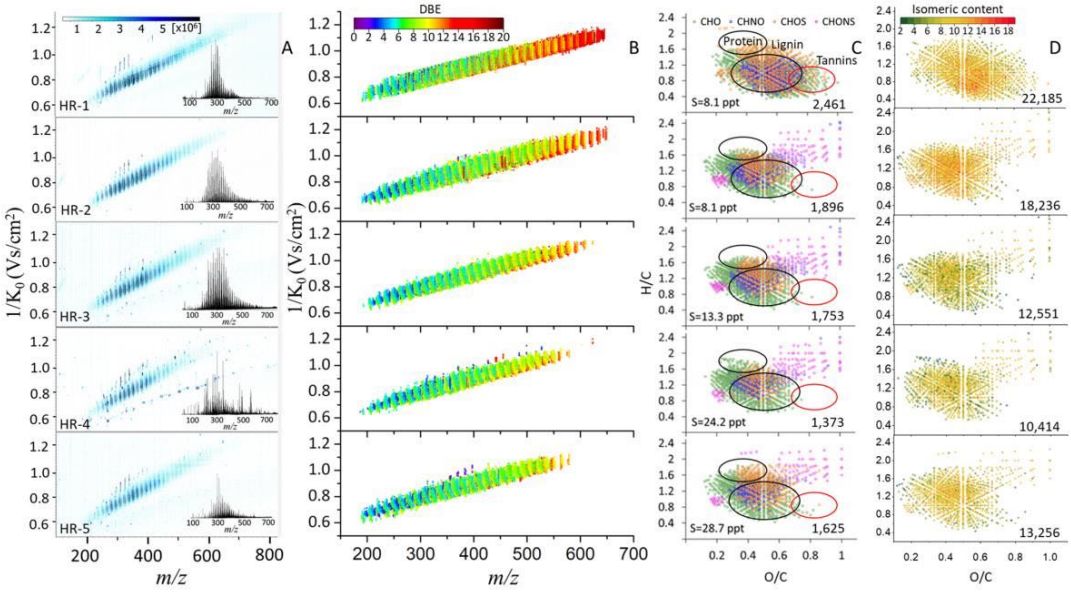


Figure 5. 2. 2D IMS-MS profiles of Harney River SPE-DOM samples obtained from ESI-TIMS-FT ICR MS (A). Broadband MS[1] projections shown as insets. 2D IMS-MS profiles including DBE distribution of molecular formulas assigned in each sample. Note the consistent decrease of highly unsaturated molecules (red color) along the transect (B). Van Krevelen plots of the assigned molecular species (chemical classes indicated by colors) denoting a change in the molecular complexity of the samples as a function of salinity (C). Van Krevelen plots indicating a decrease in the isomeric content from freshwater DOM to the estuary influenced DOM (D).

A decreasing pattern, analogous to the one observed for molecular formulas, was found for the estimated number of isomers along the salinity transect (Figure 5.3). Nearly two-fifth of the isomers detected at the freshwater end-member were not found above the signal-to-noise ratio defined or transformed downstream, in contrast with a freshwater-to-estuarine isomeric complexity enrichment in a salinity gradient in a tributary channel of the Mississippi River (based on isomeric distinctive cluster percentage reported from IM Q-TOF LC/MS by Lu et.al)[49]. Overall, the spike in isomeric complexity (over 3000 new isomers) observed at the downstream-most estuarine site (station HR-5) was highly influenced by CHON and CHOS compound classes (Table 5.1). These results suggest that

the spatial variability observed in DOM isomeric complexity along the Harney River may be highly influenced by the input from mangroves and potentially seagrass. Reports from Schmidt et.al[77] and Powers et.al[78] have suggested that sulfurization processes may contribute to the characteristic compositional signature of porewater DOM in coastal areas. Interestingly, lignin-like components hold the highest isomeric complexity across the samples with an average of 10-12 isomers per molecular formula (See Figure 5.2D).

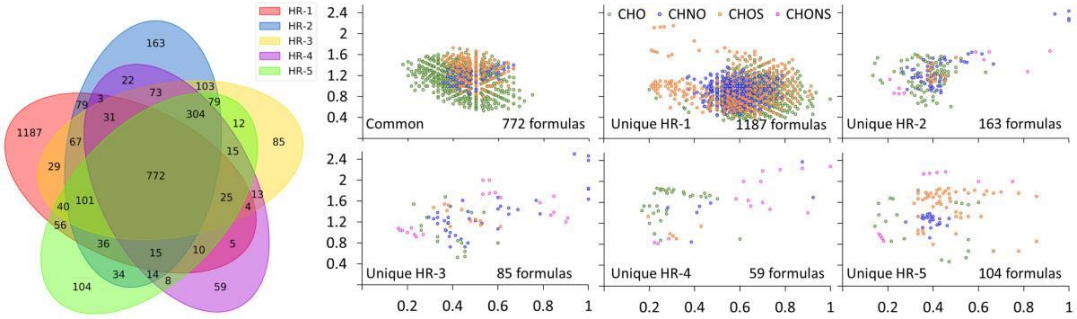


Figure 5. 3.Venn diagram indicating the common and unique molecular species identified by ESI-FT-ICR MS across the salinity transect (left). The van Krevelen plots of common and site-specific DOM molecular species (right).

The Venn diagram depicted in Figure 5.3 shows that near 800 molecular features were shared by all DOM transect samples suggesting that a group of terrestrial-derived components persisted unaltered along the Harney River. On the other hand, molecular formulas were also found to be unique to each section of the transect. While the total number of molecular formulas decreases along the salinity transect, a slight increase was found at the marine endmember compared to the sampling points at intermediate salinity. This result agrees with previous findings on the contribution of marine-derived components to the pool of DOM molecules in the Harney River.

Inspection of the DOM composition based on unique and common molecular formulas across the transect as a function of DBE (Appendix 5.1) reveals the removal of highly unsaturated components from the freshwater endmember. These results confirmed previous findings of the significant export of saturated refractory lignin-type DOM components in coastal ecosystems[22,66]. A closer view of the van Krevelen plots of molecular formulas shared by all samples (Appendix 5.2A) shows an enrichment in DOM isomeric complexity at the segment HR-1 to HR-2, followed by a pattern that exhibits a less complex DOM until the river estuary. The same analysis extended to heteroatom classes (Appendix 5.2B-D) indicated that CHO species were the main contributor to the surge in isomeric complexity at the sampling point HR-2. Since salinity at the upper estuary (e.g. HR-2) is usually higher during the dry season when freshwater discharge into the estuary is lower, DOM is exposed to higher phosphorus conditions in this 'upside down' estuary[79], and thus to different microbial conditions under less nutrient limitations. This fact might enhance degradation of DOM components that were more refractory under freshwater conditions.

Over 1100 distinctive molecular formulas, with mostly tannins-like characteristics, were identified at the freshwater sample (See van Krevelen plot of Unique HR-1 in Figure 3). Plant-derived material exported to freshwater has been recognized as the main source of the DOM signature observed at the Everglades coastal wetlands[31,32,80]. In comparison to the molecular signature of unique components detected at HR-1, the mid-salinity sampling points in the mangrove estuary revealed a major removal of tannins-like species derived from the freshwater endmember. Reports from incubation experiments have attributed the sink of such aromatic tannins-like species to several biogeochemical processes including

biodegradation[81] and photo mineralization[73,74,82]. Nevertheless, other potential mechanisms such as co-precipitation[83], flocculation[84], sorption[85], dilution[86], and non-conservative mixing are very likely to also influence the observed compositional changes.

Unique CHNO and CHOS molecular species were also detected at the freshwater end member (see Figure 5.3). Previous works have suggested that sulphate contributions from different sources (agriculture, sea spray, etc.) may contribute to the CHOS signature of DOM at the Everglades wetlands[31,32].

Few unique DOM molecular species were found across samples HR-2 to HR-5 compared to HR-1, indicating that the compositional fingerprint of DOM exported to the ocean is highly dominated by the molecular signature of marsh derived DOM from the freshwater endmember. However, a minor input of unique molecular compounds observed downstream is a clear indication of the contribution of the Everglades mangrove forest and seagrass to the estuary DOM. Although this is the first report on the molecular level characterization of DOM along the Harney River, previous studies of carbon dynamics in the same area utilizing optical properties[26,29] have suggested up to 20% DOC contributions from the fringe mangrove system. In agreement with our results, a comparative study of DOM dynamics in three coastal wetland systems across the world, using optical properties, has found some mangrove-derived DOM material at regions of intermediate salinities in the Harney River that are equivalent to our sampling points HR-2 and HR-3[7].

Interestingly, a set of common compounds to samples HR-2 to HR-5, mainly CHO and CHNOS, was found, indicating an export of lignin-like refractory DOM from the mangrove forests and/or the formation of degradation products from the freshwater endmember along the salinity gradient (Appendix 5.3). In a recent study[87], sequential photo degradation and

photo generation of CHNO and CHOS compounds were observed during a light exposed time series of freshwater DOM from the Everglades. The exact source and environmental dynamics of CHNOS DOM compounds could not be clearly determined in this study. Nevertheless, the distribution of the two CHNOS compound clusters in the compositional space (Appendix 5.3) indicates that the reduced (O/C < 0.4) and relatively unsaturated species and the oxidized (O/C > 0.8) aliphatic material might be possibly associated with sulfur containing protein degradation products[3,9,20,66] given the high diversity in reaction pathways of DOM N-bearing components in the estuary-marine interface[88].

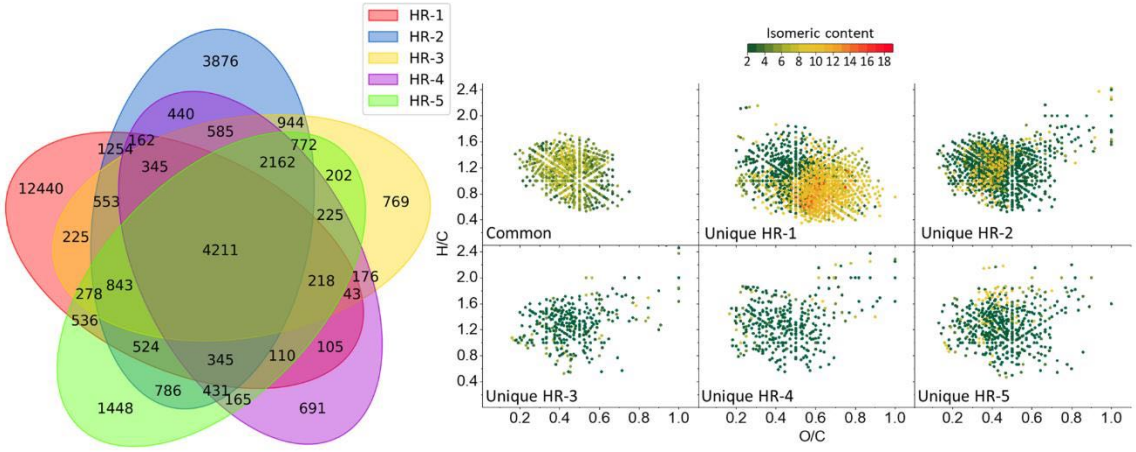


Figure 5. 4.Venn diagram indicating the common and unique isomeric species filtered by their $^{TIMS}CCS_{N2}$ values determined from ESI-TIMS-FT-ICR MS with a threshold error of <8% (left panel). The van Krevelen plots of common and unique isomers with the number of isomers plotted as a third dimension (right panel).

The ESI-TIMS-FT-ICR MS analysis provided ion mobility information at the chemical formulae level. Each isomer can be characterized by its $^{TIMS}CCS_{N2}$, which is an indication of the size and shape of the chemical structure. For example, a long aliphatic chain structure exhibits larger $^{TIMS}CCS_{N2}$ values than a more compact isomeric structure. The addition of the isomeric dimension, based on $^{TIMS}CCS_{N2}$, to each chemical formula

reveals a novel layer of structural information which is invisible in the mass domain. The chemical formulas assigned to each DOM sample were tagged with their corresponding isomeric information ($^{TIMS}CCS_{N2}$ values). A further comparison of concatenated chemical formulas with their $^{TIMS}CCS_{N2}$ values was conducted across DOM samples for the identification of characteristic and common isomeric species along the transect. Over 4,000 structural isomers linked to 772 molecular formulas shared by the five DOM samples were transported from the freshwater end member to the sea (see Venn diagram in Figure 5.4). In another study, a structural analysis based on CID fragmentation of selected DOM precursor molecules in the geographically much more extensive Delaware estuary found no significant differences among DOM structural features of molecular formulas common to all samples collected along a salinity gradient[3]. The authors suggest that these consistent DOM features along the salinity gradient could be the result of a high degree of recalcitrance of the terrestrial endmember DOM, resulting in negligible biogeochemical degradation, or that while some of the DOM components are actually removed during transport, similar compounds are produced at the same time, as previously suggested by Sleighter and Hatcher[66]. In contrast, while our results suggested a major change in the DOM isomeric pool along the salinity transect (only ~20% of the total number of isomers detected at the freshwater end member were presumably transported downstream), an analogous reasoning as the one utilized by Osterholz et.al[75] in the molecular formula domain could be extended here at the isomeric level. Thus, we hypothesize that while dilution and degradation processes may lead to a potential depletion of some isomeric species during their transport from land to sea, the input at higher salinities of isomers derived from fringe mangrove and marine end member sources share close structural

similarities to the terrestrial derived compounds found upstream (i.e., $^{TIMS}CCS_{N2}$ values within 8% sÅ^2). Such process could function as an apparent "isomeric replacement". This information, obtained for the first time at a complex aquatic ecosystem, provides new insights into the isomeric nature of DOM components. Yet, further studies are needed to clearly understand the role of input sources on shaping the isomeric signature of DOM and the potential structural features that contribute to the long-term stability of refractory DOM components.

The number of isomeric compounds unique to each DOM sample along the estuarine transect, filtered based on both chemical formula and $^{TIMS}CCS_{N2}$ values, showed a declining profile in the order HR-1>HR-2>HR-5>HR-3>HR-4. This trend resembles the one obtained for the site-specific chemical formulas (see Venn diagram of Figure 5.3). However, some slight differences can be highlighted here. First, a highly diverse isomeric content attributed to the sample HR-1 (> 12,000 unique structural isomers detected) clearly shows that degradation along the transect may have a major impact on DOM isomeric diversity. Second, the increased isomeric complexity found at HR-5 compared to the nearest upstream sites (HR-3 and HR-4), suggests a preferential input of isomeric components presumably from primary production at the marine end member.

The van Krevelen plots of common DOM isomers found across the transect (Figure 4) suggests that the main contribution to the isomeric diversity of DOM exported from the Everglades to the coastal system is preferentially derived from lignin-like species originated at the freshwater endmember (average of 6 isomers per chemical formula). Noteworthy, two isomeric diverse clusters can be distinguished from the van Krevelen diagram at sample points HR-1 and HR-5. In the first case, highly structurally complex

tannin-like molecules observed at HR-1 support our previous findings that the structural diversity of these particular components is substantially reduced though downstream estuarine biogeochemical processes. On the other hand, the protein-like signature detected at HR-5 with an average of 6-8 isomers per chemical formula (yellow dots in the van Krevelen plot) clearly suggests DOM contributions of more isomeric complex mangrove and/or marine derived DOM. The unique and common isomeric species identified by ESI-TIMS-FT-ICR MS across the samples in this study could be further explored as potential chemo-markers to fingerprinting natural sources of DOM in this coastal ecosystem.

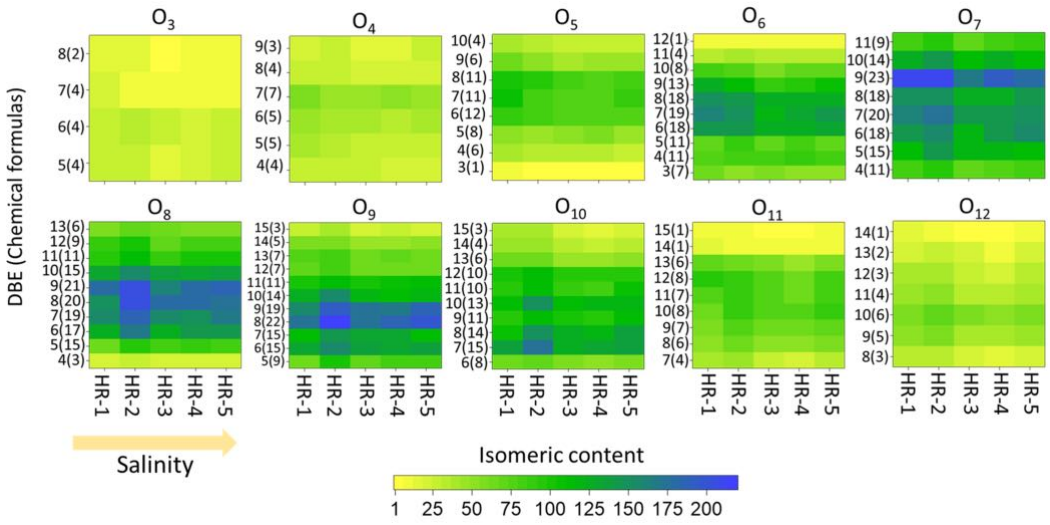


Figure 5. 5. Contour plots showing the distribution of the isomeric diversity from chemical formulas common to all sites as a function of DBE, oxygen class and salinity. Note that each box in the contour plot included the sum of the estimated number of isomers for all the chemical formulas sharing the same DBE (number in parenthesis).

Heatmaps highlighting the total isomeric content of the chemical formulas (indicated in parenthesis) common to all sites as a function of DBE, oxygen class and salinity are described in Figure 5.5. The analysis of the isomeric content along the transect (horizontal trend) as a function of the oxygen class revealed a decreasing profile in the isomeric diversity with increasing salinity for relatively poor oxygenated isomers (up to 6 oxygens).

On the contrary, more oxygenated species exhibited a varying isomeric pattern along the transect with and initial apparent isomeric enrichments at the second lowest salinity site HR-2, followed by a decreasing isomeric enrichment until a second enrichment at the highest salinity site HR-5 (see O_7-O_{10} profiles). The observed relative increase in the isomeric content observed at HR-2 and HR-5 could be associated with the input of isomers resulting from early degradation processes occurring in the upper estuary, when phosphorus limitations are first reduced, followed by isomeric inputs derived from mangrove and marine derived DOM observed at HR-5. A closer view at the Figure 5.5 shows that the isomeric diversity of more unsaturated compounds ($7 \leq DBE \leq 10$) changed with increasing salinity regardless of the chemical formula diversity and the oxygen class. In contrast, the more aliphatic structural isomers ($DBE < 6$), showed a higher resistance to molecular transformations. Note that since we are considering the sum of estimated isomers for all chemical formulas of the same O-class and DBE, other trends at the chemical formula level may not be uncovered.

The inspection of the isomeric information at the level of DOM chemical formulas can illustrate previous observed trends across the salinity transect of the Harney River (Appendix 5.4, see example for $C_{14}H_{10}O_7$ and $C_{15}H_{14}O_6$). In the examples shown, the number of isomers across the salinity transect correlates with the transformation trends described before (see Figure 5.5). That is, the higher oxygenated and higher DBE $C_{14}H_{10}O_7$ strongly varies in isomeric diversity (decreased from 12 to 5 minimum isomers along the transient), suggesting that higher oxidized and more unsaturated isomeric species are more prone to biogeochemical transformations. In contrast, other molecular species with lower oxygenation and DBE $C_{15}H_{14}O_6$ showed no variation in the number of potential isomers

along the transient. Although our results seem to contrast with the degradation patterns as a function of salinity observed by Osterholz et.al[3] for less oxygenated polyphenol chemical formulas ($0.5 < AImod \leq 0.66$), critical differences in both approaches need to be highlighted. While Osterholz's data involved normalized abundances of chemical formulas assigned from FT-ICR MS signals, our approach rests on estimated number of isomers obtained by ion mobility spectrometry. Nevertheless, changes in the relative abundance across the isomers per chemical formula observed along the transect best reflect the isomeric chemical diversity with salinity. While only a small number of isomers are accounted for in this analysis, other isomeric trends could be present and not accessible based on differences in mobility per chemical formula. Further, more in depth studies in this field are warranted to uncover additional information that isomeric data can provide to better understand biogeochemical cycling of DOM.

5.5 Conclusions

A characterization of DOM isomeric complexity using ESI-TIMS-FT ICR MS was conducted for the first time along a freshwater marsh fed mangrove estuary salinity transect (the Harney River in the Florida Everglades). Overall, while the number of detected DOM molecular formulas decreased with increasing salinity, the compositional signature of DOM varied from more unsaturated and oxygenated to more aliphatic and heteroatom diverse components along the transect. While the estimated number of isomeric DOM compounds consistently declined downstream, a relatively small set of terrestrial-derived DOM components remained structurally unchanged (isomers with $\Delta^{TIMS}CCS_{N2}$ values < 8%). Our results provide novel evidence of the increase in refractory character for DOM isomeric species as both DBE and oxygenation decreases, contrasting with unchanged

isomeric patterns detected along a similar freshwater-to-estuarine system[3]. A detailed examination of unique molecular components to each sample evidenced that DOM chemical fingerprint in this estuary is highly influenced by multiple input sources from different origins including the freshwater marsh, mangrove, and primary producers. Nevertheless, several biogeochemical degradation and transformation processes are also likely involved in the modulation of DOM isomeric molecular fingerprint along the transect. Though previous works using traditional and optical techniques (i.e., DOC and fluorescence) have provided valuable information of bulk DOM change along the Harney River freshwater-to-marine system, our results constitute the first view of DOM variability at the isomeric level. The analysis of ion mobility data from molecular formulas unique to each sample indicates that terrestrial, freshwater marsh-derived material is up to four times more isomeric complex than marine primary producers and mangrove DOM. The inspection of the isomeric information at the level of DOM chemical formulas revealed that higher oxidized and more unsaturated isomeric species are more prone to biogeochemical transformations. Moreover, the unique molecular features identified across all the samples in this study, along with their specific isomeric characteristics may be used as potential chemo-markers that can help to better elucidate the linkage between DOM molecular fingerprint and input sources and/or related biogeochemical processing during transport. Our findings on tracking the changes in chemical and isomeric signatures of DOM along coastal gradients illustrate that ESI-TIMS-FT-ICR MS is a useful tool that will help to expand the current knowledge of the carbon fate in coastal aquatic ecosystems.

5.6 References

1. Asmala, E.; Massicotte, P.; Carstensen, J. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnology and Oceanography* **2021,** *66* (4), 1381-1393.

2. Buffam, I.; Turner, M. G.; Desai, A. R.; Hanson, P. C.; Rusak, J. A.; Lottig, N. R.; Stanley, E. H.; Carpenter, S. R. Integrating aquatic and terrestrial components to construct a complete carbon budget for a north temperate lake district. *Global Change Biology* **2011,** *17* (2), 1193-1211.

3. Osterholz, H.; Kirchman, D. L.; Niggemann, J.; Dittmar, T. Environmental Drivers of Dissolved Organic Matter Molecular Composition in the Delaware Estuary. *Frontiers in Earth Science* **2016,** *4* (95), 1-14.

4. Wang, C.; Li, Y.; Li, Y.; Zhou, H.; Stubbins, A.; Dahlgren, R. A.; Wang, Z.; Guo, W. Dissolved Organic Matter Dynamics in the Epipelagic Northwest Pacific Low-Latitude Western Boundary Current System: Insights From Optical Analyses. *Journal of Geophysical Research: Oceans* **2021,** *126* (9), 1-18.

5. Berg, S. M.; Peterson, B. D.; McMahon, K. D.; Remucal, C. K. Spatial and Temporal Variability of Dissolved Organic Matter Molecular Composition in a Stratified Eutrophic Lake. *Journal of Geophysical Research: Biogeosciences* **2022,** *127* (1), 1-16.

6. Zhang, X.; Cao, F.; Huang, Y.; Tang, J. Variability of dissolved organic matter in two coastal wetlands along the Changjiang River Estuary: Responses to tidal cycles, seasons, and degradation processes. *Science of The Total Environment* **2022,** *807*, 1-14.

7. Maie, N.; Sekiguchi, S.; Watanabe, A.; Tsutsuki, K.; Yamashita, Y.; Melling, L.; Cawley, K. M.; Shima, E.; Jaffé, R. Dissolved organic matter dynamics in the oligo/meso-haline zone of wetland-influenced coastal rivers. *Journal of Sea Research* **2014,** *91*, 58-69.

8. Gonçalves-Araujo, R.; Stedmon, C. A.; Heim, B.; Dubinenkov, I.; Kraberg, A.; Moiseev, D.; Bracher, A. From Fresh to Marine Waters: Characterization and Fate of Dissolved Organic Matter in the Lena River Delta Region, Siberia. *Frontiers in Marine Science* **2015,** *2* (108), 1-13.

9. Seidel, M.; Yager, P. L.; Ward, N. D.; Carpenter, E. J.; Gomes, H. R.; Krusche, A. V.; Richey, J. E.; Dittmar, T.; Medeiros, P. M. Molecular-level changes of dissolved organic matter along the Amazon River-to-ocean continuum. *Marine Chemistry* **2015,** *177*, 218-231.

10. Hosen, J. D.; Armstrong, A. W.; Palmer, M. A. Dissolved organic matter variations in coastal plain wetland watersheds: The integrated role of hydrological connectivity, land use, and seasonality. *Hydrological Processes* **2018,** *32* (11), 1664-1681.

11. Chow, A. T.; Dai, J.; Conner, W. H.; Hitchcock, D. R.; Wang, J.-J. Dissolved organic matter and nutrient dynamics of a coastal freshwater forested wetland in Winyah Bay, South Carolina. *Biogeochemistry* **2013,** *112* (1), 571-587.

12. Fellman, J. B.; Hood, E.; Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnology and Oceanography* **2010,** *55* (6), 2452-2462.

13. Yamashita, Y.; Boyer, J. N.; Jaffé, R. Evaluating the distribution of terrestrial dissolved organic matter in a complex coastal ecosystem using fluorescence spectroscopy. *Continental Shelf Research* **2013,** *66*, 136-144.

14. Hansen, A. M.; Kraus, T. E. C.; Pellerin, B. A.; Fleck, J. A.; Downing, B. D.; Bergamaschi, B. A. Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation. *Limnology and Oceanography* **2016,** *61* (3), 1015-1032.

15. Gonçalves-Araujo, R.; Granskog, M. A.; Bracher, A.; Azetsu-Scott, K.; Dodd, P. A.; Stedmon, C. A. Using fluorescent dissolved organic matter to trace and distinguish the origin of Arctic surface waters. *Scientific Reports* **2016,** *6* (1), 33978.

16. Clark, M. G.; Biagi, K. M.; Carey, S. K. Optical properties of dissolved organic matter highlight peatland-like properties in a constructed wetland. *Science of The Total Environment* **2022,** *802*, 1-13.

17. Zhang, Y.; Zhou, Y.; Shi, K.; Qin, B.; Yao, X.; Zhang, Y. Optical properties and composition changes in chromophoric dissolved organic matter along trophic gradients: Implications for monitoring and assessing lake eutrophication. *Water Research* **2018,** *131*, 255-263.

18. Kellerman, A. M.; Kothawala, D. N.; Dittmar, T.; Tranvik, L. J. Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nature Geoscience* **2015,** *8* (6), 454-457.

19. Roebuck, J. A.; Seidel, M.; Dittmar, T.; Jaffé, R. Land Use Controls on the Spatial Variability of Dissolved Black Carbon in a Subtropical Watershed. *Environmental Science & Technology* **2018,** *52* (15), 8104-8114.

20. Seidel, M.; Beck, M.; Riedel, T.; Waska, H.; Suryaputra, I. G. N. A.; Schnetger, B.; Niggemann, J.; Simon, M.; Dittmar, T. Biogeochemistry of dissolved organic matter in an anoxic intertidal creek bank. *Geochimica et Cosmochimica Acta* **2014,** *140*, 418-434.

21. Osterholz, H.; Niggemann, J.; Giebel, H.-A.; Simon, M.; Dittmar, T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nature Communications* **2015,** *6* (1), 1-8.

22. Koch, B. P.; Witt, M.; Engbrodt, R.; Dittmar, T.; Kattner, G. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta* **2005,** *69* (13), 3299-3308.

23. Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **2013,** *10* (3), 1583-1624.

24. Kominoski, J. S.; Gaiser, E. E.; Castañeda-Moya, E.; Davis, S. E.; Dessu, S. B.; Julian II, P.; Lee, D. Y.; Marazzi, L.; Rivera-Monroy, V. H.; Sola, A.; Stingl, U.; Stumpf, S.; Surratt, D.; Travieso, R.; Troxler, T. G. Disturbance legacies increase and synchronize nutrient concentrations and bacterial productivity in coastal ecosystems. *Ecology* **2020,** *101* (5), 1-13.

25. Regier, P.; Larsen, L.; Cawley, K.; Jaffé, R. Linking Hydrology and Dissolved Organic Matter Characteristics in a Subtropical Wetland: A Long-Term Study of the Florida Everglades. *Global Biogeochemical Cycles* **2020,** *34* (12), 1-20.

26. Cawley, K. M.; Yamashita, Y.; Maie, N.; Jaffé, R. Using Optical Properties to Quantify Fringe Mangrove Inputs to the Dissolved Organic Matter (DOM) Pool in a Subtropical Estuary. *Estuaries and Coasts* **2014,** *37* (2), 399-410.

27. Yamashita, Y.; Scinto, L. J.; Maie, N.; Jaffé, R. Dissolved Organic Matter Characteristics Across a Subtropical Wetland's Landscape: Application of Optical Properties in the Assessment of Environmental Dynamics. *Ecosystems* **2010,** *13* (7), 1006-1019.

28. Chen, M.; Maie, N.; Parish, K.; Jaffé, R. Spatial and temporal variability of dissolved organic matter quantity and composition in an oligotrophic subtropical coastal wetland. *Biogeochemistry* **2013,** *115* (1), 167-183.

29. Regier, P.; Jaffé, R. Short-Term Dissolved Organic Carbon Dynamics Reflect Tidal, Water Management, and Precipitation Patterns in a Subtropical Estuary. *Frontiers in Marine Science* **2016,** *3*, 1-14.

30. Ya, C.; Anderson, W.; Jaffé, R. Assessing dissolved organic matter dynamics and source strengths in a subtropical estuary: Application of stable carbon isotopes and optical properties. *Continental Shelf Research* **2015,** *92*, 98-107.

31. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016,** *13* (8), 2257-2277.

32. Wagner, S.; Jaffé, R.; Cawley, K.; Dittmar, T.; Stubbins, A. Associations Between the Molecular and Optical Properties of Dissolved Organic Matter in the Florida Everglades, a Model Coastal Wetland System. *Frontiers in Chemistry* **2015,** *3* (66), 1-14.

33. Hawkes, J. A.; Dittmar, T.; Patriarca, C.; Tranvik, L.; Bergquist, J. Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Analytical Chemistry* **2016,** *88* (15), 7698-7704.

34. Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science* **2017,** *4* (405), 1-14.

35. Patriarca, C.; Bergquist, J.; Sjöberg, P. J. R.; Tranvik, L.; Hawkes, J. A. Online HPLC-ESI-HRMS Method for the Analysis and Comparison of Different Dissolved Organic Matter Samples. *Environmental Science & Technology* **2018,** *52* (4), 2091-2099.

36. Spranger, T.; Pinxteren, D. v.; Reemtsma, T.; Lechtenfeld, O. J.; Herrmann, H. 2D Liquid Chromatographic Fractionation with Ultra-high Resolution MS Analysis Resolves a Vast Molecular Diversity of Tropospheric Particle Organics. *Environmental Science & Technology* **2019,** *53* (19), 11353-11363.

37. Venter, P.; Muller, M.; Vestner, J.; Stander, M. A.; Tredoux, A. G. J.; Pasch, H.; de Villiers, A. Comprehensive Three-Dimensional LC × LC × Ion Mobility Spectrometry Separation Combined with High-Resolution MS for the Analysis of Complex Samples. *Analytical Chemistry* **2018,** *90* (19), 11643-11650.

38. Tose, L. V.; Benigni, P.; Leyva, D.; Sundberg, A.; Ramírez, C. E.; Ridgeway, M. E.; Park, M. A.; Romão, W.; Jaffé, R.; Fernandez-Lima, F. Coupling trapped ion mobility spectrometry to mass spectrometry: trapped ion mobility spectrometry–time-of-flight mass spectrometry versus trapped ion mobility spectrometry–Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* **2018,** *32* (15), 1287-1295.

39. Leyva, D.; Tariq, M. U.; Jaffé, R.; Saeed, F.; Lima, F. F. Unsupervised Structural Classification of Dissolved Organic Matter Based on Fragmentation Pathways. *Environmental Science & Technology* **2022,** *56* (2), 1458-1468.

40. Robinson, E. W.; Garcia, D. E.; Leib, R. D.; Williams, E. R. Enhanced Mixture Analysis of Poly(ethylene glycol) Using High-Field Asymmetric Waveform Ion

Mobility Spectrometry Combined with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2006,** *78* (7), 2190-2198.

41. Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion Mobility–Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angewandte Chemie International Edition* **2007,** *46* (42), 8001-8004.

42. Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. Petroleum Crude Oil Characterization by IMS-MS and FTICR MS. *Anal. Chem.* **2009,** *81* (24), 9941-9947.

43. Schrader, W.; Xuan, Y.; Gaspar, A. Studying ultra-complex crude oil mixtures by using High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) coupled to an ESI-LTQ-Orbitrap Mass Spectrometer. *Eur. J. Mass Spectrom.* **2014,** *20* (1), 43-49.

44. Benigni, P.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Targeted High-Resolution Ion Mobility Separation Coupled to Ultrahigh-Resolution Mass Spectrometry of Endocrine Disruptors in Complex Mixtures. *Analytical Chemistry* **2015,** *87* (8), 4321-4325.

45. Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the structural complexity of dissolved organic matter: isomeric diversity. *Faraday Discussions* **2019,** *218* (0), 431-440.

46. Leyva, D.; Jaffe, R.; Fernandez-Lima, F. Structural Characterization of Dissolved Organic Matter at the Chemical Formula Level Using TIMS-FT-ICR MS/MS. *Analytical Chemistry* **2020,** *92* (17), 11960-11966.

47. Olanrewaju, C. A.; Ramirez, C. E.; Fernandez-Lima, F. Characterization of Deasphalted Crude Oils Using Gas Chromatography–Atmospheric Pressure Laser Ionization–Trapped Ion Mobility Spectrometry–Time-of-Flight Mass Spectrometry. *Energy & Fuels* **2021,** *35* (17), 13722-13730.

48. Lu, K.; Gardner, W. S.; Liu, Z. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). *Environmental Science & Technology* **2018,** *52* (13), 7182-7191.

49. Lu, K.; Li, X.; Chen, H.; Liu, Z. Constraints on isomers of dissolved organic matter in aquatic environments: Insights from ion mobility mass spectrometry. *Geochimica et Cosmochimica Acta* **2021,** *308*, 353-372.

50. Law, K. P.; He, W.; Tao, J.; Zhang, C. Characterization of the Exometabolome of Nitrosopumilus maritimus SCM1 by Liquid Chromatography–Ion Mobility Mass Spectrometry. *Frontiers in Microbiology* **2021,** *12*, 1-20.

51. Gao, Y.; Wang, W.; He, C.; Fang, Z.; Zhang, Y.; Shi, Q. Fractionation and molecular characterization of natural organic matter (NOM) by solid-phase extraction followed by FT-ICR MS and ion mobility MS. *Analytical and Bioanalytical Chemistry* **2019**, 1-10.

52. Morrison, E. S.; Shields, M. R.; Bianchi, T. S.; Liu, Y.; Newman, S.; Tolic, N.; Chu, R. K. Multiple biomarkers highlight the importance of water column processes in treatment wetland organic matter cycling. *Water Research* **2020,** *168*, 115153.

53. He, D.; Li, P.; He, C.; Wang, Y.; Shi, Q. Eutrophication and watershed characteristics shape changes in dissolved organic matter chemistry along two river-estuarine transects. *Water Research* **2022,** *214*, 118196.

54. Benigni, P.; Fernandez-Lima, F. Oversampling Selective Accumulation Trapped Ion Mobility Spectrometry Coupled to FT-ICR MS: Fundamentals and Applications. *Analytical Chemistry* **2016,** *88* (14), 7404-7412.

55. Ridgeway, M. E.; Wolff, J. J.; Silveira, J. A.; Lin, C.; Costello, C. E.; Park, M. A. Gated trapped ion mobility spectrometry coupled to fourier transform ion cyclotron resonance mass spectrometry. *International Journal for Ion Mobility Spectrometry* **2016,** *19* (2), 77-85.

56. Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A. Gas-phase separation using a trapped ion mobility spectrometer. *International Journal for Ion Mobility Spectrometry* **2011,** *14* (2), 93-98.

57. Hernandez, D. R.; DeBord, J. D.; Ridgeway, M. E.; Kaplan, D. A.; Park, M. A.; Fernandez-Lima, F. Ion dynamics in a trapped ion mobility spectrometer. *Analyst* **2014,** *139* (8), 1913-1921.

58. Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Review of Scientific Instruments* **2011,** *82* (12), 126106.

59. Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* **2008,** *6* (6), 230-235.

60. Benigni, P.; Marin, R.; Sandoval, K.; Gardinali, P.; Fernandez-Lima, F. Chemical Analysis of Water-accommodated Fractions of Crude Oil Spills Using TIMS-FT-ICR MS. . *J. Vis. Exp.* **2017,** *121*, e55352.

61. Benigni, P.; Porter, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Increasing Analytical Separation and Duty Cycle with Nonlinear Analytical Mobility Scan Functions in TIMS-FT-ICR MS. *Analytical Chemistry* **2018,** *90* (4), 2446-2450.

62. McDaniel, E. W.; Mason, E. A. *Mobility and diffusion of ions in gases*; John Wiley and Sons, Inc: New York, 1973.

63. Herzsprung, P.; Hertkorn, N.; von Tümpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Analytical and Bioanalytical Chemistry* **2014,** *406* (30), 7977-7987.

64. Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017,** *89* (17), 9048-9055.

65. Koch, B. P.; Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry* **2006,** *20* (5), 926-932.

66. Sleighter, R. L.; Hatcher, P. G. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry* **2008,** *110* (3), 140-152.

67. Creed, I. F.; McKnight, D. M.; Pellerin, B. A.; Green, M. B.; Bergamaschi, B. A.; Aiken, G. R.; Burns, D. A.; Findlay, S. E. G.; Shanley, J. B.; Striegl, R. G.; Aulenbach, B. T.; Clow, D. W.; Laudon, H.; McGlynn, B. L.; McGuire, K. J.; Smith, R. A.; Stackpoole, S. M. The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum. *Canadian Journal of Fisheries and Aquatic Sciences* **2015,** *72* (8), 1272-1285.

68. Vannote, R. L.; Minshall, G. W.; Cummins, K. W.; Sedell, J. R.; Cushing, C. E. The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences* **1980,** *37* (1), 130-137.

69. Kujawinski, E. B.; Del Vecchio, R.; Blough, N. V.; Klein, G. C.; Marshall, A. G. Probing molecular-level transformations of dissolved organic matter: insights on photochemical degradation and protozoan modification of DOM from electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry* **2004,** *92* (1), 23-37.

70. Minor, E. C.; Dalzell, B. J.; Stubbins, A.; Mopper, K. Evaluating the photoalteration of estuarine dissolved organic matter using direct temperature-resolved mass spectrometry and UV-visible spectroscopy. *Aquatic Sciences* **2007,** *69* (4), 440-455.

71. Dalzell, B. J.; Minor, E. C.; Mopper, K. M. Photodegradation of estuarine dissolved organic matter: a multi-method assessment of DOM transformation. *Organic Geochemistry* **2009,** *40* (2), 243-257.

72. Wilske, C.; Herzsprung, P.; Lechtenfeld, O. J.; Kamjunke, N.; von Tümpling, W. Photochemically Induced Changes of Dissolved Organic Matter in a Humic-Rich and Forested Stream. *Water* **2020,** *12* (2), 1-21.

73. Gonsior, M.; Peake, B. M.; Cooper, W. T.; Podgorski, D.; D'Andrilli, J.; Cooper, W. J. Photochemically Induced Changes in Dissolved Organic Matter Identified by Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Environmental Science & Technology* **2009,** *43* (3), 698-703.

74. Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnology and Oceanography* **2010,** *55* (4), 1467-1477.

75. Osterholz, H.; Kirchman, D. L.; Niggemann, J.; Dittmar, T. Environmental Drivers of Dissolved Organic Matter Molecular Composition in the Delaware Estuary. *Frontiers in Earth Science* **2016,** *4*, 1-14.

76. Jaffé, R.; Mead, R.; Hernandez, M. E.; Peralba, M. C.; DiGuida, O. A. Origin and transport of sedimentary organic matter in two subtropical estuaries: a comparative, biomarker-based study. *Organic Geochemistry* **2001,** *32* (4), 507-526.

77. Schmidt, F.; Elvert, M.; Koch, B. P.; Witt, M.; Hinrichs, K.-U. Molecular characterization of dissolved organic matter in pore water of continental shelf sediments. *Geochimica et Cosmochimica Acta* **2009,** *73* (11), 3337-3358.

78. Powers, L. C.; Lapham, L. L.; Malkin, S. Y.; Heyes, A.; Schmitt-Kopplin, P.; Gonsior, M. Molecular and optical characterization reveals the preservation and sulfurization of chemically diverse porewater dissolved organic matter in oligohaline and brackish Chesapeake Bay sediments. *Organic Geochemistry* **2021,** *161*, 1-15.

79. Childers, D. L.; Boyer, J. N.; Davis, S. E.; Madden, C. J.; Rudnick, D. T.; Sklar, F. H. Relating precipitation and water management to nutrient concentrations in the oligotrophic "upside-down" estuaries of the Florida Everglades. *Limnology and Oceanography* **2006,** *51* (1part2), 602-616.

80. Maie, N.; Yang, C.; Miyoshi, T.; Parish, K.; Jaffé, R. Chemical characteristics of dissolved organic matter in an oligotrophic subtropical wetland/estuarine. *Limnology and Oceanography* **2005,** *50* (1), 23-35.

81. Ward, N. D.; Keil, R. G.; Medeiros, P. M.; Brito, D. C.; Cunha, A. C.; Dittmar, T.; Yager, P. L.; Krusche, A. V.; Richey, J. E. Degradation of terrestrially derived macromolecules in the Amazon River. *Nature Geoscience* **2013,** *6* (7), 530-533.

82. Ward, C. P.; Cory, R. M. Complete and Partial Photo-oxidation of Dissolved Organic Matter Draining Permafrost Soils. *Environmental Science & Technology* **2016,** *50* (7), 3545-3553.

83. Maie, N.; Pisani, O.; JaffÉ, R. Mangrove tannins in aquatic ecosystems: Their fate and possible influence on dissolved organic carbon and nitrogen cycling. *Limnology and Oceanography* **2008,** *53* (1), 160-171.

84. Asmala, E.; Kaartokallio, H.; Carstensen, J.; Thomas, D. N. Variation in Riverine Inputs Affect Dissolved Organic Matter Characteristics throughout the Estuarine Gradient. *Frontiers in Marine Science* **2016,** *2*, 1-15.

85. Groeneveld, M.; Catalán, N.; Attermeyer, K.; Hawkes, J.; Einarsdóttir, K.; Kothawala, D.; Bergquist, J.; Tranvik, L. Selective Adsorption of Terrestrial Dissolved Organic Matter to Inorganic Surfaces Along a Boreal Inland Water Continuum. *Journal of Geophysical Research: Biogeosciences* **2020,** *125* (3), 1-22.

86. Medeiros, P. M.; Seidel, M.; Ward, N. D.; Carpenter, E. J.; Gomes, H. R.; Niggemann, J.; Krusche, A. V.; Richey, J. E.; Yager, P. L.; Dittmar, T. Fate of the Amazon River dissolved organic matter in the tropical Atlantic Ocean. *Global Biogeochemical Cycles* **2015,** *29* (5), 677-690.

87. Harir, M.; Cawley, K. M.; Hertkorn, N.; Schmitt-Kopplin, P.; Jaffé, R. Molecular and spectroscopic changes of peat-derived organic matter following photo-exposure: Effects on heteroatom composition of DOM. *Science of The Total Environment* **2022,** *838*, 155790.

88. Schmidt, F.; Koch, B. P.; Elvert, M.; Schmidt, G.; Witt, M.; Hinrichs, K.-U. Diagenetic Transformation of Dissolved Organic Nitrogen Compounds under Contrasting Sedimentary Redox Conditions in the Black Sea. *Environmental Science & Technology* **2011,** *45* (12), 5223-5229.

CHAPTER VI

VI. DESCRIPTION OF DOM TRANSFORMATIONAL NETWORKS AT THE MOLECULAR LEVEL

6.1 Abstract

Dissolved Organic Matter (DOM) is a building block of the global carbon cycle. Unscrambling the structural footprint of DOM is key to understand its biogeochemical transformations at the mechanistic level. Although numerous studies have improved our knowledge of DOM chemical makeup, its three-dimensional picture remains unrevealed. In this work we compare four SPE-DOM samples (Suwanee River fulvic acid standard, Pantanal, HR-1 and HR-5) from three different ecosystems using high resolution mobility and ultra-high-resolution FT ICR MS/MS. Structural families were identified based on neutral losses at the level of nominal mass using CASI-CID FT ICR MS/MS. Comparison of the structural families indicated dissimilarities in the structural footprint (SRFA>Pantanal>HR-5>HR-1). The Cytoscape structural family representation revealed characteristic clustering patterns among the DOM samples, thus confirming their differences at the structural level (only 10% is common across the four samples). The analysis at the level of neutral loss-based functionalities suggest that hydration and carboxylation are ubiquitous transformational processes across the three ecosystems. In contrast, mechanisms involving methoxy moieties were found unique to families of SRFA and Pantanal samples. The inclusion of the isomeric content (mobility measurements at the level of chemical formula) to the structural family description suggest that additional pathways are possible and account for the dissimilarities observed. This is first molecular level characterization of DOM samples capable of providing meaningful global and chemical formula based structural information leading to the identification of unique/common transformational processes in DOM.

6.2 Introduction

Dissolved Organic Matter (DOM) is an extremely complex mixture of organic species with a wide variety of chemical signatures and is considered a fundamental piece of the biogeochemical cycle in our planet[1-6]. Due to its crucial role in the aquatic ecosystems (i.e., primary source, light attenuation, complexation, etc.), the characterization of dissolved organic matter (DOM) across different environments has been the focus of numerous studies[7-16] and can be classified in i) bulk- and ii) molecular-level characterization [10,17-20]. The molecular level characterization is gaining primary attention because the increasing need to better understand how changes in DOM makeup across environments affect critical ecological processes (e.g., the global carbon cycle)[4,21]. The introduction of ultra-high, resolution mass spectrometry (e.g., Fourier transform ion cyclotron mass spectrometry, FT-ICR MS) has significantly advance our understanding of the DOM molecular makeup[18,22]. Several thousands of chemical components distributed across various heteroatom classes can be now routinely identified from DOM samples in a single broadband FT-ICR MS analysis[23-25].

The structural characterization of DOM remains an open problem. The high chemical diversity and high isomeric complexity has significantly limited the DOM structural characterization [10-12]. Traditional NMR studies are limited by the need of pure and relative high concentrations of purified components[9,24,26,27]. On a parallel approach, chromatographic (e.g., liquid chromatography[28-30]) and gas-phase separations (e.g., ion mobility[12,25,31-33] and ultra-high resolution FTMS[13,18,34-36]) have been explored. Most of the studies attempting to gain DOM structural information have rest on identifying structural

classes by NMR (i.e., aliphatic, aromatic, olefinic, etc.)[26,27,37] and functional groups from tandem mass spectrometry (i.e., carboxylic moiety from CO_2 neutral loss)[38-40].

A fundamental aspect for the structural DOM characterization is the need to separate and identify isomeric components, as well as to identify the transformation pathways[11,30]. In 2020, trapped ion mobility spectrometry (TIMS) coupled to FT-ICR MS[41] allowed for the first time isomeric separation followed by chemical formula level fragmentation of a DOM sample and candidate isomeric structures were proposed. The integration of a score system from *in-silico* MS/MS fragmentation and structural screening based on experimental vs theoretical ion-neutral collisional cross sections (CCS), allowed for a structural assignment with reduced ambiguities. Despite the unique advantages of this methodology, its application to a large-scale study (e.g., thousands of chemical formulas) could be challenging and ultimately impractical for routine characterization. In 2022, a more feasible approach based on the neutral loss patterns of nominal mass isolated precursors using continuous accumulation of selected ions (CASI) followed by collision induced dissociation (CID) and FT ICR MS detection[42] was proposed. Over 1000 structural families of DOM related components from 110 nominal masses were identified in a SPE-DOM sample from Pantanal, Brazil. When combined with a novel visualization approach based on a Cytoscape network (Graph-DOM[42,43]), the structurally interconnected DOM families were visualized, and the potential biogeochemical transformation processes identified.

In the present work, we further interrogate the DOM structural space by identifying the transformational networks common to different DOMs. In particular, we implement for the first time a combined isomeric mobility separation and neutral loss-based MS/MS

fragmentation strategies for the identification of common and unique structural families on common DOMs. This unsupervised structural classification workflow allows for the structural footprint across aquatic DOMs. Examples are shown for the case of SPE-DOM samples from Suwannee River fulvic acid standard (SRFA), a wetland from Brazil (Pantanal), and HR-1 and HR-5 from marsh-estuary transect in the Harney River (Florida Everglades).

6.3 Experimental

6.3.1 Sample preparation

DOM samples from four aquatic environments were studied: Suwanee River Fulvic Acid, (SRFA), a wetland from Brazil (Pantanal), and HR-1 and HR-5 from marsh-estuary transect in the Harney River (Florida Everglades). Suwanee River Fulvic Acid standard was obtained from the International Humic Substances Society (IHSS, http://humic-substances.org). The SRFA stock solution (1 mg/ml) was prepared by dissolving 1 mg of SRFA standard powder in 1 ml of methanol. Pantanal (Pantanal National Park, Brazil), and HR-1 and HR-5 (Harney River, Florida Everglades) were obtained by solid phase extraction (SPE) of surface water. Details of the sampling protocol and the SPE method used are described elsewhere[9,44,45]. Briefly, one liter of water was collected using pre-cleaned plastic bottles and filtered using GFF pre-combusted 0.7 μm glass fiber filters no later than 6 h after collection. For the SPE procedure, samples were acidified (pH 2) and loaded onto a 1g Bond Elut PPL cartridge (Agilent, Santa Clara, CA, USA) conditioned with 1 cartridge volume of methanol followed by 1 cartridge volume of pH 2 Milli-Q water. The loaded PPL cartridge was then rinse with pH 2 Milli-Q water for desalting and dried under a nitrogen gas flow for five minutes, before the elution of DOM components with 20

mL of methanol. SPE-DOM extracts were stored at -20 °C in amber glass vials until further analysis. All SPE-DOM samples were diluted by dissolving 30 uL of methanol extract in 1 mL of denatured ethanol. All solvents (Optima LC-MS grade) were obtained from Fisher Scientific (Pittsburgh, PA).

6.3.2 (-)ESI-FT-ICR-MS

A SolariX 9T ESI-FT-ICR MS spectrometer (Bruker Daltonics, MA) equipped with an infinity ICR cell was optimized for high transmission of ions in the 100-1200 m/z range. Diluted DOM samples were ionized using an electrospray ionization source (Apollo II ESI design, Bruker Daltonics, Inc., MA) in negative ion mode and injected at 200 μL/h. Typical operating conditions were 3700 - 4200 V capillary voltage, 4 L/min dry gas flow rate, 1.0 bar nebulizer gas pressure, and a dry gas temperature 200 °C. Operational parameters were as follows: funnel rf amplitude 160 voltage peak-to-peak (Vpp), capillary exit -150 V, deflector plate -140 V, skimmer1 -20 V, transfer line RF 350 Vpp, octupole RF amplitude 350 Vpp and collision cell RF 1100 Vpp. Agilent ESI-L low concentration Tuning Mix calibration standard (Agilent, Santa Clara, CA, USA) was used during the instrument tuning. Broadband MS_1 spectra (200 co-added scans) of the four DOM samples were collected at 4 MW data acquisition size (mass resolution of ~ 300k at 400 *m/z*).

6.3.3 (-)ESI-FT-ICR CASI CID MS/MS

Comprehensive MS/MS data of SRFA, Pantanal, HR-1, and HR-5 DOM samples was obtained by performing ESI-FT-ICR CASI CID MS /MS (see details in ref EST paper). Briefly, odd mass ions (*m/z* range 261-477), sequentially isolated in the quadrupole at nominal mass, were accumulated for 5-7 s in the collision cell, and further subjected to CID before the analysis in the ICR cell. Tailored CID collision voltages (15 V – 27 V)

were utilized for a better fragment coverage across the mass range. Same ion optics parameters used during broadband ESI-FT-ICR MS experiments were utilized for the MS/MS analysis. Six to eight segments of 60-100 MS/MS scans each were collected across the predefined *m/z* range using the serial run mode acquisition. Tandem MS spectra of the DOM samples were collected at 2 MW data acquisition size (mass resolution of ~ 140k at 400 m/z).

6.3.4 (-)ESI-TIMS-FT-ICR MS

A custom-built Solarix 7T ESI-TIMS-FT-ICR MS spectrometer with an infinity ICR cell (Bruker Daltonics Inc., MA) was utilized to obtain isomeric information for the precursor molecules included in the structural families. The ESI operating conditions were the same as for the ESI-FT-ICR MS analysis described above. Typical transmission operational parameters included funnel rf amplitude 220 peak-to-peak voltage (Vpp), capillary exit -100 V, deflector plate -90 V, skimmer1 -60 V, transfer line rf 350 Vpp, octupole rf amplitude 350 Vpp and collision cell rf 1000 Vpp. Mass profiles of the four DOM samples were collected at 4 MW data acquisition size.

The principle of the separation of ions in the gas phase using TIMS-ESI-FT-ICR MS are described elsewhere[33,46-48]. Briefly, ions in the TIMS cartridge are hold against a nitrogen flow using an electric field applied on the electrodes. In these settings, the drag force exerted by the moving gas is counteracted by the electric field force so that the ions can be spatially separated across the TIMS analyzer axis based on their ion mobility[49-51]. To improve the trapping efficiency of the TIMS cell, ions are radially confined using a quadrupolar RF field. The ion mobility, K_0, of a charged molecule in a TIMS cell can be described by the equation (1):

$$K_0 = \frac{v_g}{E} = \frac{A}{V_{elution} - V_{out}} \qquad (1)$$

where v_g, E, A, $V_{elution}$, and V_{out} are the velocity of the gas, electric field, a calibration constant, elution voltage, and tunnel out voltage, respectively.

Ion–neutral collisional cross sections (CCS (Ω, Å^2), which are an indication of the size and shape of the charged molecules can be calculated from K_0 using the Mason–Schamp equation (2):

$$\Omega = \frac{(18\pi)^{1/2}\, z}{16(k_B T)^{1/2}} \left(\frac{1}{m_i} + \frac{1}{m_b}\right)^{1/2} \frac{1760\, T}{Ko\ P\ 273.15\ N^*} \qquad (2)$$

where z is the charge of the ion, k_B is the Boltzmann constant, N^* is the number density, and m_i and m_b are the masses of the ion and bath gas, respectively. (McDaniel and Mason 1973)

An in-house software written in National Instruments LabVIEW synchronized with the FTMS control acquisition software was used for controlling the TIMS cartridge. The collection of IMS frames was conducted in the oversampling mode previously described by our group[12,33,41,46]. Nitrogen gas at ca 300 K was employed as the bath gas and pressures P1 = 2.4 mbar, P2 = 1.0 mbar, and a 220 Vpp rf were also utilized. The TIMS cell was operated using a simultaneous fill/trap sequence synchronized with the accumulation during detect mode of the FTMS control software. A voltage difference of 4 V across the ΔE gradient and a 0.2 V stepping scan function across the total 100V ΔV range were utilized. A maximum of 1,000 IMS scans were collected per mass spectrum.

6.3.5 Data Processing

Data was processed using Data Analysis (v. 5.2, Bruker Daltonics, CA) and the plots were created using OriginPro 2016 (Originlab Co., MA). The assignment of chemical formulas was conducted using Composer software (version 1.0.6, Sierra Analytics, CA, USA) and confirmed with Data Analysis (version 5.2, Bruker Daltonics). The assignment was validated by the lowest formula errors, the confirmation of isotopologues and the removal of assigned peaks belonging to classes with only a few sparsely scattered members. Theoretical formula constraints of $C_{4\text{-}50}H_{4\text{-}100}N_{0\text{–}3}O_{0\text{–}25}S_{0\text{–}2}$, S/N>3, *m/z* range 100-800, error <1 ppm, $0<O/C\leq1.2$, $0.3\leq H/C\leq2.5$, and DBE-O ≤ 10[52] were considered. An internal walking recalibration using the oxygen homologous series (O_4-O_{20}) was performed in Composer software. An average recalibration error < 200 ppb in the mass range 200-700 Da was obtained for the MS spectra of the four DOM samples. The fragment spectra were internally calibrated using a list of exact masses of fragment ions obtained from regularly occurring neutral losses in DOM and their combinations[10,11].

Input files containing the accurate mass of assigned peaks from MS2 and MS1, the isolated nominal mass, the ion abundance, and the assigned chemical formulas were processed using our developed in-house Python code *Graph-DOM*[42,43]. Briefly, ordered fragmentation pathways considering multiples of CH_4, O, H_2O, CO, CH_2O, CH_4O, and CO_2 neutral losses and with 1 mDa tolerance error were computed using *Graph-DOM*. Families of structurally related precursors were identified using a conceptual model based on *de novo* matching of fragmentation pathways[42,43]. Networks of DOM interconnected precursors belonging to the structural families were created using Cytoscape v.3.82[53]. A comparison of the structural families obtained for each DOM sample was conducted using

new functionalities added to *Graph-DOM.* For this purpose, precursor molecules forming a family were concatenated and defined as mathematical *sets*. Common families across the four DOM samples were identified by intersecting (∩) all *sets.* Unique families to each sample were determined by subtracting the union (∪) of three of the family *sets* to the reaming *one*. For example, unique structural families to SRFA sample were determined by subtracting the union of Pantanal, HR-1, and HR-5 *sets* to the SRFA *set.*

Isomeric information for the precursor molecules of the structural were obtained from (-) ESI-TIMS-FT ICR MS. TIMS profiles of the precursor formulas were extracted using Data Analysis (v. 5.2, Bruker Daltonics, City, CA, USA). The ion mobility spectra were externally calibrated using the reported ion mobilities of the Agilent Tuning Mix calibration standard[54]. Extracted ion mobility profiles for each chemical formula were deconvoluted using the custom-built Software Assisted Molecular Elucidation (SAME) written in Python v3.7.3. The SAME package relies on noise removal, mean gap filling, asymmetric least squares smoothing for base line correction, continuous wavelet transform (CWT)-based peak detection (SciPy package), and Gaussian fitting with non-linear least squares functions[55]. Common structural isomers for precursors belonging to the families shared by all DOM samples were filtered by concatenating CCS values with chemical formulas. The same mathematical *set* approach described above for the identification of common and unique families was utilized for the identification of common structural isomers across DOM samples sharing the same structural families.

6.4 Results and discussion

The broadband (-)ESI-FT-ICR MS analysis of SRFA, Pantanal, HR-1, and HR-5 DOM samples resulted in the typical unimodal distribution of signals centered at ~340 *m/z*

(Figure 6.1, left panel). The SRFA, Pantanal, and HR-1 samples, all with wetland characteristics, showed the highest chemical diversity with an average of 2920 chemical formulas assigned in the mass range 100-800 *m/z*. On the other hand, HR-5, the estuarine DOM sample, showed a less chemically diverse fingerprint (<2000 chemical components). The distribution of heteroatoms classes across samples showed that the CHO class (58%) dominates over the CHON (19%), CHOS (12%), and CHONS (1%) classes (Appendix 6.1), in good agreement with previous reports of wetland samples[9]. Over 30% more CHO components were found for SRFA and Pantanal samples compared to HR-1 and HR-5. Moreover, HR-1 and HR-5 exhibited ~10% more enriched CHON and CHOS signatures than SRFA and Pantanal DOMs. The heteroatom pattern comparison provides a first level description of DOM signatures at the molecular level.

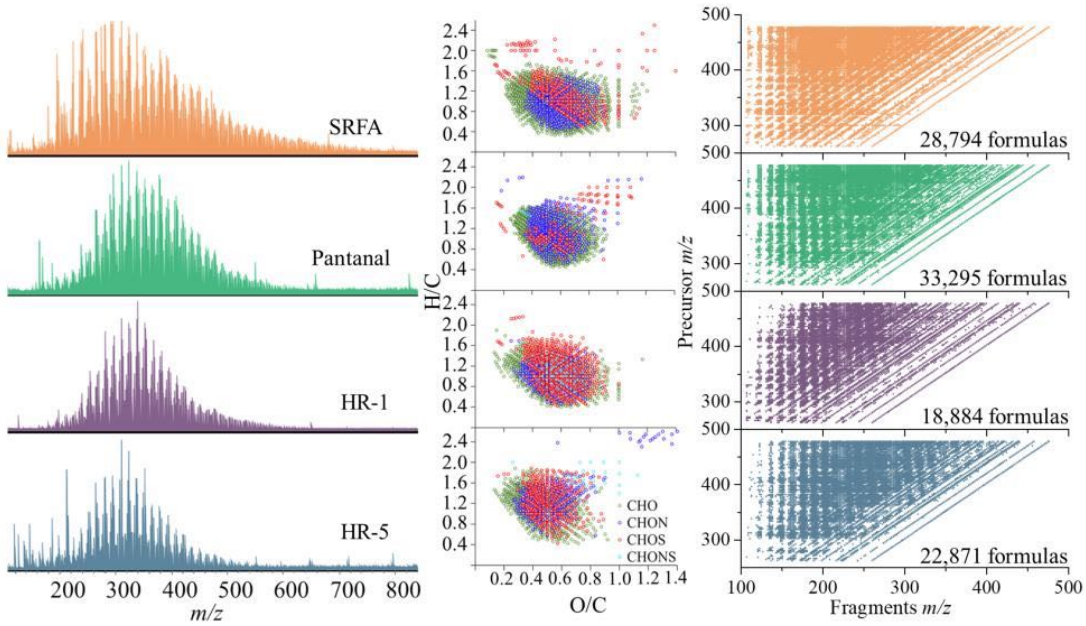


Figure 6. 1. (-)ESI-FT-ICR MS spectra of the SRFA, Pantanal, HR-1, and HR-5 DOM samples (left panel), van Krevelen plots of the DOM samples highlighting the CHO, CHON, CHOS, and CHONS heteroatom classes (center panel), and 2D MS/MS plots obtained from the analysis of the DOM samples by the (-)ESI-FT-ICR CASI-CID MS/MS.

The DOM compositional representation in the van Krevelen space showed the preponderance of lignin-like and tannins-like components (Figure 6.2, center panel). Nevertheless, it should be clearly noted that these structural-like assignments are only indicative since they are solely based on chemical composition. A further structural comparison was conducted based on the neutral loss fragmentation pattern.

Inspection of the 2D MS/MS maps for the CHO class showed a similar pattern of typical neutral loss lines (i.e., H_2O, CH_4O, CH_2O, and CO_2 losses) and their multiples across the DOM samples (Figure 6.1, right panel). The fragment assignment of over 200 MS/MS spectra per DOM sample evidenced a higher structural diversity for the CHO class of SRFA and Pantanal samples (> 10,000 fragment + precursor chemical formulas) compared with HR-1 and HR-5 samples. While this information was clearly uncovered by the CASI-CID MS/MS experiments, it was hindered at the MS^1 level since HR-1 and HR-5 samples contained even more precursor molecules than SRFA and Pantanal samples (Appendix 6.1).

Inspection of isolated mass signals at nominal mass 365 and their corresponding MS2 profiles (Appendix 6.2) revealed that all DOM samples yielded a similar MS/MS fragmentation pattern with characteristic fragments to each DOM sample. A higher number of neutral losses was observed for SRFA and Pantanal samples compared with HR-1 and HR-5. Neutral losses are typically associated with structural functionalities. We interpret these results as that SRFA and Pantanal DOM samples have a larger structural diversity than HR-1 and HR-5 DOM samples.

The *Graph-DOM* schematics for the determination of structural families per DOM sample is visualized in Figure 6. 2 (top panel). When applied to the CHO class, near two-

fold more fragmentation pathways were obtained for SRFA and Pantanal samples in comparison to HR-1 and HR-5 (Appendix 6.3 top panel). In particular, SRFA and HR-1 exhibited the highest (10^8) and lowest (10^6) number of pathways, respectively. The number of structural families showed a similar trend across the DOMs: SRFA>Pantanal>HR-5>HR-1. An average family size distribution with 4-5 precursors was observed across all DOM samples (Appendix 6.3 center panel); up to seven family members were found only for SRFA as an indication of its higher structural complexity.

Inspection of the CHO families showed that the top precursor (highest m/z) within a family for the Pantanal and HR-5 families showed 6-10 oxygens when compared with the SRFA and HR-1 structural families with over >12 oxygens (Appendix 6.3 bottom). These results are in good agreement with previous reports where a higher structural diversity has been associated with a higher number of oxygens[42]. In particular, the SRFA higher structural diversity is exemplified by the larger family sizes (up to 7 precursors) and top family precursors with up to 15 oxygens. The *Graph-DOM* method resulted effective for the identification of structural commonalities and dissimilarities across DOMs. Further interpretation of the structural families can provide insight towards the correlation between DOC, DOM source, biogeochemical transformations, among others[11].

The observation of a single complex interconnected cluster of DOM components in the Cytoscape structural network of SRFA is consistent with its higher structural complexity (Figure 6.2, bottom panel). Although, Pantanal's structural network (three interconnected clusters) looks slightly different than SRFA's network, they both shared similar interconnection complexities.

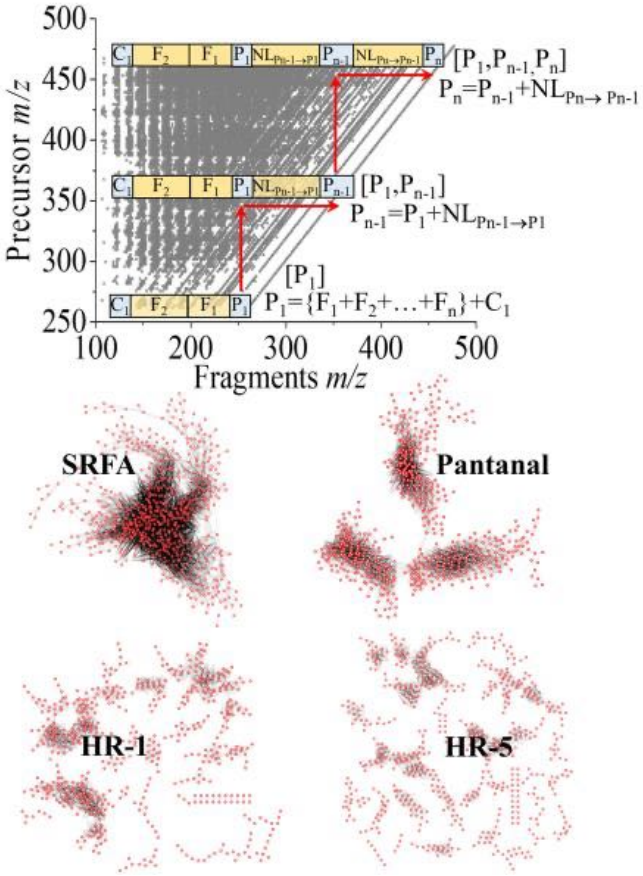


Figure 6. 2. Schematic representation of Graph-DOM models used for computing fragmentation pathways (top left panel) and identifying structural families (bottom left panel) in a DOM sample. Cytoscape networks of interconnected precursors belonging to the identified structural families for SRFA, Pantanal, HR-1, and HR-5 DOM samples (right panel).

The scattered clusters observed for HR-1 and HR-5 indicate that these two samples are structurally more similar between each other, and less complex than Pantanal and SRFA, respectively. The structural similarity of HR-1 and HR-5 could be explained by the fact that both samples belong to the same ecosystem (Harney River, Everglades). The differences observed in the structural networks (HR-1 contains more complex clusters than HR-5) evidence different DOM biogeochemical transformations and different inputs associated to the different mangrove and sea ecosystems (Appendix 6.4). While this

comparison of structural families was limited to the CHO class (most abundant class), further inspection of other heteroatom classes can provide additional information.

The *Graph-DOM* provided a list of structural families for each DOM sample. A comparison of the structural families is shown using a Venn diagram in Figure 6.3. Note that the comparison using family *sets* (Appendix 6.5) enabled the identification of common and unique structural families across the four DOM samples. Results showed that 8% of the CHO families across the four DOM samples are common; that is, over 200 families of structurally related CHO precursors are ubiquitous across distinctive aquatic environments.

The number of unique structural CHO families from the four DOM samples follows the same observed trend as the one from the number of fragmentation pathways and number of families (SRFA>Pantanal>HR-5>HR-1). A high structural complexity of SRFA (>2500 unique families) and a high similarity with Pantanal sample (>800 shared families) was observed. The Cytoscape structural network of the 218 CHO families common to all DOM samples showed a pattern of disconnected multi-clusters (Figure 6.3 bottom panel). The lack of connectivity among structural families common to all samples could be associated with a lower structural diversity of HR-1 and HR-5 families compared to SRFA and Pantanal. We interpret these results as dissimilar degradation and mixing processes across aquatic environments. A closer look at the network of intersected families of SRFA and Pantanal but excluding both HR-1 and HR-5 families (Appendix 6.6), confirmed the structural similarities of both SRFA and Pantanal samples. The exclusion of SRFA and Pantanal CHO families from the intersection of HR-1 and HR-5 CHO families resulted in a network that resembles the everglades ecosystem (Appendix 6.6). These results showed

that *Graph-DOM* method is effective for the determination of structural fingerprints associated to a DOM origin.

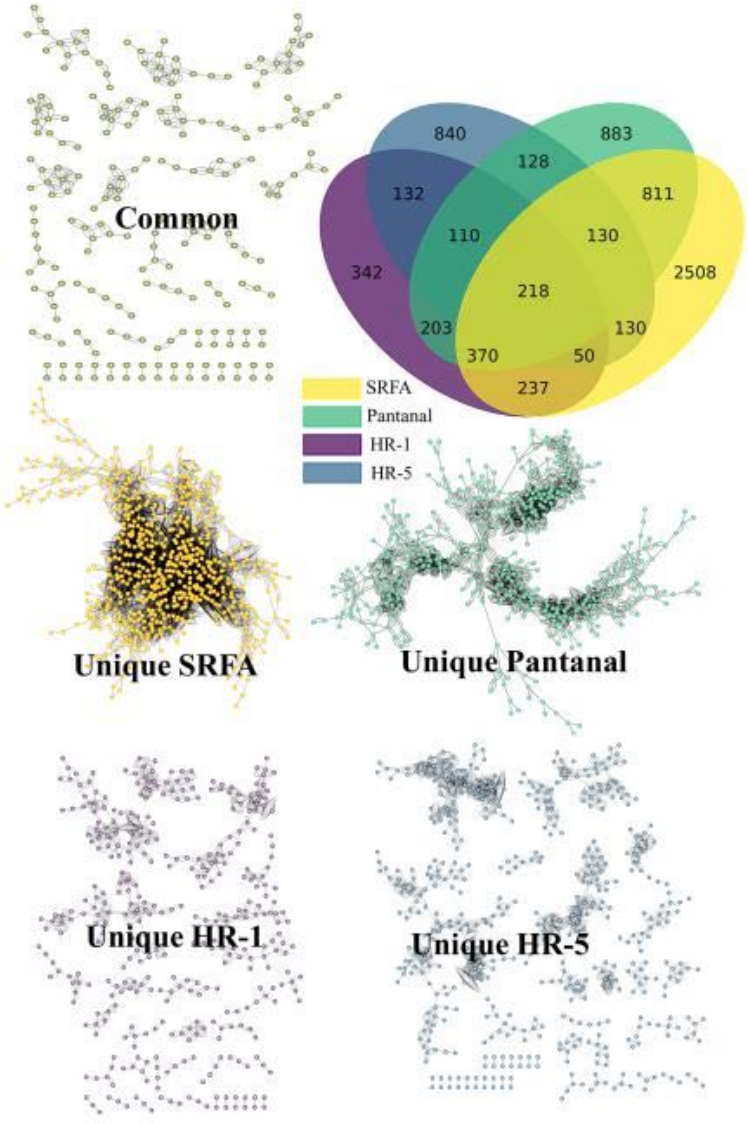


Figure 6. 3. Van Krevelen plot showing common structural families across DOM samples and unique families to SRFA, Pantanal, HR-1, and HR-5 determined by the comparison of families using *Graph-DOM* (top right panel). Cytoscape structural networks of common families found across the DOM samples (top left panel) and unique families to each sample (bottom).

Inspection of the compositional characteristics of the CHO precursors within the common CHO families showed transformational processes associated to hydration and

carboxylation steps. Three examples of common CHO structural families involving hydroxylic (H_2O loss) and carboxylic (CO_2 loss) structural transitions among precursors is depicted in Figure 6.4. These transitions can be viewed as both synthetic and degradation pathways that illustrate some of complex biogeochemical mechanisms involved in DOM structural transformation. The absence of methoxy (CH_4O loss)-based transformational processes in the common CHO structural families suggests that the mechanisms triggering the addition/subtraction of this functionality are characteristic to certain aquatic ecosystems. Examination of the neutral loss sequences among precursors in the families (data not shown) evidenced that at least one methoxy moiety was found in ~ 50% of both SRFA and Pantanal families. In contrast, lower abundance of this functionality was observed across HR-1 (<7%) and HR-5 (<3%) structural families, thus confirming the similarities between SRFA/Pantanal and HR-1/HR-5 pairs at the structural level.

The three examples of common families superimposed on a van Krevelen plot (Figure 6.4) revealed continuous (orange and green) and discontinuous (pink) lines describing potential transformational pathways. These patterns illustrate that, in the context of our study (selected samples, conceptual model, and experimental set up), a set of constrained reactions pathways involving only addition/subtraction of hydroxyl and carboxylic moieties could be responsible for the ubiquitous pool of DOM structures across aquatic environments.

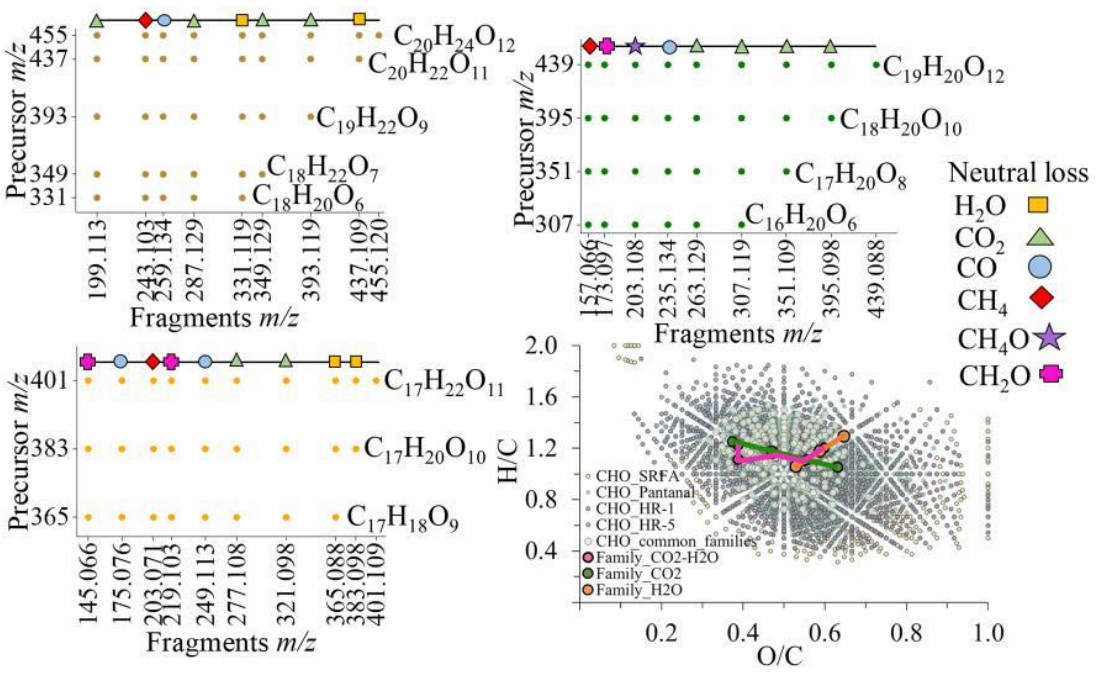


Figure 6. 4. Representation of three common CHO families. The H_2O-CO_2-CO_2-H_2O (top left), CO_2-CO_2-CO_2 (top right), and H_2O-H_2O (bottom left) families are depicted with one homologous fragmentation pathways in the form of 2D MS/MS fragments *m/z* vs precursor *m/z* plots. A van Krevelen plot (bottom right) showcases the compositional pattern of the CHO class components from the SRFA, Pantanal, HR-1, and HR-5 samples, the CHO precursors from the structural families shared by all DOM samples, and the three selected structural families.

Complementary structural information can be derived from the use of mobility (TIMS) and neutral loss (CASI-CID FT ICR MS/MS) experiments. A model describing the integration of the precursor isomeric information and neutral loss-based family identification is described in Appendix 6.7. Briefly, the list of unique precursor formulas from the shared families was tagged with the isomeric information ($^{TIMS}CCS_{N2}$ values) from each DOM sample. The comparison of the new sets enabled the identification of common and unique precursor isomers shared by the common structural families (see Venn diagram in Figure 6.5 top). A total of 450 isomeric precursors were shared by the common

CHO structural families. Up to 4 isomers and an average of two isomeric species were found for precursors in the common families. Interestingly, the number of unique isomers found across precursors from common families exhibited a decreasing trend in the order HR-5>HR-1>SRFA>Pantanal. These differences in isomeric content among common precursors could be one of the possible explanations behind the dissimilarities observed between the structural families of SRFA/Pantanal and HR-1/HR-5 pairs. However, it should be noted that the isomeric diversity provided in this study is a lower estimate[12]; that is, mobility unresolved isomers should be accounted for.

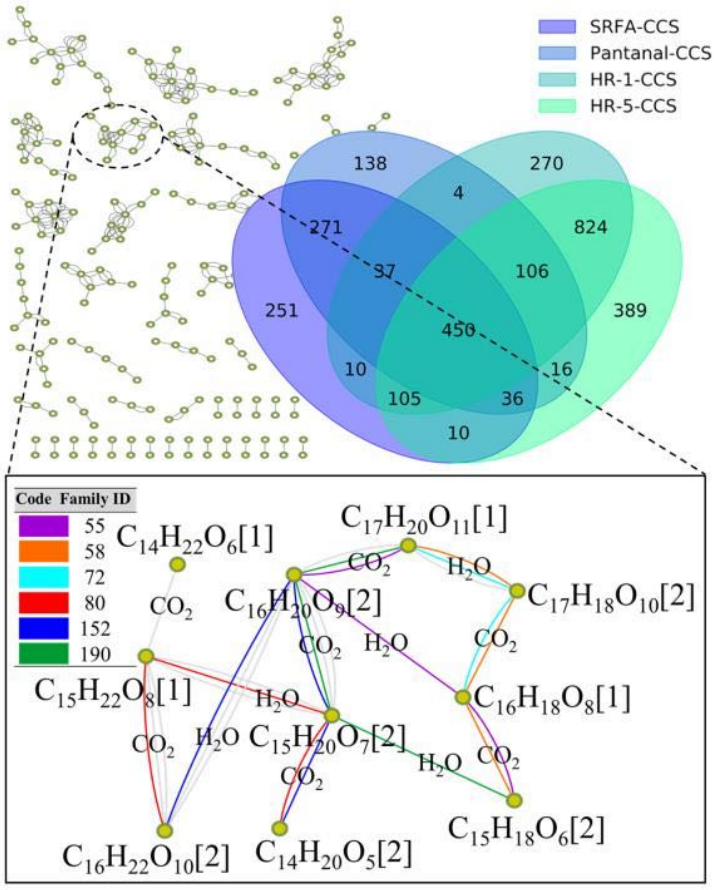


Figure 6. 5. Venn diagram describing common and unique isomers for the precursors of the families shared by all DOM samples (top). Section of the structural network of families common to all samples (bottom) highlighting the interconnected precursor formulas (nodes), the family IDs (color code), the neutral based functionalities (edge label),

and the number of common isomers across DOM samples per precursor (pink bars). Note that grey lines are associated with other not labeled shorter families.

An expanded view of a cluster from the network of common CHO families shows intra-connections of up to four precursors within a family (colored lines) and the complex inter-connection among all families (Figure 6.5 bottom). The layer of isomeric information added on top of the family domain shows the presence of up to two isomers per precursor (number in brackets). In our previous report[42], we suggested that the lines interconnecting two precursors in a network could be an indication of isomeric diversity based solely on the neutral loss information. A closer look at the expanded network on the bottom of Figure 5 reveals differences between the number of isomers estimated based on a neutral loss model (colored +grey lines) and mobility separation (number in brackets). For example, at least four isomers can be estimated for the $C_{15}H_{20}O_7$ precursor based on its CO_2 loss interconnections with $C_{16}H_{20}O_9$. On the other hand, the estimated lower number of isomers for $C_{15}H_{20}O_7$ based on the SAME fitting was two. Considering the potential ambiguities and shortcoming of each approach, the results are remarkable. Further integration of data using the chemical formula mobility selected MS/MS analytical workflow will be helpful in providing a training set of DOM unambiguous fragmentation information that can be incorporated in a machine learning algorityhm[41].

The results presented in this study show the potential of data mining of ultra-high resolution neutral loss fragmentation patterns using the *Graph-DOM* method in the description of DOM transformational networks at the molecular level. The role of the isomeric diversity at the molecular level was successfully integrated for the first time with the structural family description. The application of this methodology to four aquatic

ecosystems allowed for the identification of common and unique DOM transformational networks. For the first time, evidence of common transformational processes is provided across the four DOMs. Future studies including samples of DOM photo/biodegradation experiments and samples along salinity transects could provide further insights into the DOM composition and correlation between the chemical components. All the *Graph-DOM* functionalities developed and described in this paper are freely accessible at https://github.com/Usman095/Graph-DOM.

6.5 References

1. Chantigny, M. H. Dissolved and water-extractable organic matter in soils: a review on the influence of land use and management practices. *Geoderma* **2003,** *113* (3), 357-380.

2. Tremblay, L.; Gagné, J.-P. Organic matter distribution and reactivity in the waters of a large estuarine system. *Marine Chemistry* **2009,** *116* (1), 1-12.

3. Ridgwell, A.; Arndt, S. Chapter 1 - Why Dissolved Organics Matter: DOC in Ancient Oceans and Past Climate Change. In *Biogeochemistry of Marine Dissolved Organic Matter (Second Edition)*, Hansell, D. A.; Carlson, C. A., Eds.; Academic Press: Boston, 2015; pp 1-20.

4. Moran, M. A.; Kujawinski, E. B.; Stubbins, A.; Fatland, R.; Aluwihare, L. I.; Buchan, A.; Crump, B. C.; Dorrestein, P. C.; Dyhrman, S. T.; Hess, N. J.; Howe, B.; Longnecker, K.; Medeiros, P. M.; Niggemann, J.; Obernosterer, I.; Repeta, D. J.; Waldbauer, J. R. Deciphering ocean carbon in a changing world. *Proceedings of the National Academy of Sciences* **2016,** *113* (12), 3143-3151.

5. Lønborg, C.; Carreira, C.; Jickells, T.; Álvarez-Salgado, X. A. Impacts of Global Change on Ocean Dissolved Organic Carbon (DOC) Cycling. *Frontiers in Marine Science* **2020,** *7*.

6. Fakhraee, M.; Tarhan, L. G.; Planavsky, N. J.; Reinhard, C. T. A largely invariant marine dissolved organic carbon reservoir across Earth's history. *Proceedings of the National Academy of Sciences* **2021,** *118* (40), e2103511118.

7. Wagner, S.; Jaffé, R.; Cawley, K.; Dittmar, T.; Stubbins, A. Associations Between the Molecular and Optical Properties of Dissolved Organic Matter in the Florida Everglades, a Model Coastal Wetland System. *Frontiers in Chemistry* **2015,** *3* (66), 1-14.

8. Osterholz, H.; Kirchman, D. L.; Niggemann, J.; Dittmar, T. Environmental Drivers of Dissolved Organic Matter Molecular Composition in the Delaware Estuary. *Frontiers in Earth Science* **2016,** *4* (95), 1-14.

9. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016,** *13* (8), 2257-2277.

10. Zark, M.; Christoffers, J.; Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Marine Chemistry* **2017,** *191*, 9-15.

11. Zark, M.; Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nature Communications* **2018,** *9* (1), 3178.

12. Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the structural complexity of dissolved organic matter: isomeric diversity. *Faraday Discussions* **2019,** *218* (0), 431-440.

13. Xu, W.; Gao, Q.; He, C.; Shi, Q.; Hou, Z.-Q.; Zhao, H.-Z. Using ESI FT-ICR MS to Characterize Dissolved Organic Matter in Salt Lakes with Different Salinity. *Environmental Science & Technology* **2020,** *54* (20), 12929-12937.

14. Asmala, E.; Massicotte, P.; Carstensen, J. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnology and Oceanography* **2021,** *66* (4), 1381-1393.

15. He, D.; Li, P.; He, C.; Wang, Y.; Shi, Q. Eutrophication and watershed characteristics shape changes in dissolved organic matter chemistry along two river-estuarine transects. *Water Research* **2022,** *214*, 118196.

16. Seidel, M.; Beck, M.; Riedel, T.; Waska, H.; Suryaputra, I. G. N. A.; Schnetger, B.; Niggemann, J.; Simon, M.; Dittmar, T. Biogeochemistry of dissolved organic matter in an anoxic intertidal creek bank. *Geochimica et Cosmochimica Acta* **2014,** *140*, 418-434.

17. Antony, R.; Willoughby, A. S.; Grannas, A. M.; Catanzano, V.; Sleighter, R. L.; Thamban, M.; Hatcher, P. G.; Nair, S. Molecular Insights on Dissolved Organic Matter Transformation by Supraglacial Microbial Communities. *Environmental Science & Technology* **2017,** *51* (8), 4328-4337.

18. Kim, S.; Kim, D.; Jung, M.-J.; Kim, S. Analysis of environmental organic matters by Ultrahigh-Resolution mass spectrometry—A review on the development of analytical methods. *Mass Spectrometry Reviews* **2021,** *n/a* (n/a).

19. Zhang, Y.; Zhou, Y.; Shi, K.; Qin, B.; Yao, X.; Zhang, Y. Optical properties and composition changes in chromophoric dissolved organic matter along trophic gradients:

Implications for monitoring and assessing lake eutrophication. *Water Research* **2018,** *131*, 255-263.

20. O'Connor, J. A.; Lu, K.; Guo, L.; Rosenheim, B. E.; Liu, Z. Composition and lability of riverine dissolved organic matter: Insights from thermal slicing ramped pyrolysis GC–MS, amino acid, and stable isotope analyses. *Organic Geochemistry* **2020,** *149*, 104100.

21. Gonsior, M.; Powers, L.; Lahm, M.; McCallister, S. L. New Perspectives on the Marine Carbon Cycle–The Marine Dissolved Organic Matter Reactivity Continuum. *Environmental Science & Technology* **2022,** *56* (9), 5371-5380.

22. Smith, D. F.; Podgorski, D. C.; Rodgers, R. P.; Blakney, G. T.; Hendrickson, C. L. 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures. *Analytical Chemistry* **2018,** *90* (3), 2041-2047.

23. Berg, S. M.; Peterson, B. D.; McMahon, K. D.; Remucal, C. K. Spatial and Temporal Variability of Dissolved Organic Matter Molecular Composition in a Stratified Eutrophic Lake. *Journal of Geophysical Research: Biogeosciences* **2022,** *127* (1), 1-16.

24. Seidel, M.; Vemulapalli, S. P. B.; Mathieu, D.; Dittmar, T. Marine Dissolved Organic Matter Shares Thousands of Molecular Formulae Yet Differs Structurally across Major Water Masses. *Environmental Science & Technology* **2022,** *56* (6), 3758-3769.

25. Lu, K.; Liu, Z. Molecular Level Analysis Reveals Changes in Chemical Composition of Dissolved Organic Matter From South Texas Rivers After High Flow Events. *Frontiers in Marine Science* **2019,** *6* (673).

26. Hertkorn, N.; Harir, M.; Schmitt-Kopplin, P. Nontarget analysis of Murchison soluble organic matter by high-field NMR spectroscopy and FTICR mass spectrometry. *Magnetic Resonance in Chemistry* **2015,** *53* (9), 754-768.

27. Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **2013,** *10* (3), 1583-1624.

28. Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science* **2017,** *4* (405), 1-14.

29. Qi, Y.; Ma, C.; Chen, S.; Ge, J.; Hu, Q.; Li, S.-L.; Volmer, D. A.; Fu, P. Online Liquid Chromatography and FT-ICR MS Enable Advanced Separation and Profiling of Organosulfates in Dissolved Organic Matter. *ACS ES&T Water* **2021,** *1* (8), 1975-1982.

30. Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J. R.; Tranvik, L. J.; Bergquist, J. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnology and Oceanography Letters* **2018,** *3* (2), 21-30.

31. Lu, K.; Li, X.; Chen, H.; Liu, Z. Constraints on isomers of dissolved organic matter in aquatic environments: Insights from ion mobility mass spectrometry. *Geochimica et Cosmochimica Acta* **2021,** *308*, 353-372.

32. Gao, Y.; Wang, W.; He, C.; Fang, Z.; Zhang, Y.; Shi, Q. Fractionation and molecular characterization of natural organic matter (NOM) by solid-phase extraction followed by FT-ICR MS and ion mobility MS. *Analytical and Bioanalytical Chemistry* **2019**.

33. Tose, L. V.; Benigni, P.; Leyva, D.; Sundberg, A.; Ramírez, C. E.; Ridgeway, M. E.; Park, M. A.; Romão, W.; Jaffé, R.; Fernandez-Lima, F. Coupling trapped ion mobility spectrometry to mass spectrometry: trapped ion mobility spectrometry–time-of-flight mass spectrometry versus trapped ion mobility spectrometry–Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* **2018,** *32* (15), 1287-1295.

34. Powers, L. C.; Lapham, L. L.; Malkin, S. Y.; Heyes, A.; Schmitt-Kopplin, P.; Gonsior, M. Molecular and optical characterization reveals the preservation and sulfurization of chemically diverse porewater dissolved organic matter in oligohaline and brackish Chesapeake Bay sediments. *Organic Geochemistry* **2021,** *161*, 1-15.

35. Pan, Q.; Zhuo, X.; He, C.; Zhang, Y.; Shi, Q. Validation and Evaluation of High-Resolution Orbitrap Mass Spectrometry on Molecular Characterization of Dissolved Organic Matter. *ACS Omega* **2020,** *5* (10), 5372-5379.

36. Hawkes, J. A.; Dittmar, T.; Patriarca, C.; Tranvik, L.; Bergquist, J. Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Analytical Chemistry* **2016,** *88* (15), 7698-7704.

37. Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a major refractory component of marine dissolved organic matter. *Geochimica et Cosmochimica Acta* **2006,** *70* (12), 2990-3010.

38. Cortés-Francisco, N.; Caixach, J. Fragmentation studies for the structural characterization of marine dissolved organic matter. *Analytical and Bioanalytical Chemistry* **2015,** *407* (9), 2455-2462.

39. LeClair, J. P.; Collett, J. L.; Mazzoleni, L. R. Fragmentation Analysis of Water-Soluble Atmospheric Organic Matter Using Ultrahigh-Resolution FT-ICR Mass Spectrometry. *Environmental Science & Technology* **2012,** *46* (8), 4312-4322.

40. Osterholz, H.; Niggemann, J.; Giebel, H.-A.; Simon, M.; Dittmar, T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nature Communications* **2015,** *6* (1), 1-8.

41. Leyva, D.; Jaffe, R.; Fernandez-Lima, F. Structural Characterization of Dissolved Organic Matter at the Chemical Formula Level Using TIMS-FT-ICR MS/MS. *Analytical Chemistry* **2020,** *92* (17), 11960-11966.

42. Leyva, D.; Tariq, M. U.; Jaffé, R.; Saeed, F.; Lima, F. F. Unsupervised Structural Classification of Dissolved Organic Matter Based on Fragmentation Pathways. *Environmental Science & Technology* **2022,** *56* (2), 1458-1468.

43. Tariq, M. U.; Leyva, D.; Lima, F. A. F.; Saeed, F. In *Graph Theoretic Approach for the Analysis of Comprehensive Mass-Spectrometry (MS/MS) Data of Dissolved Organic Matter*, 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 9-12 Dec. 2021; 2021; pp 3742-3746.

44. Li, Y.; Harir, M.; Lucio, M.; Kanawati, B.; Smirnov, K.; Flerus, R.; Koch, B. P.; Schmitt-Kopplin, P.; Hertkorn, N. Proposed Guidelines for Solid Phase Extraction of Suwannee River Dissolved Organic Matter. *Analytical Chemistry* **2016,** *88* (13), 6680-6688.

45. Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* **2008,** *6* (6), 230-235.

46. Benigni, P.; Fernandez-Lima, F. Oversampling Selective Accumulation Trapped Ion Mobility Spectrometry Coupled to FT-ICR MS: Fundamentals and Applications. *Analytical Chemistry* **2016,** *88* (14), 7404-7412.

47. Benigni, P.; Sandoval, K.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Gardinali, P.; Fernandez-Lima, F. Analysis of Photoirradiated Water Accommodated Fractions of Crude Oils Using Tandem TIMS and FT-ICR MS. *Environmental Science & Technology* **2017,** *51* (11), 5978-5988.

48. Benigni, P.; Porter, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Increasing Analytical Separation and Duty Cycle with Nonlinear Analytical Mobility Scan Functions in TIMS-FT-ICR MS. *Analytical Chemistry* **2018,** *90* (4), 2446-2450.

49. Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A. Gas-phase separation using a trapped ion mobility spectrometer. *International Journal for Ion Mobility Spectrometry* **2011,** *14* (2), 93-98.

50. Fernandez-Lima, F. A.; Kaplan, D. A.; Park, M. A. Note: Integration of trapped ion mobility spectrometry with mass spectrometry. *Review of Scientific Instruments* **2011,** *82* (12), 126106.

51. Hernandez, D. R.; DeBord, J. D.; Ridgeway, M. E.; Kaplan, D. A.; Park, M. A.; Fernandez-Lima, F. Ion dynamics in a trapped ion mobility spectrometer. *Analyst* **2014,** *139* (8), 1913-1921.

52. Herzsprung, P.; Hertkorn, N.; von Tümpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Analytical and Bioanalytical Chemistry* **2014,** *406* (30), 7977-7987.

53. Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P.-L.; Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2010,** *27* (3), 431-432.

54. Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility–Mass Spectrometry Collision Cross Section Measurements. *Analytical Chemistry* **2017,** *89* (17), 9048-9055.

55. Benigni, P.; Marin, R.; Sandoval, K.; Gardinali, P.; Fernandez-Lima, F. Chemical Analysis of Water-accommodated Fractions of Crude Oil Spills Using TIMS-FT-ICR MS. . *J. Vis. Exp.* **2017,** *121*, e55352.

CHAPTER VII

VII. CONCLUSIONS

In this dissertation an advanced analytical toolbox based on TIMS-FT-ICR MS/MS and computational algorithms to address the molecular complexity and diversity of dissolved organic matter was developed. Two novel applications of the developed workflows for the study of DOM structural complexity across various aquatic ecosystems are presented. In a first contribution published in the journal *Faraday Discussions*, the advantages of untargeted TIMS-FT-ICR MS/MS to address the isomeric content of wetland DOM samples is illustrated. The analysis permitted the identification of near 3000 chemical components and an average estimate of 8 structural isomers per chemical formula, in a single direct infusion experiment. Highly oxygenated components (up to O_{20}) were the most abundant chemical compounds (~80%) in the wetland DOM sample. The ESI-q-FT-ICR MS/MS analysis performed at the level of nominal mass, allowed for an estimation of the number of structural isomers based on unique fragmentation patterns. In summary, the comparison of mobility and fragmentation data showed that multiple structural isomers could share very close mobility values, thus demanding the use of novel analytical procedures based on ultrahigh resolution TIMS mobility in tandem with FT-ICR MS/MS.

A novel ESI-TIMS-FT-ICR MS/MS workflow for the structural analysis of complex DOM samples at the level of chemical formula was developed in a second work showcased in the journal *Analytical Chemistry*. High-resolution ion mobility combined with single precursor ion isolation in the ICR with a very narrow $\Delta m/z$ of 0.036, and fragment assignment with high mass accuracy, permitted the structural assignment of single species in a complex mixture (i.e., mixture containing isomeric interferences). Single species from a model mixture containing 4-methoxy-1-naphthoic acid, 2-methoxy-1-

naphthoic acid, and decanedioic acid (isomers and isobars, respectively), were unambiguously identified either via ion mobility or characteristic fragmentation spectra. Therefore, the extraordinary capabilities of the procedure in comparison with traditional nominal mass MS/MS schemes (co-isolation and fragmentation of several precursors) were demonstrated. The MS/MS analysis of a selected chemical formula (*m/z* 393.0828) from a DOM sample revealed that our protocol can effectively isolate and fragment a single compound from several isobaric species in the ICR cell. The IMS information obtained for the isolated compound $[C_{18}H_{18}O_{10}\text{-}H]^-$ was correlated with the fragmentation data generated by CHEF+SHOTS-SORI-CID MS/MS yielding candidate isomeric structures from the PubChem database screened based on their ion mobility and MS/MS matching score. This work provided a proof of concept for the structural analysis of DOM at the chemical formula level integrating high-resolution ion mobility separations with single precursor fragmentation in the ICR cell.

The application of our CHEF+SHOTS-SORI-CID MS/MS procedure to obtain structural information from thousands of chemical species typically found in a DOM sample is a very challenging task. Therefore, simplified strategies capable of establishing structural patterns across DOM molecules using untargeted ultrahigh resolution MS and tandem MS/MS information with shorter experimental and processing time scales, are needed. An unsupervised ESI-FT-ICR MS/MS workflow relying on CASI-CID of precursor ions at nominal mass combined with a data mining approach was developed to better understand the structural complexity of a wetland DOM sample. The structural complexity obtained from families of structurally connected precursors using neutral mass loss patterns (Pn-1+F1:n+C) across a 2D MS/MS space was studied. Our methodology was capable to

identify over 1,000 structural families of DOM component based on precursor and neutral losses (H_2O, CH_4O and CO_2). A high degree of isomeric content (numerous identical fragmentation pathways), hindered at the precursor level, was observed. A connectivity map generated for the structural families highlighted potential biogeochemical processes that DOM undergoes throughout its lifetime. This work, published in *Environmental Science and Technology*, demonstrates that advanced computational algorithms combined with comprehensive high-resolution MS/MS approaches are useful tools for unscrambling the extreme molecular complexity of DOM. In combination, our developed procedures constitute a novel and versatile toolbox capable of providing 1) candidate isomeric structures for DOM compounds filtered by accurate *m/z*, fragment score, and collisional cross sections, and at a larger scale, 2) families of structurally related DOM compounds.

The last two chapters of this dissertation provide the practical application of the novel *GraphDOM* workflows to the molecular level characterization of DOM across various ecosystems. In the first case, the DOM isomeric complexity along a freshwater marsh fed mangrove estuary salinity transect at the Harney River, Florida Everglades using untargeted ESI-TIMS-FT ICR MS was studied for the first time. Our protocol was able to detect a decrease in the number of chemical components and isomeric species along the increasing salinity gradient. A trend of more unsaturated and oxygenated to more aliphatic and heteroatom diverse components downstream was also observed. Novel evidence about the isomeric refractory character of terrestrial-derived DOM components was provided based on a set of structurally unchanged isomers ($\Delta^{TIMS}CCS_{N2}$ values < 8%) found across the transect. The IMS data obtained for common chemical formulas along the transect suggested that freshwater marsh-derived material was up to four times more isomeric

complex than marine primary producers and mangrove DOM. The unique chemical and isomeric fingerprints obtained across samples in this study, are proposed as potential markers that could improve our understanding of how input sources and biogeochemical transformations shape the DOM molecular makeup. Our results illustrate that ESI-TIMS-FT-ICR MS is an important tool that will contribute to expand the current knowledge of the carbon fate in coastal aquatic ecosystems.

In the second case, the structural diversity of four common SPE-DOM samples (Suwanee River fulvic acid standard, Pantanal, HR-1 and HR-5) from different aquatic ecosystems was used. The *GraphDOM* was further developed to integrate CASI-CID MS/MS methodology and TIMS-FT-ICR MS information. Clear structural dissimilarities based on fragmentation pathways and compound families across the DOM samples in the order SRFA>Pantanal>HR-5>HR-1 were uncovered. Characteristic graphical patterns observed across the structural networks confirmed the structural differences among samples. Moreover, similarities across samples based on over 200 shared structural families were identified for the first time. The neutral loss fragmentation analysis showed evidenced of ubiquitous hydration and carboxylation transformational processes across ecosystems, and unique methoxy-based mechanisms for SRFA and Pantanal samples. The analysis of the isomeric content across common families showed a new layer of similarities and differences that could be used for further fingerprinting purposes.

In combination, our workflows are a powerful platform that will certainly expand our knowledge about the chemical complexity of DOM. Studies including changes in DOM structural fingerprint due to photo and biodegradation, and mixing of input sources, could potentially benefit from our novel approach. However, improvements in separation

methodologies, availability of DOM-like standards, fragment assignment, customized statistical tools, and computing power will considerably help in advancing this field to a much higher level. For example, exploring chromatographic approaches such as hydrophilic interaction (HILIC) or size exclusion (SEC) will provide an alternative orthogonal layer of separation in the polarity and size domain respectively. The integration of these platforms with high resolution ion mobility, namely TIMS and structures for lossless ion manipulations (SLIM), will further enable the characterization of DOM at the structural level. A natural push would include testing the effectiveness of these hyphenated approaches in the separation of a complex isomeric mixture of standards, mimicking DOM isomeric diversity. Moreover, minimizing the level of ambiguity during fragment assignment using CASI-CID will enhance the computation of fragmentation pathways and structural families using our *Graph-DOM* tool.

APPENDICIES

Appendix 2. 1. FT ICR MS/MS fragmentation data for 391 m/z with nominal mass isolation from DOM sample PAN-S.

Measured m/z	Ion Formula	Theoretical m/z	err [ppm]
155.0349	$C_7H_7O_4$	155.0350	0.7
161.0607	$C_{10}H_9O_2$	161.0608	0.7
163.0763	$C_{10}H_{11}O_2$	163.0765	0.8
165.0192	$C_8H_5O_4$	165.0193	0.7
165.0556	$C_9H_9O_3$	165.0557	0.8
167.0349	$C_8H_7O_4$	167.0350	0.5
167.0712	$C_9H_{11}O_3$	167.0714	0.8
171.0814	$C_{12}H_{11}O$	171.0815	0.9
173.0607	$C_{11}H_9O_2$	173.0608	0.6
175.04	$C_{10}H_7O_3$	175.0401	0.5
175.0763	$C_{11}H_{11}O_2$	175.0765	0.7
177.0556	$C_{10}H_9O_3$	177.0557	0.6
177.092	$C_{11}H_{13}O_2$	177.0921	0.8
179.0348	$C_9H_7O_4$	179.0350	0.8
179.0712	$C_{10}H_{11}O_3$	179.0714	0.8
181.0505	$C_9H_9O_4$	181.0506	0.7
181.0869	$C_{10}H_{13}O_3$	181.0870	0.6
183.0298	$C_8H_7O_5$	183.0299	0.6
183.045	$C_{12}H_7O_2$	183.0452	0.7
183.0814	$C_{13}H_{11}O$	183.0815	1.0
185.0607	$C_{12}H_9O_2$	185.0608	0.7
187.04	$C_{11}H_7O_3$	187.0401	0.6
187.0763	$C_{12}H_{11}O_2$	187.0765	0.7
189.0556	$C_{11}H_9O_3$	189.0557	0.7
189.0919	$C_{12}H_{13}O_2$	189.0921	0.9
191.0348	$C_{10}H_7O_4$	191.0350	0.8
191.0712	$C_{11}H_{11}O_3$	191.0714	0.8
191.1076	$C_{12}H_{15}O_2$	191.1078	0.7
193.0141	$C_9H_5O_5$	193.0142	0.7
193.0505	$C_{10}H_9O_4$	193.0506	0.7
193.0869	$C_{11}H_{13}O_3$	193.0870	0.8
195.0297	$C_9H_7O_5$	195.0299	0.9

195.0661	$C_{10}H_{11}O_4$	195.0663	0.8
197.0454	$C_9H_9O_5$	197.0455	0.8
197.0607	$C_{13}H_9O_2$	197.0608	0.7
197.0818	$C_{10}H_{13}O_4$	197.0819	0.6
199.0399	$C_{12}H_7O_3$	199.0401	0.8
199.0763	$C_{13}H_{11}O_2$	199.0765	0.7
201.0192	$C_{11}H_5O_4$	201.0193	0.7
201.0555	$C_{12}H_9O_3$	201.0557	0.9
202.9984	$C_{10}H_3O_5$	202.9986	0.8
203.0348	$C_{11}H_7O_4$	203.0350	0.7
203.0712	$C_{12}H_{11}O_3$	203.0714	0.8
203.1076	$C_{13}H_{15}O_2$	203.1078	0.8
205.014	$C_{10}H_5O_5$	205.0142	1.0
205.0505	$C_{11}H_9O_4$	205.0506	0.5
205.0869	$C_{12}H_{13}O_3$	205.0870	0.8
205.1232	$C_{13}H_{17}O_2$	205.1234	0.9
207.0297	$C_{10}H_7O_5$	207.0299	0.9
207.0661	$C_{11}H_{11}O_4$	207.0663	0.8
207.1025	$C_{12}H_{15}O_3$	207.1027	0.6
209.0454	$C_{10}H_9O_5$	209.0455	0.7
209.0818	$C_{11}H_{13}O_4$	209.0819	0.8
211.061	$C_{10}H_{11}O_5$	211.0612	0.9
213.0555	$C_{13}H_9O_3$	213.0557	0.9
215.0348	$C_{12}H_7O_4$	215.0350	0.8
215.0712	$C_{13}H_{11}O_3$	215.0714	0.8
215.1076	$C_{14}H_{15}O_2$	215.1078	0.7
215.144	$C_{15}H_{19}O$	215.1441	0.6
217.0141	$C_{11}H_5O_5$	217.0142	0.7
217.0505	$C_{12}H_9O_4$	217.0506	0.8
217.0868	$C_{13}H_{13}O_3$	217.0870	0.8
217.1233	$C_{14}H_{17}O_2$	217.1234	0.7
219.0297	$C_{11}H_7O_5$	219.0299	0.8
219.0661	$C_{12}H_{11}O_4$	219.0663	0.7
219.1025	$C_{13}H_{15}O_3$	219.1027	0.9

221.009	$C_{10}H_5O_6$	221.0092	0.7
221.0454	$C_{11}H_9O_5$	221.0455	0.7
221.0818	$C_{12}H_{13}O_4$	221.0819	0.7
221.1181	$C_{13}H_{17}O_3$	221.1183	0.9
223.0246	$C_{10}H_7O_6$	223.0248	1.1
223.0611	$C_{11}H_{11}O_5$	223.0612	0.6
223.0974	$C_{12}H_{15}O_4$	223.0976	0.7
223.1338	$C_{13}H_{19}O_3$	223.1340	0.8
225.0767	$C_{11}H_{13}O_5$	225.0768	0.6
227.0348	$C_{13}H_7O_4$	227.0350	0.8
227.0712	$C_{14}H_{11}O_3$	227.0714	0.8
227.1076	$C_{15}H_{15}O_2$	227.1078	0.8
229.0141	$C_{12}H_5O_5$	229.0142	0.8
229.0505	$C_{13}H_9O_4$	229.0506	0.8
229.0868	$C_{14}H_{13}O_3$	229.0870	0.8
231.0298	$C_{12}H_7O_5$	231.0299	0.6
231.0661	$C_{13}H_{11}O_4$	231.0663	0.8
231.1025	$C_{14}H_{15}O_3$	231.1027	0.9
233.0454	$C_{12}H_9O_5$	233.0455	0.8
233.0818	$C_{13}H_{13}O_4$	233.0819	0.7
233.1181	$C_{14}H_{17}O_3$	233.1183	0.8
233.1545	$C_{15}H_{21}O_2$	233.1547	1.1
235.0246	$C_{11}H_7O_6$	235.0248	0.8
235.061	$C_{12}H_{11}O_5$	235.0612	0.8
235.0974	$C_{13}H_{15}O_4$	235.0976	0.8
235.1338	$C_{14}H_{19}O_3$	235.1340	0.8
237.0403	$C_{11}H_9O_6$	237.0405	0.8
237.0767	$C_{12}H_{13}O_5$	237.0768	0.8
237.113	$C_{13}H_{17}O_4$	237.1132	0.9
239.0349	$C_{14}H_7O_4$	239.0350	0.5
239.056	$C_{11}H_{11}O_6$	239.0561	0.4
239.0712	$C_{15}H_{11}O_3$	239.0714	0.7
239.0923	$C_{12}H_{15}O_5$	239.0925	0.9
241.014	$C_{13}H_5O_5$	241.0142	0.9

241.0504	$C_{14}H_9O_4$	241.0506	0.8
241.0868	$C_{15}H_{13}O_3$	241.0870	0.7
241.1232	$C_{16}H_{17}O_2$	241.1234	0.7
243.0297	$C_{13}H_7O_5$	243.0299	0.7
243.0661	$C_{14}H_{11}O_4$	243.0663	0.8
243.1025	$C_{15}H_{15}O_3$	243.1027	0.8
243.1389	$C_{16}H_{19}O_2$	243.1391	0.6
245.009	$C_{12}H_5O_6$	245.0092	0.8
245.0818	$C_{14}H_{13}O_4$	245.0819	0.6
245.118	$C_{15}H_{17}O_3$	245.1183	1.1
247.0246	$C_{12}H_7O_6$	247.0248	0.9
247.061	$C_{13}H_{11}O_5$	247.0612	0.8
247.0974	$C_{14}H_{15}O_4$	247.0976	0.8
247.1337	$C_{15}H_{19}O_3$	247.1340	1.0
249.0403	$C_{12}H_9O_6$	249.0405	0.6
249.0767	$C_{13}H_{13}O_5$	249.0768	0.8
249.113	$C_{14}H_{17}O_4$	249.1132	0.9
249.1494	$C_{15}H_{21}O_3$	249.1496	0.7
251.0559	$C_{12}H_{11}O_6$	251.0561	0.7
251.0923	$C_{13}H_{15}O_5$	251.0925	0.9
251.1286	$C_{14}H_{19}O_4$	251.1289	1.0
253.0504	$C_{15}H_9O_4$	253.0506	0.9
253.0717	$C_{12}H_{13}O_6$	253.0718	0.4
253.1079	$C_{13}H_{17}O_5$	253.1081	0.8
255.0297	$C_{14}H_7O_5$	255.0299	0.9
256.0375	$C_{13}H_2N_7$	256.0377	0.8
257.0453	$C_{14}H_9O_5$	257.0455	0.8
257.0817	$C_{15}H_{13}O_4$	257.0819	0.8
259.0246	$C_{13}H_7O_6$	259.0248	0.9
259.061	$C_{14}H_{11}O_5$	259.0612	0.8
259.0974	$C_{15}H_{15}O_4$	259.0976	0.8
259.1338	$C_{16}H_{19}O_3$	259.1340	0.7
259.1702	$C_{17}H_{23}O_2$	259.1704	0.6
260.0563	$C_{13}H_{10}NO_5$	260.0564	0.7

260.0928	$C_{14}H_{14}NO_4$	260.0928	0.3
261.0403	$C_{13}H_9O_6$	261.0405	0.8
261.0767	$C_{14}H_{13}O_5$	261.0768	0.8
261.113	$C_{15}H_{17}O_4$	261.1132	0.9
263.0195	$C_{12}H_7O_7$	263.0197	0.9
263.0559	$C_{13}H_{11}O_6$	263.0561	0.8
263.0923	$C_{14}H_{15}O_5$	263.0925	0.8
263.1286	$C_{15}H_{19}O_4$	263.1289	0.9
265.0352	$C_{12}H_9O_7$	265.0354	0.6
265.0715	$C_{13}H_{13}O_6$	265.0718	0.8
265.1079	$C_{14}H_{17}O_5$	265.1081	0.8
265.1443	$C_{15}H_{21}O_4$	265.1445	0.9
265.1808	$C_{16}H_{25}O_3$	265.1809	0.5
267.0297	$C_{15}H_7O_5$	267.0299	0.7
267.0507	$C_{12}H_{11}O_7$	267.0510	1.3
267.0661	$C_{16}H_{11}O_4$	267.0663	0.6
267.0872	$C_{13}H_{15}O_6$	267.0874	0.8
267.1024	$C_{17}H_{15}O_3$	267.1027	1.0
267.1236	$C_{14}H_{19}O_5$	267.1238	0.8
267.1387	$C_{18}H_{19}O_2$	267.1391	1.2
267.16	$C_{15}H_{23}O_4$	267.1602	0.7
269.1028	$C_{13}H_{17}O_6$	269.1031	0.9
271.0246	$C_{14}H_7O_6$	271.0248	0.8
271.061	$C_{15}H_{11}O_5$	271.0612	0.8
271.0974	$C_{16}H_{15}O_4$	271.0976	0.8
271.1337	$C_{17}H_{19}O_3$	271.1340	0.8
271.1703	$C_{18}H_{23}O_2$	271.1704	0.2
272.0563	$C_{14}H_{10}NO_5$	272.0564	0.7
273.0038	$C_{13}H_5O_7$	273.0041	1.0
273.0402	$C_{14}H_9O_6$	273.0405	0.8
273.0766	$C_{15}H_{13}O_5$	273.0768	0.7
273.1131	$C_{16}H_{17}O_4$	273.1132	0.3
275.0559	$C_{14}H_{11}O_6$	275.0561	0.9
275.0923	$C_{15}H_{15}O_5$	275.0925	0.8

275.1287	$C_{16}H_{19}O_4$	275.1289	0.8
277.0351	$C_{13}H_9O_7$	277.0354	0.9
277.0715	$C_{14}H_{13}O_6$	277.0718	0.8
277.1079	$C_{15}H_{17}O_5$	277.1081	0.8
277.1443	$C_{16}H_{21}O_4$	277.1445	0.7
279.0508	$C_{13}H_{11}O_7$	279.0510	0.8
279.0872	$C_{14}H_{15}O_6$	279.0874	0.8
279.1236	$C_{15}H_{19}O_5$	279.1238	0.8
279.1599	$C_{16}H_{23}O_4$	279.1602	1.0
281.0665	$C_{13}H_{13}O_7$	281.0667	0.8
281.1029	$C_{14}H_{17}O_6$	281.1031	0.7
281.1393	$C_{15}H_{21}O_5$	281.1394	0.5
282.9885	$C_{14}H_3O_7$	282.9884	-0.4
283.0246	$C_{15}H_7O_6$	283.0248	0.7
283.061	$C_{16}H_{11}O_5$	283.0612	0.8
283.082	$C_{13}H_{15}O_7$	283.0823	1.2
283.264	$C_{18}H_{35}O_2$	283.2643	0.8
285.0402	$C_{15}H_9O_6$	285.0405	0.8
285.0766	$C_{16}H_{13}O_5$	285.0768	0.8
285.113	$C_{17}H_{17}O_4$	285.1132	0.9
285.1494	$C_{18}H_{21}O_3$	285.1496	0.6
287.0195	$C_{14}H_7O_7$	287.0197	0.8
287.0559	$C_{15}H_{11}O_6$	287.0561	0.9
287.0922	$C_{16}H_{15}O_5$	287.0925	0.9
287.1287	$C_{17}H_{19}O_4$	287.1289	0.7
289.0351	$C_{14}H_9O_7$	289.0354	0.9
289.0715	$C_{15}H_{13}O_6$	289.0718	0.9
289.1079	$C_{16}H_{17}O_5$	289.1081	0.7
291.0143	$C_{13}H_7O_8$	291.0146	1.0
291.0508	$C_{14}H_{11}O_7$	291.0510	0.8
291.0872	$C_{15}H_{15}O_6$	291.0874	0.7
291.1236	$C_{16}H_{19}O_5$	291.1238	0.8
291.1599	$C_{17}H_{23}O_4$	291.1602	0.9
293.0301	$C_{13}H_9O_8$	293.0303	0.7

293.0664	$C_{14}H_{13}O_7$	293.0667	0.9
293.1028	$C_{15}H_{17}O_6$	293.1031	0.9
293.1392	$C_{16}H_{21}O_5$	293.1394	0.8
293.1756	$C_{17}H_{25}O_4$	293.1758	0.7
295.0457	$C_{13}H_{11}O_8$	295.0459	0.7
295.0821	$C_{14}H_{15}O_7$	295.0823	0.8
295.1184	$C_{15}H_{19}O_6$	295.1187	1.0
297.0402	$C_{16}H_9O_6$	297.0405	0.8
297.0766	$C_{17}H_{13}O_5$	297.0768	0.8
297.0978	$C_{14}H_{17}O_7$	297.0980	0.7
303.0508	$C_{15}H_{11}O_7$	303.0510	0.9
303.0621	$C_{14}H_{11}N_2O_6$	303.0623	0.6
303.0871	$C_{16}H_{15}O_6$	303.0874	0.9
303.1235	$C_{17}H_{19}O_5$	303.1238	0.9
303.1599	$C_{18}H_{23}O_4$	303.1602	0.8
304.046	$C_{14}H_{10}NO_7$	304.0463	0.8
304.0824	$C_{15}H_{14}NO_6$	304.0827	0.8
304.1189	$C_{16}H_{18}NO_5$	304.1190	0.6
305.0664	$C_{15}H_{13}O_7$	305.0667	0.8
305.1028	$C_{16}H_{17}O_6$	305.1031	0.8
309.0613	$C_{14}H_{13}O_8$	309.0616	0.9
309.0977	$C_{15}H_{17}O_7$	309.0980	0.9
309.1341	$C_{16}H_{21}O_6$	309.1344	0.8
309.1705	$C_{17}H_{25}O_5$	309.1707	0.8
311.0558	$C_{17}H_{11}O_6$	311.0561	1.0
311.077	$C_{14}H_{15}O_8$	311.0772	0.9
311.0922	$C_{18}H_{15}O_5$	311.0925	0.9
311.1134	$C_{15}H_{19}O_7$	311.1136	0.9
311.1497	$C_{16}H_{23}O_6$	311.1500	0.9
311.1861	$C_{17}H_{27}O_5$	311.1864	0.8
315.0143	$C_{15}H_7O_8$	315.0146	0.9
315.0508	$C_{16}H_{11}O_7$	315.0510	0.9
315.0872	$C_{17}H_{15}O_6$	315.0874	0.8
315.1235	$C_{18}H_{19}O_5$	315.1238	1.1

315.1598	$C_{19}H_{23}O_4$	315.1602	1.2
317.03	$C_{15}H_9O_8$	317.0303	1.0
317.0663	$C_{16}H_{13}O_7$	317.0667	1.2
317.1029	$C_{17}H_{17}O_6$	317.1031	0.5
319.0457	$C_{15}H_{11}O_8$	319.0459	0.9
319.082	$C_{16}H_{15}O_7$	319.0823	1.1
319.1185	$C_{17}H_{19}O_6$	319.1187	0.7
319.1549	$C_{18}H_{23}O_5$	319.1551	0.7
321.025	$C_{14}H_9O_9$	321.0252	0.6
321.0613	$C_{15}H_{13}O_8$	321.0616	1.0
321.0977	$C_{16}H_{17}O_7$	321.0980	1.0
321.1341	$C_{17}H_{21}O_6$	321.1344	0.9
323.0406	$C_{14}H_{11}O_9$	323.0409	0.8
323.0769	$C_{15}H_{15}O_8$	323.0772	0.9
323.1133	$C_{16}H_{19}O_7$	323.1136	0.9
323.1497	$C_{17}H_{23}O_6$	323.1500	0.9
325.0927	$C_{15}H_{17}O_8$	325.0929	0.7
327.0143	$C_{16}H_7O_8$	327.0146	1.1
327.1083	$C_{15}H_{19}O_8$	327.1085	0.7
329.03	$C_{16}H_9O_8$	329.0303	0.8
329.0664	$C_{17}H_{13}O_7$	329.0667	0.9
329.1028	$C_{18}H_{17}O_6$	329.1031	0.8
329.1392	$C_{19}H_{21}O_5$	329.1394	0.8
331.0456	$C_{16}H_{11}O_8$	331.0459	1.0
335.0404	$C_{15}H_{11}O_9$	335.0409	1.2
335.077	$C_{16}H_{15}O_8$	335.0772	0.6
335.1133	$C_{17}H_{19}O_7$	335.1136	1.0
335.1497	$C_{18}H_{23}O_6$	335.1500	0.8
337.0562	$C_{15}H_{13}O_9$	337.0565	0.8
337.0926	$C_{16}H_{17}O_8$	337.0929	0.8
337.1289	$C_{17}H_{21}O_7$	337.1293	1.0
337.1653	$C_{18}H_{25}O_6$	337.1657	1.1
341.03	$C_{17}H_9O_8$	341.0303	0.9
341.0876	$C_{15}H_{17}O_9$	341.0878	0.7

341.1241	$C_{16}H_{21}O_8$	341.1242	0.3
347.0405	$C_{16}H_{11}O_9$	347.0409	1.0
347.0769	$C_{17}H_{15}O_8$	347.0772	0.9
347.0881	$C_{16}H_{15}N_2O_7$	347.0885	1.0
347.1133	$C_{18}H_{19}O_7$	347.1136	0.9
347.1245	$C_{17}H_{19}N_2O_6$	347.1249	1.1
347.1497	$C_{19}H_{23}O_6$	347.1500	0.9
347.1861	$C_{20}H_{27}O_5$	347.1864	0.9
348.0721	$C_{16}H_{14}NO_8$	348.0725	1.0
348.1085	$C_{17}H_{18}NO_7$	348.1089	1.2
353.051	$C_{15}H_{13}O_{10}$	353.0514	1.1
353.0875	$C_{16}H_{17}O_9$	353.0878	0.9
353.1239	$C_{17}H_{21}O_8$	353.1242	0.9
353.1602	$C_{18}H_{25}O_7$	353.1606	0.9
353.1966	$C_{19}H_{29}O_6$	353.1970	1.0
355.0456	$C_{18}H_{11}O_8$	355.0459	0.8
355.0667	$C_{15}H_{15}O_{10}$	355.0671	1.0
355.0667	$C_{13}H_3N_{14}$	355.0671	0.9
355.082	$C_{19}H_{15}O_7$	355.0823	0.8
355.1031	$C_{16}H_{19}O_9$	355.1035	0.9
355.1395	$C_{17}H_{23}O_8$	355.1398	0.9
355.1759	$C_{18}H_{27}O_7$	355.1762	0.9
355.2123	$C_{19}H_{31}O_6$	355.2126	0.9
355.3214	$C_{22}H_{43}O_3$	355.3218	1.0
356.0985	$C_{15}H_{18}NO_9$	356.0987	0.6
356.1347	$C_{16}H_{22}NO_8$	356.1351	1.1
359.0042	$C_{16}H_7O_{10}$	359.0045	0.8
359.0405	$C_{17}H_{11}O_9$	359.0409	0.9
359.0769	$C_{18}H_{15}O_8$	359.0772	1.0
359.0881	$C_{17}H_{15}N_2O_7$	359.0885	1.1
373.0198	$C_{17}H_9O_{10}$	373.0201	0.9
373.0562	$C_{18}H_{13}O_9$	373.0565	0.9
373.0925	$C_{19}H_{17}O_8$	373.0929	1.0
373.1289	$C_{20}H_{21}O_7$	373.1293	0.9

391.0303	$C_{17}H_{11}O_{11}$	391.0307	1.0
391.0303	$C_{30}H_{3}N_{2}$	391.0302	-0.3
391.0667	$C_{18}H_{15}O_{10}$	391.0671	1.0
391.1031	$C_{19}H_{19}O_{9}$	391.1035	1.0
391.1395	$C_{20}H_{23}O_{8}$	391.1398	1.0
391.1758	$C_{21}H_{27}O_{7}$	391.1762	1.0
391.2123	$C_{22}H_{31}O_{6}$	391.2126	0.9

Appendix 2. 2. Neutral losses considered in the fragmentation analysis of sample PAN-S

Neutral loss	Mass
CH_2	14.01565
CH_3	15.02348
O	15.99492
CH_4	16.03130
H_2O	18.01057
CO	27.99492
$2CH_2$	28.03130
$2CH_3$	30.04695
O_2	31.98983
$2CH_4$	32.06260
$2H_2O$	36.02113
$3CH_2$	42.04695
CO_2	43.98983
$3H_2O$	54.03169
2CO	55.98983
$4CH_2$	56.06260
3CO	83.98474
$2CO_2$	87.97966
4CO	111.97966
$3CO_2$	131.96949
$4CO_2$	175.95932

Appendix 2. 3. Fragmentation channels obtained using data from Appendices 1 and 2. Notice that for each core fragment, all potential neutral loss combinations are considered.

Pathway/Functionalities							Core Fragment	Structural
CH_2	CH_3	O	CH_4	H_2O	CO	CO_2	m/z	isomers
5					1	3	161.0607 $C_{10}H_9O_2$	13
4				1	4	1		
5		2			3	1		
4		1		1	5			
2			1	1	6			
5		1			2	2		
2	2	1			4	1		
	2		1		5	1		
1			2		2	1		
3		1	1		4	7		
5		3			4			
2	2				3	2		
3			1		3	2		
4		1			4	1	163.0763 $C_{10}H_{11}O_2$	7
3				1	6			
4		2			5			
2			1		5	1		
6		1			3	1		
1	2				5	1		
4					3	2		
5			1		5		165.0192 $C_8H_5O_4$	3
7					3	1		
7		1			4			
5					4	1	165.056 $C_9H_9O_3$	2
5		1			5			
6					5		167.0349 $C_8H_7O_4$	1
2				2	4	1	171.0814 $C_{12}H_{11}O$	23
3				1	1	3		
3		2		1	3	1		
2		1		2	5			
3		1		1	2	2		
2		1	1		1	3		
1		1	1	1	4	1		
1			1	1	3	2		
1	2	1			1	3		
		2	2		4	1		
2		2	1		2	2		
1	2	2			2	2		
	2	1		1	4	1		
4		4			2	1		
	2			1	3	2		
2		3	1		3	1		

1	2	3			3	1		
1		2	1	1	5			
		1	2		3	2		
			2		2	3		
	2	2		1	5			
4		5			3			
			1	2	6			
5		1				3		
2			1	1	4	1		
3			1		1	3		
1	2			1	4	1		
2	2				1	3		
4				1	2	2		
4		2		1	4			
			2	1	6			
5		2			1	2		
	2	1	1		4	1		
1		1	2		4	1		
5		3			2	1	173.0607 $C_{11}H_9O_2$	23
5		4			3			
1			2		3	2		
2	2	1			2	2		
3		1	1		2	2		
3		2	1		3	1		
2	2	2			3	1		
1	2	1		1	5			
2		1	1	1	5			
	2		1		3	2		
3				2	5			
4		1		1	3	1		
6						3		
5				1	3	1		
3			1	1	5			
6		1			1	2		
5		1		1	4			
2	2			1	5			
6		2			2	1		
2			2		4	1	175.0400 $C_{10}H_7O_3$	15
1	2		1		4	1		
6		3			3			
4			1		2	2		
3	2				2	2		
	4				4	1		
4		1	1		3	1		
3	2	1			3	1		
2			1	2	3	2	183.0450 $C_{12}H_7O_2$	40
3			1	1		3		

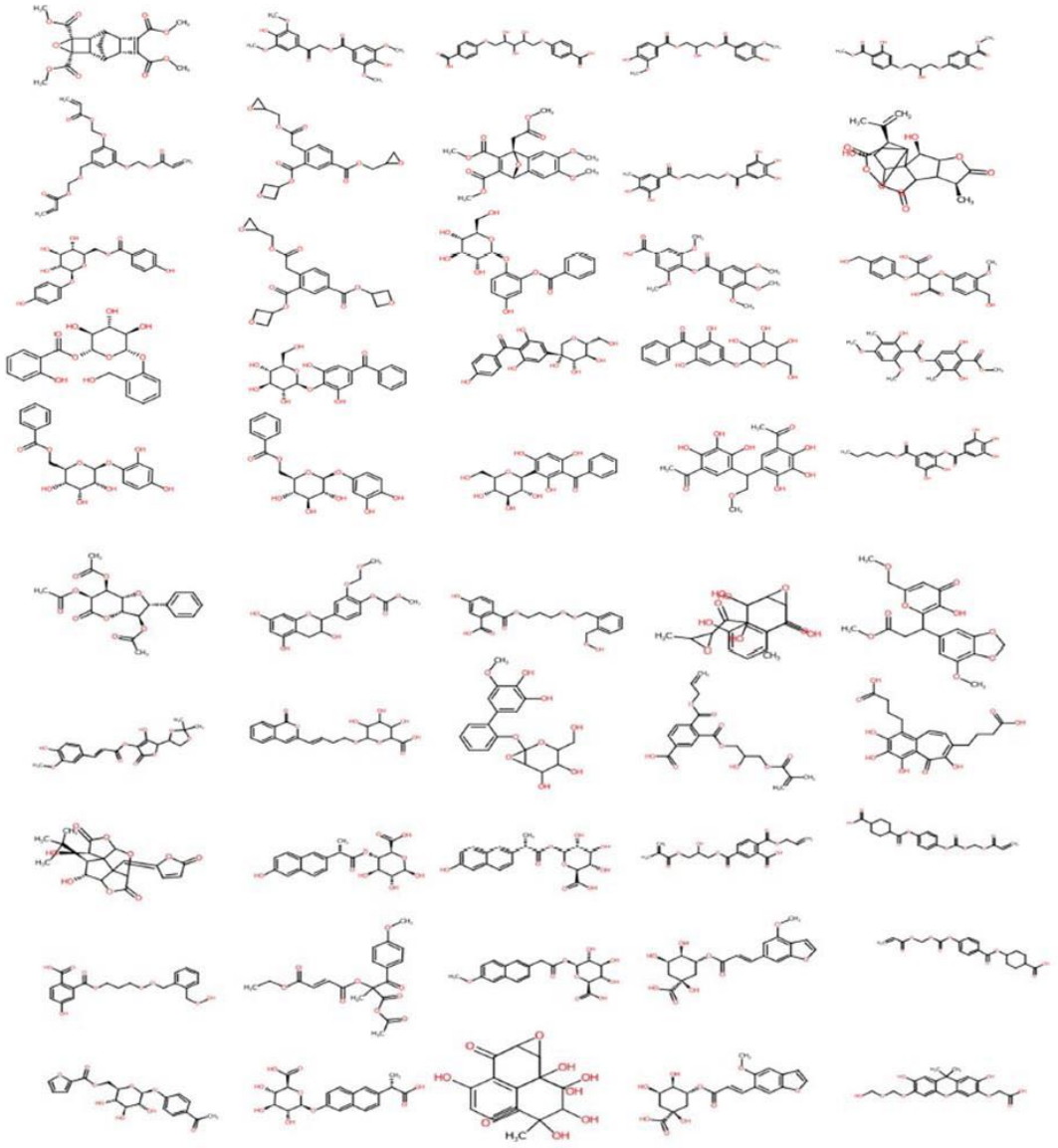
2	2			1		3		
1		1	2	1	3	1		
3		2	1	1	2	1		
1			2	1	2	2		
5		2		1		2		
4		1		2	2	1		
3				3	4			
			3		1	3		
3		1	1	1	1	2		
1	2			2	3	1		
2	2	1		1	1	2		
	2	1	1	1	3	1		
5		3		1	1	1		
6		5				1		
5		4		1	2			
4				2	1	2		
2	2	2		1	2	1		
	2		1	1	2	2		
1		2	2	1	4			
2	2	3		1	3			
3		3	1	1	3			
2		2	2		1	2		
1	2	2	1		1	2		
4		4	1		1	1		
3	2	4			1	1		
2		3	2		2	1		
1	2	3	1		2	1		
4		2		2	3			
2		1	1	2	4			
1	2	1		2	4			
6		6			1			
		2	3		3	1		
		1	3		2	2		
	4	2			1	2		
	2	2	1	1	4			
4		5	1	1	2	1		
			2	2	5			
	4	3			2	1		
1			1	1	1	3	183.0814 $C_{13}H_{11}O$	25
1		2	1	1	3	1		
			1	2	4	1		
2		1		2	3	1		
2				2	2	2		
1				3	5			
3		1		1		3		
	2	1		1	2	2		
	2	2		1	3	1		

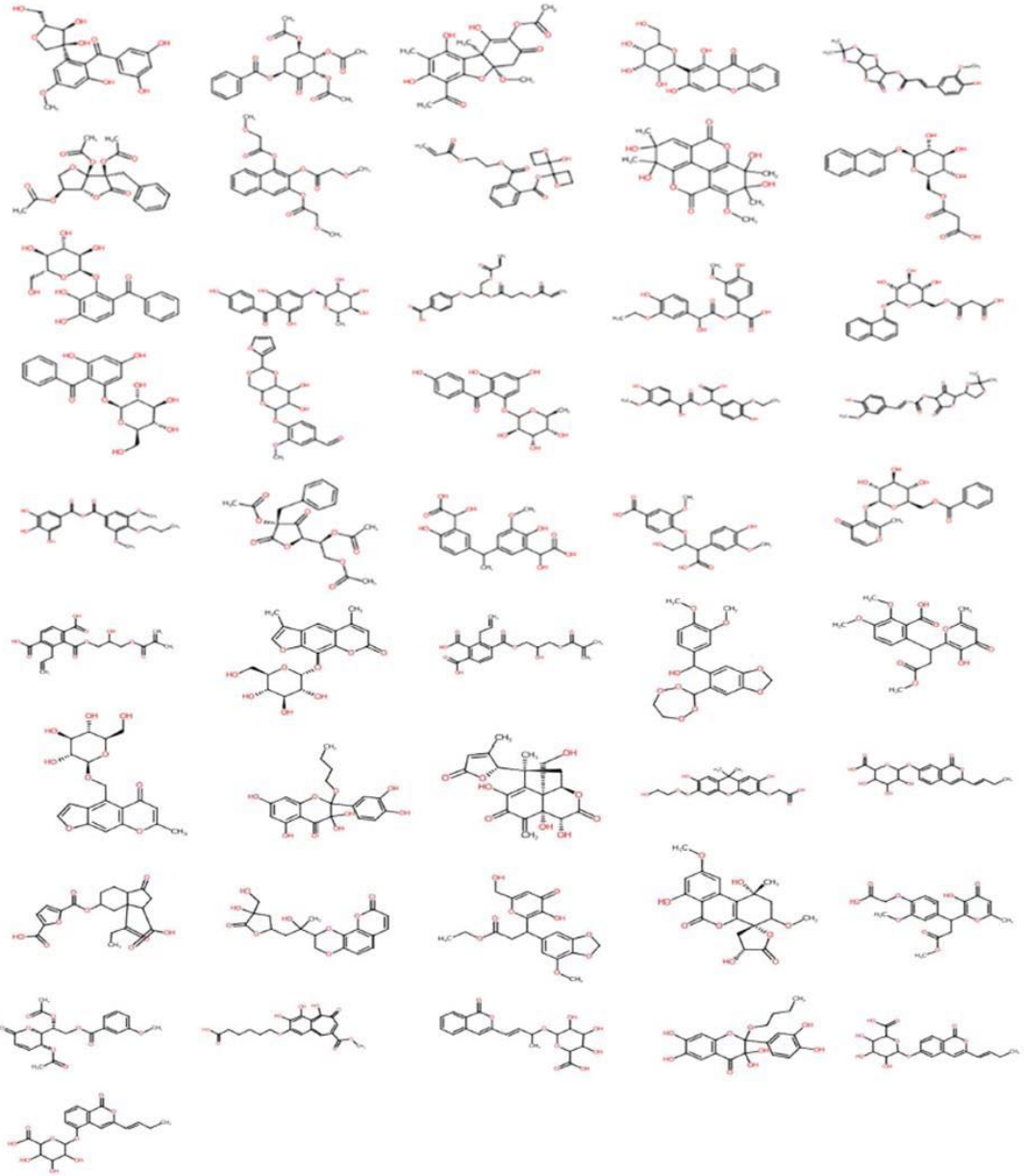
1		3	1	1	4			
		2	2		2	2		
2		3	1		1	2		
3		3		1	2	1		
3		2		1	1	2		
4		5			1	1		
1	2	4			2	1		
2		4	1		2	1		
2		2		2	4			
		3	2		3	1		
1		1	1	1	2	2		
1	2	3			1	2		
	2	3		1	4			
4		6			2			
		1	2		1	3		
		1	1	2	5			
4				1		3		
3				2	3	1		
1	2	1		1	3	1		
4		1		1	1	2		
			2	1	4	1		
1	2			1	2	2		
4		2		1	2	1		
1			2		1	3		
3		1		2	4			
	2		1		1	3		
2			1	1	2	2		
5		5			2			
5		3				2		
1		1	2		2	2		
4		3		1	3		185.0607 $C_{12}H_9O_2$	29
3		2	1		1	2		
5		4			1	1		
3		3	1		2	1		
2	2	3			2	1		
1		2	2		3	1		
1			1	2	5			
2		2	1	1	4			
1	2	2		1	4			
2	2	2			1	2		
2		1	1	1	3	1		
	2	2	1		3	1		
	2	1	1		2	2		
	2			2	5			
		1	2	1	5			
3			1	1	3	1	187.0400 $C_{11}H_7O_3$	25
2	2			1	3	1		

4		1	1		1	2		
4				2	4			
5				1	1	2		
5		1		1	2	1		
5		2		1	3			
3		1	1	1	4			
6		2				2		
			3		4	1		
6		4			2			
1	2		1		2	2		
2			2		2	2		
4		2	1		2	1		
1			2	1	5			
6		3			1	1		
3	2	2			2	1		
2	2	1		1	4			
3	2	1			1	2		
1	2	1	1		3	1		
2		1	2		3	1		
	4				2	2		
	2		1	1	5			
4		3	1		3			
	4	1			3	1		
4			1	1	2	1		
6				1		2		
4		1	1	1	3			
5				2	3			
6		2		1	2			
5		3	1		2			
5		1	1			2		
6		1		1	1	1		
3	2			1	2	1		
2			2	1	4			
3	2	1		1	3			
4	2	2			1	1	201.0192 $C_{11}H_5O_4$	25
5		2	1		1	1		
3			2		1	2		
3		1	2		2	1		
1	2		1	1	4			
7		4			1			
2	2		1		1	2		
1	4	1			2	1		
2	2	1	1		2	1		
1			3		3	1		
	2		2		3	1		
	4			1	4			
4	2	3			2			

1	4				1	2		
5			1	1	3		202.9984 $C_{10}H_3O_5$	15
6			1			2		
7				1	1	1		
8		2				1		
6		2	1		2			
5	2	2			2			
5	2					2		
4			2		2	1		
4	2			1	3			
7		1		1	2			
8		3			1			
5	2	1			1	1		
3	2		1		2	1		
2	4				2	1		
6		1	1		1	1		
6				1	3		205.0140 $C_{10}H_5O_5$	7
7						2		
7		1			1	1		
7		2			2			
4	2				2	1		
5			1		2	1		
5		1	1		3			
5				2		1	241.0140 $C_{13}H_5O_5$	7
2			2	1	1	1		
1	2		1	1	1	1		
4		2	1	1	1			
5		1		2	1			
3			1	2	2			
2	2			2	2			

Appendix 2. 4.Candidate structures generated using *in silico* fragmentation (MetFrag software) of 391.1031*m/z* ($C_{19}H_{19}O_9$) and precursors from the PubChem database.





Appendix 3. 1. Fragmentation data of the precursor ion $[C_{18}H_{18}O_{10}\text{-}H]^-$ from Pantanal sample using the ESI-TIMS-q-FT-ICR-MS/MS (q-CID) procedure.

IMS band	m/z	Error [ppm]	I	Ion Formula
	157.065894	<0.1	63718	C11H9O
	173.060851	-0.3	83989	C11H9O2
	175.040104	-0.2	68935	C10H7O3
	185.060805	<0.1	163144	C12H9O2
	187.040077	-0.1	99444	C11H7O3
	187.076471	-0.1	69982	C12H11O2
	189.055764	-0.2	63018	C11H9O3
	199.040062	<0.1	70186	C12H7O3
	199.076458	<0.1	66613	C13H11O2
	199.112801	0.2	59744	C14H15O
	201.019313	0.1	56105	C11H5O4
	201.055738	-0.1	153414	C12H9O3
I	201.092080	0.1	80240	C13H13O2
	203.034960	0.1	68772	C11H7O4
	203.071402	-0.2	81930	C12H11O3
	211.040053	0.1	79163	C13H7O3
	213.055718	<0.1	63379	C13H9O3
	215.071353	0.1	58164	C13H11O3
	217.050629	<0.1	99845	C12H9O4
	217.087041	-0.1	133246	C13H13O3
	217.123349	0.2	84725	C14H17O2
	219.066292	<0.1	69037	C12H11O4
	225.055668	0.2	54588	C14H9O3
	227.034891	0.4	64373	C13H7O4
	229.050644	-0.1	180699	C13H9O4
	229.086997	0.1	104408	C14H13O3
	229.123439	-0.2	60261	C15H17O2
	231.029929	-0.1	61446	C12H7O5
	231.066251	0.1	68975	C13H11O4
	243.029935	-0.2	66242	C13H7O5
	243.066254	0.1	120814	C14H11O4
	243.102717	-0.2	111831	C15H15O3
	243.139161	-0.4	79340	C16H19O2
	245.045558	<0.1	133561	C13H9O5
	245.081990	-0.2	84381	C14H13O4
	255.029904	<0.1	77431	C14H7O5
	255.066299	-0.1	68158	C15H11O4
	257.045508	0.1	57679	C14H9O5

259.024636	0.7	54863	C13H7O6
259.061188	<0.1	59736	C14H11O5
261.040463	0	73005	C13H9O6
261.076815	0.1	175946	C14H13O5
261.113299	-0.3	95627	C15H17O4
261.149682	-0.2	60838	C16H21O3
273.040514	-0.2	169105	C14H9O6
273.076825	0.1	123358	C15H13O5
275.056023	0.3	73883	C14H11O6
287.056112	<0.1	107654	C15H11O6
287.092485	<0.1	96295	C16H15O5
287.128872	<0.1	84856	C17H19O4
289.035344	0.1	62856	C14H9O7
305.066652	0.1	204885	C15H13O7
305.103047	<0.1	154727	C16H17O6
305.139445	<0.1	117904	C17H21O5
317.030000	0.7	83356	C15H9O8
317.066546	0.4	93977	C16H13O7
331.045977	-0.1	93192	C16H11O8
331.082290	0.1	92134	C17H15O7
331.118651	0.2	86171	C18H19O6
349.056456	0.1	191783	C16H13O9
349.092869	0.1	212960	C17H17O8
349.129304	-0.1	193001	C18H21O7
361.020330	-0.6	57619	C16H9O10
361.056436	0.2	67877	C17H13O9
375.035915	-0.4	136486	C17H11O10
375.072297	-0.4	88791	C18H15O9
393.046333	<0.1	204985	C17H13O11
393.082742	-0.1	357306	C18H17O10
393.119110	<0.1	411035	C19H21O9
393.155560	-0.2	126325	C20H25O8

157.065917	-0.2	76607	C11H9O
159.045176	-0.1	77461	C10H7O2
161.024457	-0.2	68034	C9H5O3
171.045165	-0.1	66434	C11H7O2
173.024399	0.1	60548	C10H5O3
173.060830	-0.2	128200	C11H9O2
175.040078	-0.1	99258	C10H7O3
175.076490	-0.2	59025	C11H11O2

	177.055681	0.2	59502	C10H9O3
	183.029946	-0.3	70739	C8H7O5
	183.045214	-0.3	64167	C12H7O2
	185.024441	-0.1	65372	C11H5O3
	185.060807	<0.1	247767	C12H9O2
	185.097245	-0.3	81715	C13H13O
	187.040080	-0.1	148754	C11H7O3
	187.076475	-0.1	120542	C12H11O2
	189.055765	-0.2	83996	C11H9O3
	191.034987	<0.1	64027	C10H7O4
II	191.071388	-0.1	71733	C11H11O3
	197.060789	0.1	59786	C13H9O2
	199.040094	-0.1	109857	C12H7O3
	199.076444	<0.1	90219	C13H11O2
	199.112812	0.1	61586	C14H15O
	201.019329	<0.1	76574	C11H5O4
	201.055723	<0.1	244272	C12H9O3
	201.092112	<0.1	136685	C13H13O2
	203.034975	<0.1	133510	C11H7O4
	203.071371	<0.1	125035	C12H11O3
	205.014303	-0.3	60826	C10H5O5
	205.050636	<0.1	89949	C11H9O4
	211.040125	-0.3	100352	C13H7O3
	211.076472	-0.1	72823	C14H11O2
	213.019347	-0.1	80863	C12H5O4
	213.055706	0.1	123996	C13H9O3
	215.034967	0.1	79056	C12H7O4
	215.071357	0.1	80877	C13H11O3
	217.014249	<0.1	63783	C11H5O5
	217.050668	-0.2	134586	C12H9O4
	217.087026	<0.1	144215	C13H13O3
	217.123428	-0.1	111660	C14H17O2
	217.159711	0.4	76251	C15H21O
	219.029877	0.1	96650	C11H7O5
	219.066271	<0.1	96078	C12H11O4
	219.102661	<0.1	67907	C13H15O3
	221.045606	-0.3	57213	C11H9O5
	225.055701	0.1	58566	C14H9O3
	227.034966	0.1	81658	C13H7O4
	227.071343	0.1	92577	C14H11O3
	227.107765	-0.1	61906	C15H15O2
	229.014171	0.3	78638	C12H5O5

229.050635	<0.1	331797	C13H9O4
229.087006	0.1	217414	C14H13O3
229.123416	-0.1	205422	C15H17O2
231.029878	0.1	102043	C12H7O5
231.066281	<0.1	123808	C13H11O4
231.102645	0.1	79754	C14H15O3
233.045460	0.4	68882	C12H9O5
233.081837	0.4	64425	C13H13O4
241.014197	0.2	82934	C13H5O5
241.050659	-0.1	105033	C14H9O4
243.029887	<0.1	97405	C13H7O5
243.066303	-0.1	183850	C14H11O4
243.102686	-0.1	128522	C15H15O3
243.139102	-0.2	135643	C16H19O2
245.009146	0.1	70011	C12H5O6
245.045539	<0.1	218578	C13H9O5
245.081952	-0.1	156034	C14H13O4
245.118301	0.1	98375	C15H17O3
247.024807	<0.1	68636	C12H7O6
247.061223	-0.1	78352	C13H11O5
255.029868	0.1	104248	C14H7O5
255.066296	-0.1	122428	C15H11O4
257.009217	-0.2	72049	C13H5O6
257.04546	0.3	78742	C14H9O5
259.024797	0.1	71728	C13H7O6
259.06116	0.1	111473	C14H11O5
259.097371	0.8	59939	C15H15O4
259.133985	-0.1	63827	C16H19O3
261.040438	0.1	109195	C13H9O6
261.055743	-0.1	83647	C17H9O3
261.07684	<0.1	227432	C14H13O5
261.113257	-0.1	153412	C15H17O4
261.14965	-0.1	111189	C16H21O3
263.092585	-0.3	58856	C14H15O5
269.045485	0.2	69499	C15H9O5
269.118428	-0.4	74594	C17H17O3
273.040478	-0.1	320153	C14H9O6
273.076857	<0.1	330842	C15H13O5
273.113255	-0.1	120567	C16H17O4
275.01968	0.2	59612	C13H7O7
275.056135	-0.1	102139	C14H11O6
275.092453	0.2	62868	C15H15O5

285.040504	-0.1	64605	C15H9O6
287.019556	0.6	69341	C14H7O7
287.056082	0.1	138598	C15H11O6
287.092488	<0.1	139087	C16H15O5
287.128890	<0.1	166856	C17H19O4
287.165267	<0.1	77106	C18H23O3
289.035379	<0.1	92119	C14H9O7
289.071782	-0.1	88176	C15H13O6
289.108145	<0.1	68587	C16H17O5
299.019696	0.1	63404	C15H7O7
299.056148	-0.1	84560	C16H11O6
305.045469	0.3	102680	C18H9O5
305.066670	<0.1	315113	C15H13O7
305.103062	<0.1	261962	C16H17O6
305.139450	<0.1	234476	C17H21O5
305.175842	<0.1	88905	C18H25O4
317.030215	0.2	116408	C15H9O8
317.066626	0.2	197269	C16H13O7
317.102930	0.4	64502	C17H17O6
331.045883	0.2	142820	C16H11O8
331.082290	0.1	119957	C17H15O7
331.118699	<0.1	138798	C18H19O6
331.155064	0.1	78712	C19H23O5
349.035299	0.2	101241	C19H9O7
349.056481	0.1	265378	C16H13O9
349.092862	0.1	333747	C17H17O8
349.129289	<0.1	371652	C18H21O7
349.165689	-0.1	161141	C19H25O6
361.056496	<0.1	138663	C17H13O9
361.092875	<0.1	64461	C18H17O8
375.035885	-0.3	175756	C17H11O10
375.072142	<0.1	127165	C18H15O9
375.108550	<0.1	78534	C19H19O8
393.046338	<0.1	198413	C17H13O11
393.082722	<0.1	506335	C18H17O10
393.119136	-0.1	791104	C19H21O9
393.155533	-0.1	430785	C20H25O8

135.045183	-0.2	50094	C8H7O2
143.050282	-0.3	49620	C10H7O
145.029528	-0.2	57369	C9H5O2

	147.045194	-0.3	65931	C9H7O2
	149.060887	-0.6	58520	C9H9O2
	157.065921	-0.2	85432	C11H9O
	159.045196	-0.3	82646	C10H7O2
	161.024458	-0.3	72686	C9H5O3
	161.060815	-0.1	66070	C10H9O2
	169.065913	-0.1	63798	C12H9O
	171.045180	-0.2	95567	C11H7O2
	173.024415	<0.1	65050	C10H5O3
	173.060812	-0.1	143007	C11H9O2
	173.097242	-0.3	55925	C12H13O
	175.040091	-0.1	118450	C10H7O3
	175.076492	-0.2	72500	C11H11O2
	177.019341	<0.1	61603	C9H5O4
	177.055747	-0.2	61285	C10H9O3
	177.092058	0.3	54368	C11H13O2
III	183.029959	-0.3	87895	C8H7O5
	183.045169	-0.1	65014	C12H7O2
	183.066341	-0.3	48866	C9H11O4
	183.081527	0.1	67211	C13H11O
	185.024427	-0.1	83682	C11H5O3
	185.060813	-0.1	216720	C12H9O2
	185.097217	-0.2	114564	C13H13O
	186.068648	-0.1	66294	C12H10O2
	187.040084	-0.1	167879	C11H7O3
	187.076485	-0.2	171566	C12H11O2
	187.112814	0.1	66195	C13H15O
	189.019315	0.1	72925	C10H5O4
	189.055736	-0.1	111763	C11H9O3
	191.034991	<0.1	102238	C10H7O4
	191.071375	<0.1	78409	C11H11O3
	193.050715	-0.4	58485	C10H9O4
	197.024497	-0.4	58659	C12H5O3
	197.060805	<0.1	151542	C13H9O2
	199.040087	-0.1	176531	C12H7O3
	199.076474	-0.1	109738	C13H11O2
	199.112888	-0.2	59793	C14H15O
	201.019329	<0.1	87283	C11H5O4
	201.055721	<0.1	289332	C12H9O3
	201.092124	-0.1	172582	C13H13O2
	201.128571	-0.4	63711	C14H17O
	203.034976	<0.1	205855	C11H7O4

203.071378	-0.1	167145	C12H11O3
203.107766	-0.1	87758	C13H15O2
205.014242	<0.1	94862	C10H5O5
205.050600	0.2	131279	C11H9O4
205.087054	-0.2	76242	C12H13O3
211.040095	-0.1	106899	C13H7O3
211.076448	<0.1	138868	C14H11O2
211.112900	-0.3	53568	C15H15O
213.019328	<0.1	95711	C12H5O4
213.055711	<0.1	215739	C13H9O3
213.092103	<0.1	77896	C14H13O2
214.027140	0.1	87608	C12H6O4
215.034992	<0.1	120229	C12H7O4
215.071349	0.1	114727	C13H11O3
215.107707	0.2	62057	C14H15O2
215.144175	-0.2	51929	C15H19O
217.014306	-0.3	72453	C11H5O5
217.050664	-0.1	200686	C12H9O4
217.087001	0.1	123851	C13H13O3
217.123469	-0.3	91176	C14H17O2
217.159695	0.4	89545	C15H21O
219.029853	0.2	135127	C11H7O5
219.066262	0.1	129652	C12H11O4
219.102616	0.2	86924	C13H15O3
221.045479	0.3	69391	C11H9O5
223.040052	0.1	66589	C14H7O3
225.055710	<0.1	71522	C14H9O3
227.034959	0.1	80624	C13H7O4
227.071358	<0.1	136753	C14H11O3
227.107727	0.1	90190	C15H15O2
229.014171	0.3	94341	C12H5O5
229.050635	<0.1	365757	C13H9O4
229.087022	<0.1	365294	C14H13O3
229.123410	<0.1	343103	C15H17O2
231.029878	0.1	137388	C12H7O5
231.066303	-0.1	187788	C13H11O4
231.102661	<0.1	128489	C14H15O3
231.139008	0.2	55935	C15H19O2
233.045503	0.2	81349	C12H9O5
233.081877	0.2	75540	C13H13O4
233.118241	0.3	50616	C14H17O3
235.097591	<0.1	56160	C13H15O4

239.034947	0.1	56642	C14H7O4
241.014227	0.1	113726	C13H5O5
241.050640	<0.1	257931	C14H9O4
241.086976	0.2	86095	C15H13O3
243.029881	0.1	146863	C13H7O5
243.066301	-0.1	201155	C14H11O4
243.102677	<0.1	144820	C15H15O3
243.139079	-0.1	163464	C16H19O2
243.175427	<0.1	59187	C17H23O
245.009124	0.2	70988	C12H5O6
245.045557	<0.1	278683	C13H9O5
245.081950	-0.1	245612	C14H13O4
245.118278	0.2	164548	C15H17O3
247.024801	<0.1	100999	C12H7O6
247.061195	<0.1	123719	C13H11O5
247.097548	0.1	72072	C14H15O4
249.040447	0.1	61113	C12H9O6
249.055827	-0.4	55409	C16H9O3
249.076845	<0.1	64478	C13H13O5
253.050646	-0.1	56128	C15H9O4
255.029882	0.1	104570	C14H7O5
255.066313	-0.1	213766	C15H11O4
255.102576	0.4	95215	C16H15O3
257.009198	-0.1	88020	C13H5O6
257.045512	0.1	162400	C14H9O5
257.081850	0.3	90925	C15H13O4
257.118232	0.3	55902	C16H17O3
259.024789	0.1	79740	C13H7O6
259.061196	<0.1	126072	C14H11O5
259.097529	0.2	71065	C15H15O4
259.133950	0.1	82750	C16H19O3
261.040481	-0.1	118745	C13H9O6
261.055727	<0.1	115509	C17H9O3
261.113267	-0.1	214106	C15H17O4
261.149613	<0.1	152235	C16H21O3
261.185963	0.2	81314	C17H25O2
263.056184	-0.3	68104	C13H11O6
263.092521	-0.1	71963	C14H15O5
269.045443	0.4	60606	C15H9O5
269.082047	-0.4	59002	C16H13O4
269.118374	-0.2	82779	C17H17O3
269.154648	0.2	56010	C18H21O2

271.024968	-0.6	70176	C14H7O6
271.061169	0.1	91331	C15H11O5
271.097739	-0.6	65107	C16H15O4
273.040491	-0.1	411591	C14H9O6
273.076869	-0.1	657944	C15H13O5
273.113235	<0.1	275113	C16H17O4
273.149576	0.2	96202	C17H21O3
275.019773	-0.2	96669	C13H7O7
275.056114	<0.1	159194	C14H11O6
275.092506	<0.1	88563	C15H15O5
275.128902	-0.1	73588	C16H19O4
277.050665	-0.1	82795	C17H9O4
277.071616	0.5	66431	C14H13O6
277.108260	-0.4	55990	C15H17O5
277.144584	-0.2	59389	C16H21O4
285.004133	-0.2	62160	C14H5O7
285.040451	<0.1	129803	C15H9O6
287.019447	1	72037	C14H7O7
287.056142	-0.1	155128	C15H11O6
287.092476	0.1	164505	C16H15O5
287.128874	<0.1	217866	C17H19O4
287.165265	<0.1	167467	C18H23O3
289.035429	-0.2	117802	C14H9O7
289.071763	<0.1	151407	C15H13O6
289.108132	0.1	95472	C16H17O5
289.144489	0.2	70269	C17H21O4
291.050980	0.2	68828	C14H11O7
299.019803	-0.3	78375	C15H7O7
299.056097	0.1	115472	C16H11O6
299.092364	0.4	79349	C17H15O5
299.129028	-0.5	61761	C18H19O4
303.050932	0.3	83731	C15H11O7
303.087337	0.2	68388	C16H15O6
303.123605	0.6	68569	C17H19O5
305.045525	0.1	183977	C18H9O5
305.066652	0.1	342138	C15H13O7
305.081942	<0.1	67314	C19H13O4
305.103078	-0.1	460859	C16H17O6
305.139418	0.1	338029	C17H21O5
305.175813	0.1	216742	C18H25O4
313.144364	0.5	79030	C19H21O4
317.030245	0.1	155946	C15H9O8

317.066677	<0.1	411612	C16H13O7
317.103095	-0.1	200749	C17H17O6
317.139455	<0.1	92070	C18H21O5
331.045889	0.2	150467	C16H11O8
331.082342	<0.1	145828	C17H15O7
331.118697	<0.1	149109	C18H19O6
331.155085	<0.1	215687	C19H23O5
333.025132	0.2	62096	C15H9O9
333.061795	-0.6	73807	C16H13O8
343.045997	-0.2	69775	C17H11O8
347.077170	0.2	60689	C17H15O8
349.035251	0.4	172122	C19H9O7
349.056466	0.1	246435	C16H13O9
349.071726	0.1	75211	C20H13O6
349.092889	<0.1	532196	C17H17O8
349.129300	-0.1	525061	C18H21O7
349.165684	-0.1	519815	C19H25O6
349.201935	0.3	63826	C20H29O5
361.020084	0.1	63666	C16H9O10
361.056508	<0.1	228379	C17H13O9
361.092849	0.1	143621	C18H17O8
361.129225	0.1	70854	C19H21O7
375.035823	-0.1	201976	C17H11O10
375.072073	0.2	196345	C18H15O9
375.108512	0.1	106234	C19H19O8
375.144850	0.2	96664	C20H23O7
393.025170	0.1	110114	C20H9O9
393.040426	0.1	94067	C24H9O6
393.046280	0.1	121451	C17H13O11
393.082750	-0.1	682932	C18H17O10
393.119156	-0.1	964598	C19H21O9
393.155538	-0.1	1032887	C20H25O8
393.191894	<0.1	169704	C21H29O7

157.065907	-0.1	78100	C11H9O
159.045180	-0.2	85056	C10H7O2
161.024486	-0.4	63552	C9H5O3
161.060817	-0.1	72839	C10H9O2
169.065921	-0.2	76357	C12H9O
171.045171	-0.1	104108	C11H7O2
171.081546	<0.1	57973	C12H11O

	173.024429	-0.1	71781	C10H5O3
	173.060818	-0.1	138899	C11H9O2
	175.040086	-0.1	130691	C10H7O3
	175.076504	-0.3	75860	C11H11O2
	183.029986	-0.5	74121	C8H7O5
	183.081555	-0.1	70571	C13H11O
	185.024424	<0.1	84564	C11H5O3
	185.060826	-0.1	187602	C12H9O2
	185.097202	-0.1	118700	C13H13O
	187.040077	<0.1	186294	C11H7O3
	187.076488	-0.2	179208	C12H11O2
	187.112820	0.1	68692	C13H15O
	189.019336	<0.1	80916	C10H5O4
	189.055733	-0.1	118110	C11H9O3
	189.092086	0.1	64794	C12H13O2
	191.034997	-0.1	101285	C10H7O4
	191.071332	0.2	72473	C11H11O3
	193.050685	-0.3	62613	C10H9O4
	197.060805	<0.1	189428	C13H9O2
	199.040082	-0.1	192369	C12H7O3
IV	199.076481	-0.1	104332	C13H11O2
	201.019324	<0.1	97967	C11H5O4
	201.055730	-0.1	258578	C12H9O3
	201.092125	-0.1	167064	C13H13O2
	201.128583	-0.5	67621	C14H17O
	203.034970	0.1	212894	C11H7O4
	203.071364	<0.1	175763	C12H11O3
	203.107749	<0.1	94079	C13H15O2
	205.014245	<0.1	98482	C10H5O5
	205.050616	0.1	138722	C11H9O4
	205.087045	-0.1	80849	C12H13O3
	211.040051	0.1	112079	C13H7O3
	211.076458	<0.1	148252	C14H11O2
	211.112877	-0.2	63173	C15H15O
	213.019322	<0.1	103007	C12H5O4
	213.055707	<0.1	258554	C13H9O3
	213.092117	-0.1	76893	C14H13O2
	215.034975	<0.1	122013	C12H7O4
	215.071360	<0.1	127182	C13H11O3
	215.107754	<0.1	66268	C14H15O2
	217.014277	-0.1	71108	C11H5O5
	217.050646	-0.1	208590	C12H9O4

217.086999	0.1	126371	C13H13O3
217.123446	-0.2	79372	C14H17O2
217.159701	0.4	72635	C15H21O
219.029852	0.2	130035	C11H7O5
219.066261	0.1	133260	C12H11O4
219.102643	0.1	94071	C13H15O3
221.045470	0.3	70687	C11H9O5
221.081975	-0.2	59658	C12H13O4
223.040057	<0.1	69853	C14H7O3
225.055705	0.1	81438	C14H9O3
227.034951	0.1	72003	C13H7O4
227.071359	<0.1	119114	C14H11O3
227.107743	<0.1	92698	C15H15O2
229.014202	0.2	96149	C12H5O5
229.050644	-0.1	332086	C13H9O4
229.087024	<0.1	398439	C14H13O3
229.123405	<0.1	325439	C15H17O2
231.029890	<0.1	133637	C12H7O5
231.066295	-0.1	212933	C13H11O4
231.102668	<0.1	129058	C14H15O3
233.045519	0.1	85776	C12H9O5
233.081892	0.2	77825	C13H13O4
241.014268	-0.1	99732	C13H5O5
241.050640	<0.1	324125	C14H9O4
241.087031	-0.1	106697	C15H13O3
243.029861	0.1	162269	C13H7O5
243.066302	-0.1	191542	C14H11O4
243.102709	-0.2	125682	C15H15O3
243.139082	-0.1	151907	C16H19O2
243.175451	-0.1	67907	C17H23O
245.045542	<0.1	270590	C13H9O5
245.081961	-0.1	269662	C14H13O4
245.118280	0.2	166567	C15H17O3
245.154632	0.3	65005	C16H21O2
247.024782	0.1	110778	C12H7O6
247.061198	<0.1	124842	C13H11O5
247.097570	0.1	81511	C14H15O4
249.040425	0.1	68524	C12H9O6
249.076902	-0.2	72317	C13H13O5
253.050618	0.1	62623	C15H9O4
255.029878	0.1	97251	C14H7O5
255.066293	<0.1	235524	C15H11O4

255.102633	0.1	119972	C16H15O3
257.009177	-0.1	82902	C13H5O6
257.045530	0.1	189670	C14H9O5
257.081878	0.2	96339	C15H13O4
259.024807	<0.1	81581	C13H7O6
259.061203	<0.1	131835	C14H11O5
259.097573	<0.1	72604	C15H15O4
259.133971	<0.1	86421	C16H19O3
261.040476	-0.1	121917	C13H9O6
261.055759	-0.2	119941	C17H9O3
261.076862	-0.1	185282	C14H13O5
261.113257	-0.1	219162	C15H17O4
261.149607	<0.1	151531	C16H21O3
261.185958	0.2	96520	C17H25O2
263.056174	-0.2	81697	C13H11O6
263.092501	<0.1	66786	C14H15O5
269.045642	-0.4	67742	C15H9O5
269.082051	-0.4	65132	C16H13O4
269.118321	<0.1	73343	C17H17O3
269.154633	0.3	68278	C18H21O2
271.024899	-0.3	78265	C14H7O6
271.061190	<0.1	96472	C15H11O5
271.097651	-0.3	76829	C16H15O4
271.133904	0.2	61896	C17H19O3
273.040507	-0.2	381262	C14H9O6
273.076866	-0.1	729935	C15H13O5
273.113228	<0.1	315648	C16H17O4
273.149602	0.1	142849	C17H21O3
275.019749	-0.1	103114	C13H7O7
275.056095	0.1	175299	C14H11O6
275.092490	<0.1	113232	C15H15O5
275.128914	-0.1	84625	C16H19O4
277.050673	-0.1	79057	C17H9O4
277.071631	0.5	60793	C14H13O6
277.144605	-0.3	62991	C16H21O4
285.004029	0.2	61346	C14H5O7
285.040432	0.1	173835	C15H9O6
285.076890	-0.2	73150	C16H13O5
287.056119	<0.1	141330	C15H11O6
287.092519	-0.1	161123	C16H15O5
287.128864	0.1	194301	C17H19O4
287.165261	<0.1	207742	C18H23O3

289.035392	-0.1	122130	C14H9O7
289.071759	<0.1	172219	C15H13O6
289.108114	0.1	102884	C16H17O5
289.144452	0.3	87573	C17H21O4
291.051039	<0.1	75663	C14H11O7
299.019794	-0.2	69256	C15H7O7
299.056063	0.2	118108	C16H11O6
299.092446	0.2	92001	C17H15O5
299.128986	-0.3	74258	C18H19O4
301.035285	0.3	67185	C15H9O7
301.071812	-0.2	65123	C16H13O6
303.050983	0.1	86877	C15H11O7
303.087350	0.2	71366	C16H15O6
303.123646	0.5	66206	C17H19O5
303.160037	0.5	60387	C18H23O4
305.045529	0.1	197128	C18H9O5
305.066637	0.1	312857	C15H13O7
305.081964	-0.1	79618	C19H13O4
305.103086	-0.1	487787	C16H17O6
305.139429	0.1	325900	C17H21O5
305.175820	<0.1	274717	C18H25O4
313.071893	-0.4	60499	C17H13O6
313.144422	0.4	78285	C19H21O4
317.030245	0.1	142174	C15H9O8
317.066698	-0.1	492736	C16H13O7
317.103106	-0.1	259985	C17H17O6
317.139454	<0.1	128474	C18H21O5
319.009551	<0.1	86767	C14H7O9
319.045941	<0.1	70688	C15H11O8
331.045930	<0.1	135156	C16H11O8
331.082345	-0.1	156637	C17H15O7
331.118704	<0.1	140070	C18H19O6
331.155094	<0.1	253792	C19H23O5
333.025154	0.2	63708	C15H9O9
333.061684	-0.3	77279	C16H13O8
343.009600	-0.1	60732	C16H7O9
343.045996	-0.2	76241	C17H11O8
347.077198	0.1	66565	C17H15O8
349.035263	0.3	194687	C19H9O7
349.056427	0.2	209584	C16H13O9
349.071773	<0.1	105061	C20H13O6
349.092892	<0.1	599361	C17H17O8

	349.129299	-0.1	518051	C18H21O7
	349.165681	-0.1	645604	C19H25O6
	349.201985	0.2	107017	C20H29O5
	361.020213	-0.3	71158	C16H9O10
	361.056495	<0.1	253878	C17H13O9
	361.092857	0.1	194575	C18H17O8
	361.129145	0.4	95284	C19H21O7
	365.030099	0.5	66894	C19H9O8
	375.035792	-0.1	206651	C17H11O10
	375.072089	0.2	228412	C18H15O9
	375.108471	0.2	113262	C19H19O8
	375.144893	0.1	113103	C20H23O7
	393.025208	<0.1	130130	C20H9O9
	393.040419	0.1	117739	C24H9O6
	393.046062	0.7	106679	C17H13O11
	393.082767	-0.1	678004	C18H17O10
	393.119150	-0.1	935925	C19H21O9
	393.155531	-0.1	1194103	C20H25O8
	393.191926	-0.1	295560	C21H29O7
	145.029488	0.1	55456	C9H5O2
	151.040012	0.4	60959	C8H7O3
	157.065864	0.2	62780	C11H9O
	159.045167	-0.1	78080	C10H7O2
	161.024455	-0.2	70205	C9H5O3
	161.060793	0.1	66315	C10H9O2
	169.065921	-0.2	64396	C12H9O
	171.045145	<0.1	84263	C11H7O2
	171.081533	<0.1	54253	C12H11O
	173.024468	-0.3	58957	C10H5O3
	173.060829	-0.2	108206	C11H9O2
V	175.040088	-0.1	110287	C10H7O3
	175.076490	-0.2	67421	C11H11O2
	183.045180	-0.1	63536	C12H7O2
	183.081590	-0.3	58539	C13H11O
	185.024378	0.2	63813	C11H5O3
	185.060838	-0.2	128091	C12H9O2
	185.097180	<0.1	93249	C13H13O
	187.040076	<0.1	146383	C11H7O3
	187.076490	-0.2	129001	C12H11O2
	187.112894	-0.3	58799	C13H15O

189.019364	-0.2	59378	C10H5O4
189.055731	-0.1	107467	C11H9O3
189.092089	0.1	70997	C12H13O2
191.034993	-0.1	92483	C10H7O4
191.071308	0.3	70601	C11H11O3
193.014270	-0.1	59409	C9H5O5
197.060801	<0.1	225188	C13H9O2
199.040063	<0.1	165743	C12H7O3
199.076469	-0.1	95523	C13H11O2
201.019318	0.1	90480	C11H5O4
201.055730	-0.1	180467	C12H9O3
201.092122	-0.1	112551	C13H13O2
201.128493	<0.1	69989	C14H17O
203.034983	<0.1	161317	C11H7O4
203.071344	0.1	155395	C12H11O3
203.107735	0.1	102974	C13H15O2
205.014242	<0.1	73053	C10H5O5
205.050615	0.1	113268	C11H9O4
205.087002	0.1	76458	C12H13O3
205.123361	0.2	57445	C13H17O2
211.040076	<0.1	69099	C13H7O3
211.076456	<0.1	89148	C14H11O2
213.019340	<0.1	92252	C12H5O4
213.055724	<0.1	221568	C13H9O3
213.092113	<0.1	90261	C14H13O2
215.034995	-0.1	96694	C12H7O4
215.071400	-0.1	104535	C13H11O3
215.107712	0.2	64028	C14H15O2
217.014172	0.3	66621	C11H5O5
217.050644	-0.1	159557	C12H9O4
217.087019	<0.1	96898	C13H13O3
217.123433	-0.1	76279	C14H17O2
219.029870	0.1	98481	C11H7O5
219.066266	0.1	130350	C12H11O4
219.102632	0.2	69407	C13H15O3
221.045518	0.1	63869	C11H9O5
221.081932	<0.1	65467	C12H13O4
223.040095	-0.1	62558	C14H7O3
225.055721	<0.1	87339	C14H9O3
227.034919	0.3	79333	C13H7O4
227.071378	<0.1	81912	C14H11O3
227.107764	<0.1	70646	C15H15O2

229.014287	-0.2	70089	C12H5O5
229.050625	<0.1	203574	C13H9O4
229.087008	<0.1	309222	C14H13O3
229.123409	<0.1	177023	C15H17O2
229.159771	0.1	82498	C16H21O
231.029902	<0.1	95600	C12H7O5
231.066291	<0.1	181357	C13H11O4
231.102660	<0.1	122658	C14H15O3
231.139075	-0.1	62186	C15H19O2
233.045539	<0.1	96606	C12H9O5
233.081933	<0.1	72746	C13H13O4
233.118326	<0.1	59685	C14H17O3
241.014294	-0.2	73295	C13H5O5
241.050631	<0.1	403861	C14H9O4
241.087012	<0.1	168270	C15H13O3
241.123434	-0.1	66930	C16H17O2
243.029911	-0.1	130118	C13H7O5
243.066231	0.2	185527	C14H11O4
243.102727	-0.2	116127	C15H15O3
243.139110	-0.2	90004	C16H19O2
243.175460	-0.1	62151	C17H23O
245.045554	<0.1	162744	C13H9O5
245.081939	<0.1	229056	C14H13O4
245.118317	<0.1	153281	C15H17O3
245.154733	-0.1	93861	C16H21O2
247.024801	<0.1	77270	C12H7O6
247.061182	0.1	130385	C13H11O5
247.097537	0.2	80062	C14H15O4
247.133948	0.1	56484	C15H19O3
249.040428	0.1	79108	C12H9O6
249.076969	-0.5	72055	C13H13O5
249.113057	0.7	61646	C14H17O4
255.029893	<0.1	83043	C14H7O5
255.066272	<0.1	185204	C15H11O4
255.102648	0.1	123708	C16H15O3
255.139060	<0.1	119489	C17H19O2
257.009187	-0.1	81081	C13H5O6
257.045532	0.1	186336	C14H9O5
257.081971	-0.1	104525	C15H13O4
257.118256	0.2	87702	C16H17O3
258.053383	<0.1	86035	C14H10O5
259.024860	-0.2	83156	C13H7O6

259.061240	-0.2	109187	C14H11O5
259.097602	-0.1	76982	C15H15O4
259.133954	0.1	75112	C16H19O3
259.170487	-0.5	57163	C17H23O2
261.040480	-0.1	94390	C13H9O6
261.055789	-0.3	87219	C17H9O3
261.076875	-0.1	114495	C14H13O5
261.113232	<0.1	129832	C15H17O4
261.149589	0.1	93149	C16H21O3
261.185915	0.3	76649	C17H25O2
263.056135	-0.1	74449	C13H11O6
263.128870	<0.1	72356	C15H19O4
269.154796	-0.3	66528	C18H21O2
271.024790	0.1	67758	C14H7O6
271.061170	0.1	100711	C15H11O5
271.097665	-0.3	71604	C16H15O4
271.133924	0.2	75849	C17H19O3
273.040478	-0.1	201812	C14H9O6
273.076861	-0.1	652405	C15H13O5
273.113240	<0.1	322006	C16H17O4
273.149624	<0.1	263863	C17H21O3
275.019731	<0.1	68089	C13H7O7
275.056094	0.1	187556	C14H11O6
275.092543	-0.2	140621	C15H15O5
275.128870	<0.1	96106	C16H19O4
277.050574	0.2	85205	C17H9O4
285.040441	0.1	198382	C15H9O6
285.076867	-0.1	142111	C16H13O5
285.113281	-0.2	59963	C17H17O4
287.056089	0.1	103585	C15H11O6
287.092576	-0.3	115046	C16H15O5
287.128884	<0.1	109615	C17H19O4
287.165290	-0.1	173204	C18H23O3
289.035337	0.1	97268	C14H9O7
289.071778	-0.1	194173	C15H13O6
289.108136	<0.1	98118	C16H17O5
289.144527	<0.1	111483	C17H21O4
291.050967	0.2	69887	C14H11O7
299.056058	0.2	108948	C16H11O6
299.092489	<0.1	89221	C17H15O5
299.128945	-0.2	100799	C18H19O4
299.165255	<0.1	70606	C19H23O3

303.051016	<0.1	78460	C15H11O7
305.045531	0.1	164454	C18H9O5
305.066677	<0.1	142471	C15H13O7
305.081966	-0.1	141928	C19H13O4
305.103099	-0.1	352521	C16H17O6
305.139423	0.1	257706	C17H21O5
305.175803	0.1	307755	C18H25O4
305.212235	-0.1	72085	C19H29O3
317.030196	0.3	92447	C15H9O8
317.045588	-0.1	80735	C19H9O5
317.066689	<0.1	477516	C16H13O7
317.103093	-0.1	348813	C17H17O6
317.139439	<0.1	248687	C18H21O5
317.175718	0.4	73362	C19H25O4
329.030099	0.6	70915	C16H9O8
331.045981	-0.1	78013	C16H11O8
331.082359	-0.1	121390	C17H15O7
331.118667	0.1	117483	C18H19O6
331.155099	<0.1	264672	C19H23O5
331.191486	<0.1	119346	C20H27O4
333.061595	<0.1	71799	C16H13O8
343.045793	0.4	81951	C17H11O8
349.035282	0.3	152148	C19H9O7
349.056329	0.5	115196	C16H13O9
349.071791	-0.1	179509	C20H13O6
349.092885	<0.1	515774	C17H17O8
349.129257	0.1	450719	C18H21O7
349.165695	-0.1	690912	C19H25O6
349.202043	<0.1	309121	C20H29O5
361.056492	<0.1	221416	C17H13O9
361.092889	<0.1	249097	C18H17O8
361.129242	0.1	169890	C19H21O7
375.035795	-0.1	128690	C17H11O10
375.072123	0.1	192172	C18H15O9
375.108586	-0.1	99349	C19H19O8
375.144855	0.2	96914	C20H23O7
375.181278	0.1	65037	C21H27O6
393.025161	0.1	107929	C20H9O9
393.040439	0.1	115025	C24H9O6
393.061683	-0.2	181874	C21H13O8
393.077008	-0.4	83319	C25H13O5
393.082746	-0.1	458123	C18H17O10

393.097697	0.7	89959	C22H17O7
393.119135	-0.1	695404	C19H21O9
393.155521	-0.1	1147021	C20H25O8
393.191901	-0.1	753742	C21H29O7
393.228264	-0.1	90640	C22H33O6

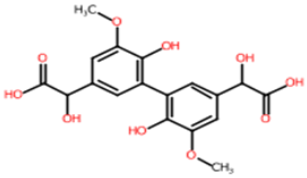
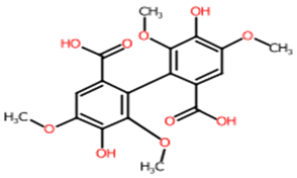
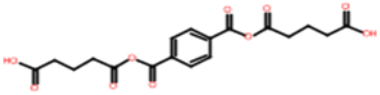
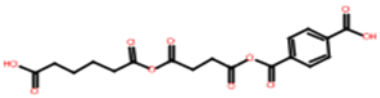
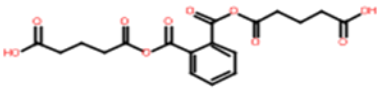
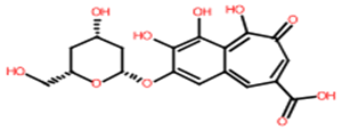
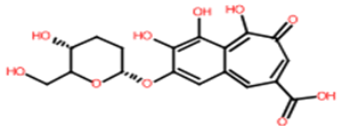
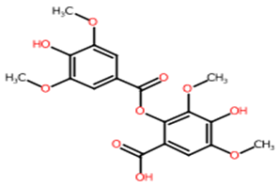
Appendix 3. 2. Fragmentation data of the precursor ion $[C_{18}H_{18}O_{10}\text{-}H]^-$ from Pantanal sample using the ESI-TIMS-FT-ICR MS/MS (CHEF-SORI-CID) procedure.

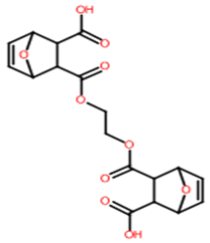
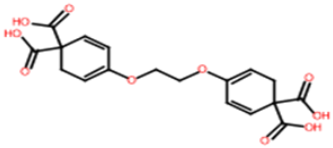
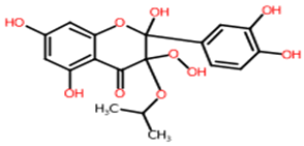
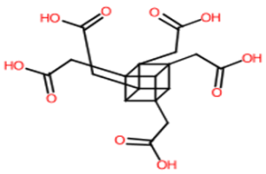
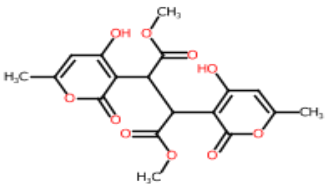
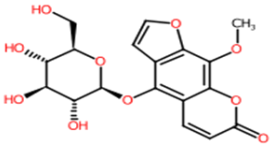
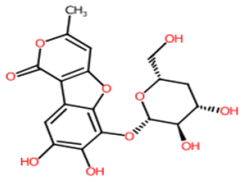
IMS band	m/z	Error [ppm]	I	Ion Formula
I	229.087096	-0.3	158108	C14H13O3
	273.076852	<0.1	181455	C15H13O5
	287.092317	0.6	179244	C16H15O5
	305.103150	-0.3	202259	C16H17O6
	317.066753	-0.2	191733	C16H13O7
	331.082395	-0.2	182547	C17H15O7
	349.092927	-0.1	494722	C17H17O8
	361.056454	0.1	235864	C17H13O9
	375.072296	-0.4	228862	C18H15O9
	393.082591	0.3	1062751	C18H17O10
II	187.076414	0.2	163289	C12H11O2
	197.060779	0.1	162865	C13H9O2
	201.092131	-0.1	168549	C13H13O2
	211.076507	-0.3	201832	C14H11O2
	213.055729	-0.1	220940	C13H9O3
	229.086994	0.1	494729	C14H13O3
	241.050590	0.2	190598	C14H9O4
	245.081884	0.2	237643	C14H13O4
	255.066234	0.2	272264	C15H11O4
	261.113176	0.2	202421	C15H17O4
	273.076849	<0.1	675673	C15H13O5
	287.092525	-0.1	314842	C16H15O5
	305.103065	<0.1	499301	C16H17O6
	317.066670	<0.1	668923	C16H13O7
	331.082379	-0.2	355835	C17H15O7
	349.092873	0.1	1265238	C17H17O8
	361.056540	-0.1	848577	C17H13O9
	375.072169	<0.1	380328	C18H15O9
	393.082714	<0.1	1853922	C18H17O10
	143.050270	-0.2	160000	C10H7O
	145.02951	-0.1	152558	C9H5O2
	159.045174	-0.1	211663	C10H7O2
	169.065882	<0.1	197634	C12H9O
	171.045193	-0.2	181194	C11H7O2
	173.060805	<0.1	167662	C11H9O2
	185.060870	-0.4	188063	C12H9O2
	187.040032	0.2	164033	C11H7O3
	197.060847	-0.2	362771	C13H9O2

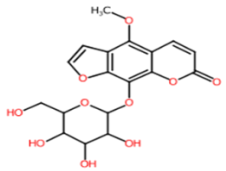
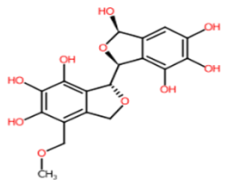
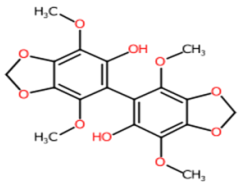
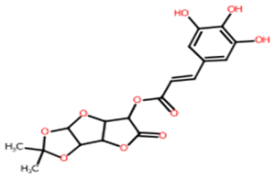
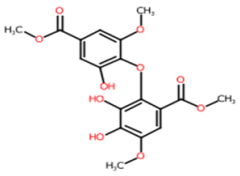
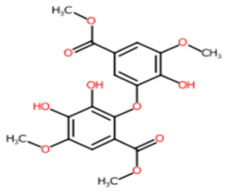
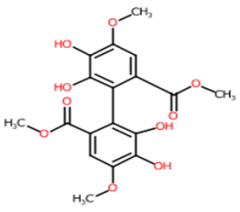
	199.040116	-0.2	267423	C12H7O3
	201.055764	-0.2	177832	C12H9O3
	201.092106	<0.1	148715	C13H13O2
	211.076465	-0.1	193108	C14H11O2
	213.055759	-0.2	381127	C13H9O3
	215.034848	0.6	183548	C12H7O4
	217.050492	0.6	165295	C12H9O4
	219.066416	-0.6	161006	C12H11O4
III	229.050634	<0.1	198874	C13H9O4
	229.086980	0.2	564683	C14H13O3
	231.066278	<0.1	277400	C13H11O4
	241.050658	-0.1	583292	C14H9O4
	245.081968	-0.1	246986	C14H13O4
	247.061285	-0.4	161056	C13H11O5
	255.066359	-0.3	306227	C15H11O4
	257.045561	-0.1	220660	C14H9O5
	261.113249	-0.1	180111	C15H17O4
	273.076901	-0.2	1149462	C15H13O5
	275.056141	-0.1	208049	C14H11O6
	285.040550	-0.3	401552	C15H9O6
	287.056136	-0.1	165508	C15H11O6
	287.092575	-0.3	237008	C16H15O5
	289.071824	-0.2	249628	C15H13O6
	299.056296	-0.6	257011	C16H11O6
	305.103086	-0.1	640043	C16H17O6
	317.066768	-0.3	1457538	C16H13O7
	331.082457	-0.4	382044	C17H15O7
	343.046056	-0.3	176366	C17H11O8
	349.092923	-0.1	2157626	C17H17O8
	361.056542	-0.1	1546083	C17H13O9
	375.072239	-0.2	566723	C18H15O9
	393.082752	-0.1	2868990	C18H17O10
	197.060809	<0.1	309407	C13H9O2
	213.055749	-0.1	325821	C13H9O3
	229.050530	0.4	182594	C13H9O4
	229.087013	<0.1	271172	C14H13O3
	241.050687	-0.2	499407	C14H9O4
	257.045515	0.1	211750	C14H9O5
IV	273.076863	-0.1	768448	C15H13O5
	285.040463	<0.1	352002	C15H9O6
	289.071884	-0.4	192303	C15H13O6

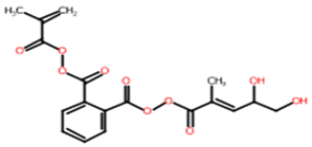
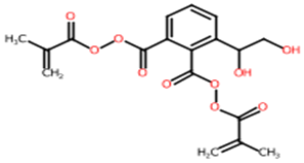
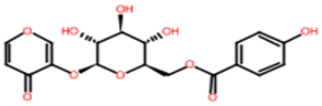
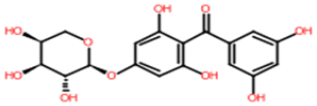
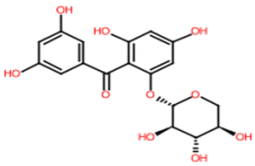
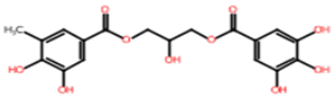
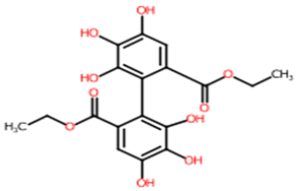
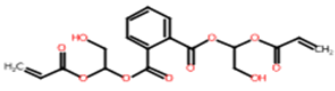
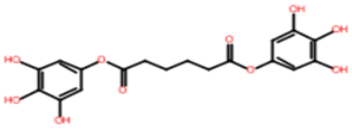
	305.103129	-0.2	390440	C16H17O6
	317.066739	-0.2	1056900	C16H13O7
	331.082256	0.2	227991	C17H15O7
	349.092934	-0.1	1439135	C17H17O8
	361.056481	0.1	1131515	C17H13O9
	375.072310	-0.4	295243	C18H15O9
	393.082650	0.2	1354904	C18H17O10
	197.060782	0.1	176377	C13H9O2
	241.050630	<0.1	267977	C14H9O4
	273.076779	0.2	383532	C15H13O5
V	285.040351	0.4	253340	C15H9O6
	317.066703	-0.1	727371	C16H13O7
	349.092935	-0.1	826272	C17H17O8
	361.056573	-0.2	708726	C17H13O9
	393.082718	<0.1	690109	C18H17O10

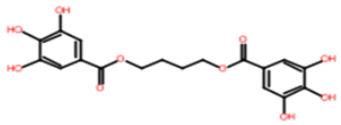
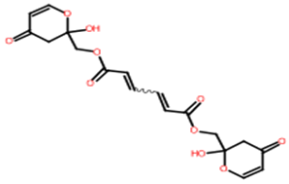
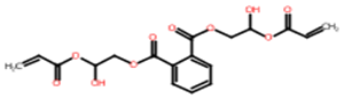
Appendix 3. 3. Candidate isomeric structures retrieved from PubChem database for $C_{18}H_{18}O_{10}$ and filtered based on ESI-TIMS-FT-ICR MS/MS (CHEF-SORI-CID) fragmentation data. Theoretical Collisional Cross Sections calculated using the software IMoS.

ID	Structure	CCS ($Å^2$)
1		191.2
2		190.9
3		188.5
4		201.9
5		184.9
6		191.7
7		193.3
8		193.2

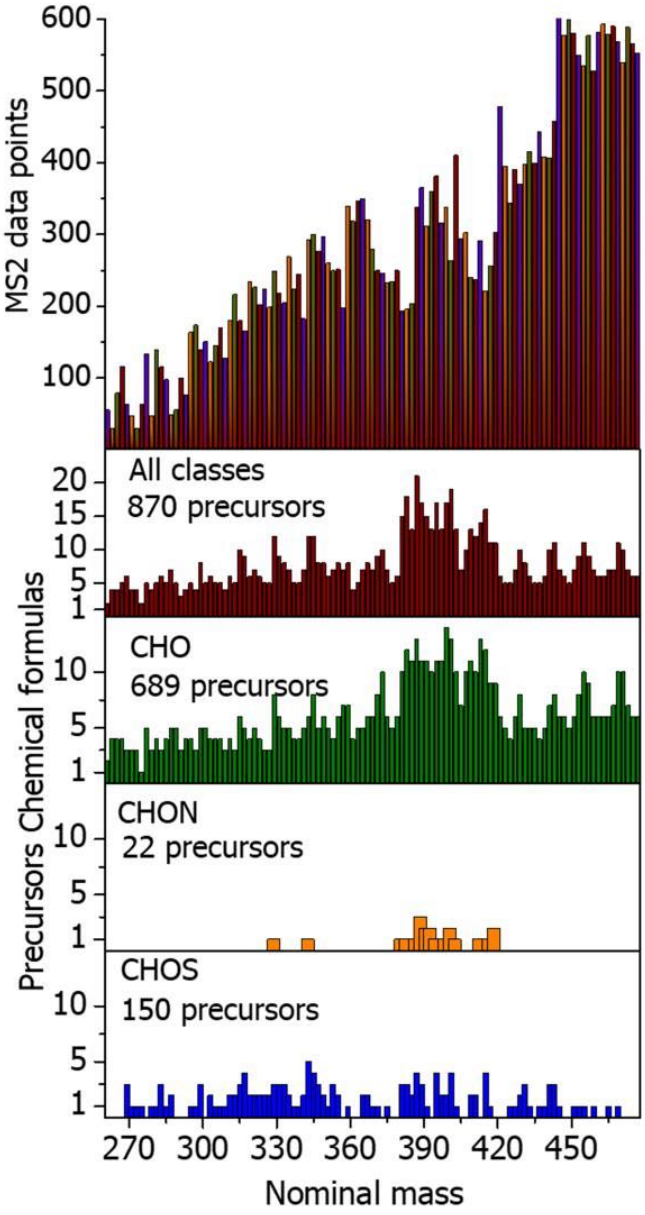
9		187.6
10		186.1
11		191.4
12		189.5
13		194.2
14		189.1
15		191.0

16		186.0
17		188.6
18		192.4
19		200.3
20		198.4
21		197.6
22		192.8

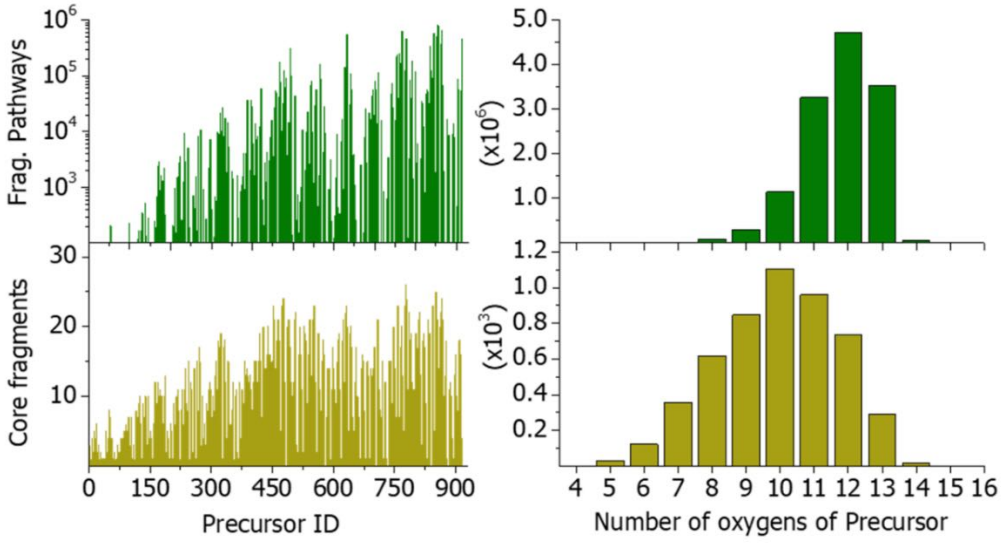
23		197.6
24		196.2
25		211.9
26		199.7
27		188.2
28		179.0
29		193.5
30		200.9
31		182.6

32		184.1
33		200.0
34		209.0

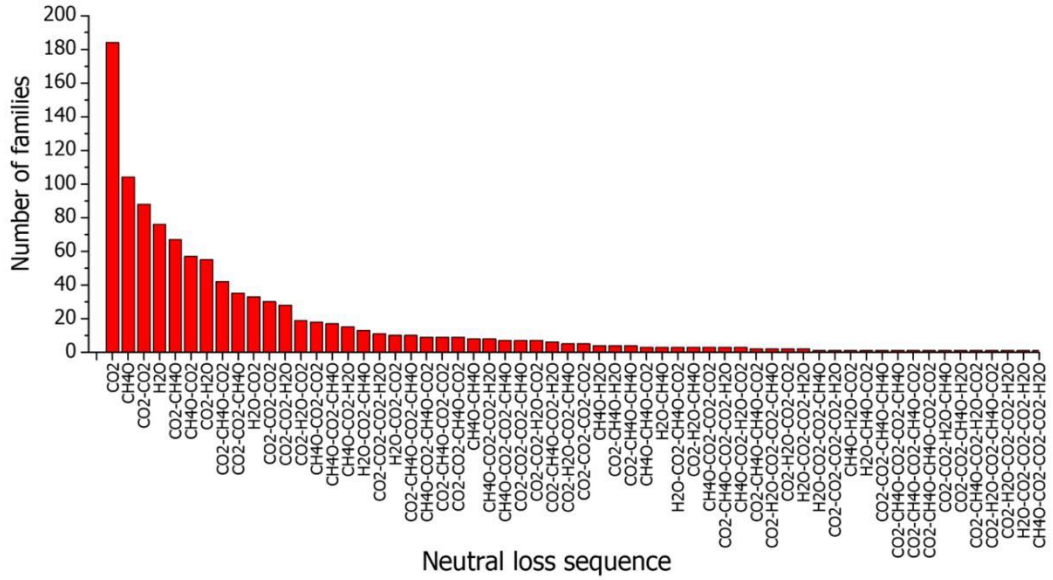
Appendix 4. 1.MS/MS data points (S/N> 3) collected per nominal m/z (top) and number of precursors chemical formulas assigned per nominal m/z for the CHO, CHON and CHOS heteroatom classes (bottom).



Appendix 4. 2. Distribution of number of fragmentation pathways and core fragments per precursor ID and oxygen class of the precursor for the CHO class (green) and all compound classes (purple).



Appendix 4. 3.Distribution of the number of structural families of CHO compounds per unique neutral loss sequence found.

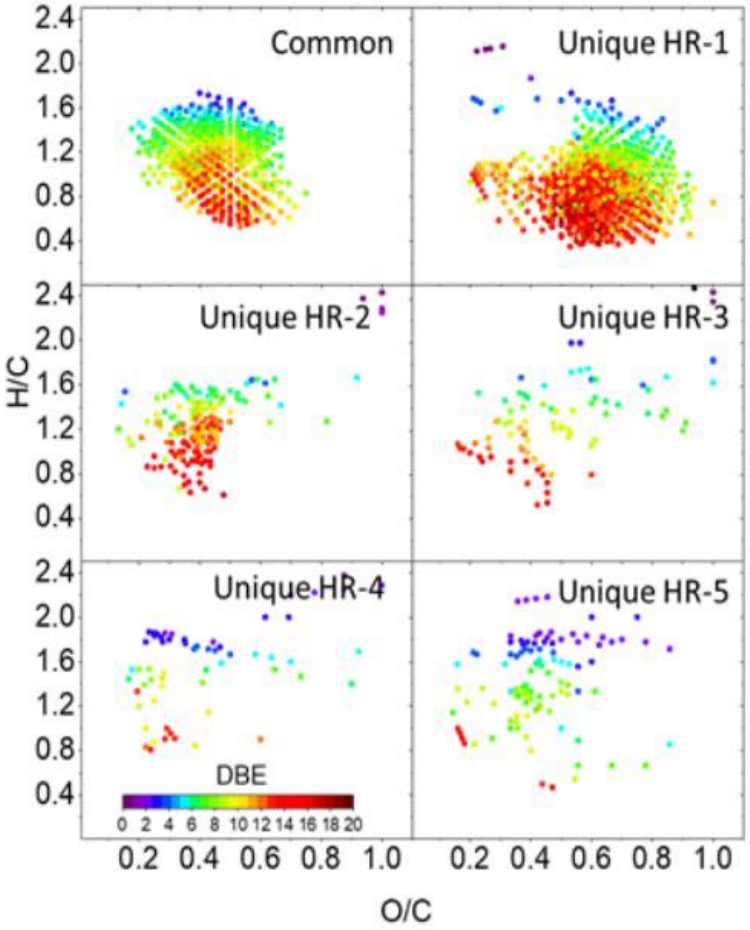


Appendix 4. 4.Comparison on fragmentation pathways obtained from ESI-FT ICR CASI-CID vs ESI-FT-ICR CHEF-SORI-CID using the Graph-DOM code.

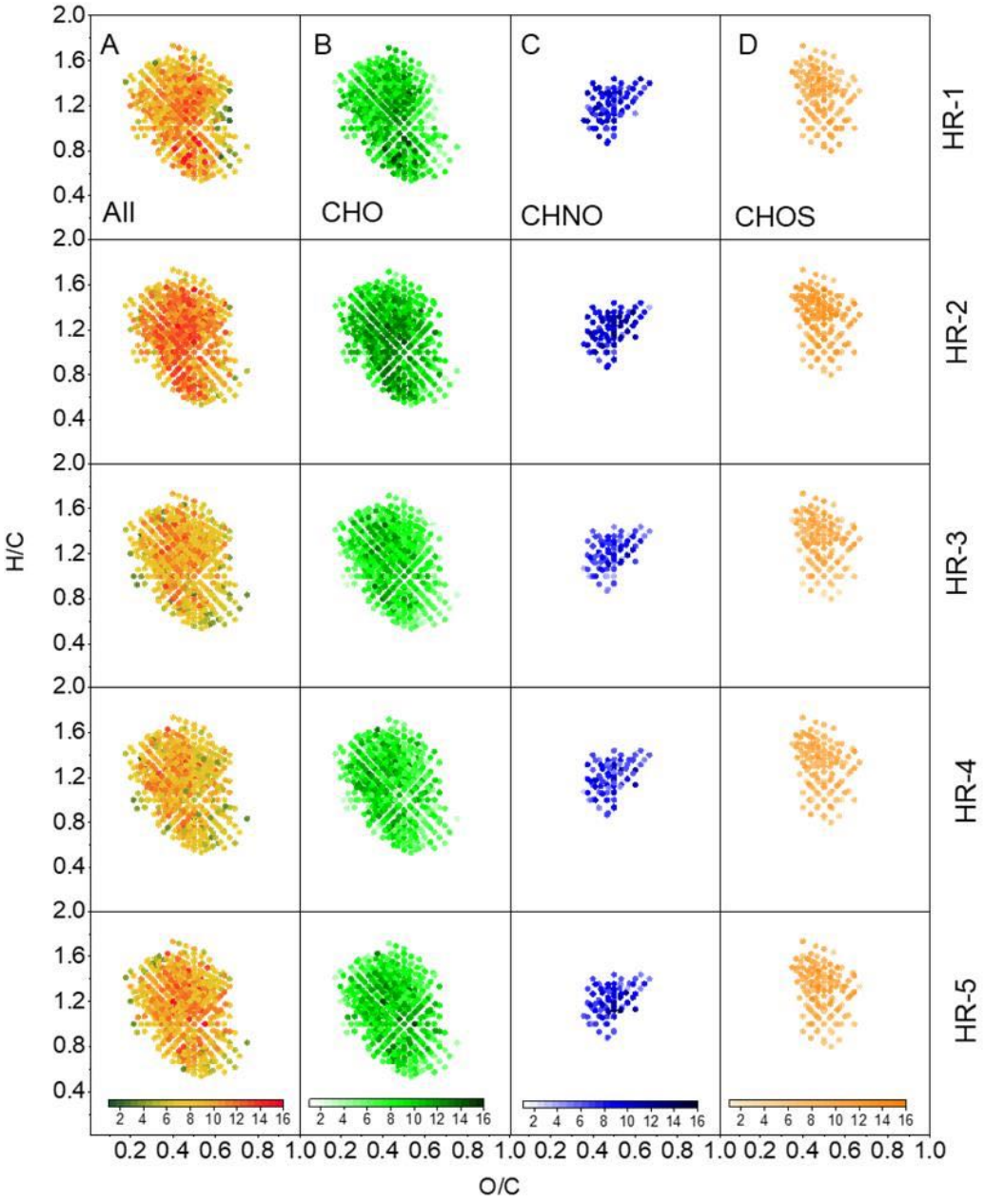
ESI-FT-ICR CHEF-SORI-CID				
Precursor *m/z*	Fragment *m/z*	Abundance	Chemical formula	Fragmentation pathways
267.087412 $C_{13}H_{15}O_6$	121.065869	2319218	C_8H_9O	$[C_{13}H_{13}O_5\text{-}C_{12}H_{13}O_3\text{-}C_{11}H_{13}O]$ $[C_{12}H_{15}O_4\text{-}C_{12}H_{13}O_3\text{-}C_{11}H_{13}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{11}H_{13}O]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{10}H_{11}O_2\text{-}C_{10}H_{11}O]$
	137.060775	2232673	$C_8H_9O_2$	
	147.081507	2881464	$C_{10}H_{11}O$	
	161.097190	3524949	$C_{11}H_{13}O$	
	163.076453	2921173	$C_{10}H_{11}O_2$	
	179.107743	8029749	$C_{11}H_{15}O_2$	
	191.071333	7041986	$C_{11}H_{11}O_3$	
	205.087027	4331752	$C_{12}H_{13}O_3$	
	223.097579	13365070	$C_{12}H_{15}O_4$	
	235.061154	6594586	$C_{12}H_{11}O_5$	
	249.076737	1962522	$C_{13}H_{13}O_5$	
ESI-FT-ICR CASI-CID				
267.087377 $C_{13}H_{15}O_6$	107.050229	1172882	C_7H_7O	$[C_{13}H_{13}O_5\text{-}C_{13}H_{11}O_4]$ $[C_{13}H_{13}O_5\text{-}C_{12}H_{13}O_3\text{-}C_{11}H_{13}O]$ $[C_{12}H_{15}O_4\text{-}C_{12}H_{13}O_3\text{-}C_{11}H_{13}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{11}H_{13}O]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_9O_4\text{-}C_{10}H_9O_2]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_9O_2]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_9O_2]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_7O_2]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_7O_2]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{11}H_{15}O_1\text{-}C_{10}H_{11}O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{10}H_{11}O_2\text{-}C_{10}H_{11}O]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_9H_7O_2\text{-}C_8H_7O]$ $[C_{12}H_{11}O_5\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_9H_{11}O\text{-}C_8H_7O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_9H_7O_2\text{-}C_8H_7O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{11}O_3\text{-}C_{10}H_{11}O_2\text{-}C_9H_{11}O\text{-}C_8H_7O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{10}H_{11}O_2\text{-}C_9H_7O_2\text{-}C_8H_7O]$ $[C_{12}H_{15}O_4\text{-}C_{11}H_{15}O_2\text{-}C_{10}H_{11}O_2\text{-}C_9H_{11}O\text{-}C_8H_7O]$
	119.013844	1697719	$C_7H_3O_2$	
	119.050241	1658533	C_8H_7O	
	121.029502	6533664	$C_7H_5O_2$	
	121.065878	1345438	C_8H_9O	
	135.045151	4004030	$C_8H_7O_2$	
	135.081539	3482818	$C_9H_{11}O$	
	137.024437	1138394	$C_7H_5O_3$	
	137.060794	1726911	$C_8H_9O_2$	
	147.008775	2315746	$C_8H_3O_3$	
	147.045153	4937069	$C_9H_7O_2$	
	147.081544	1167991	$C_{10}H_{11}O$	
	149.024438	1511373	$C_8H_5O_3$	
	149.060794	1682393	$C_9H_9O_2$	
	151.040067	5223788	$C_8H_7O_3$	
	153.019313	1299564	$C_7H_5O_4$	
	159.045146	1124616	$C_{10}H_7O_2$	
	161.060788	1851764	$C_{10}H_9O_2$	
	161.09718	2097796	$C_{11}H_{13}O$	
	163.003675	5962847	$C_8H_3O_4$	
	163.040076	2217199	$C_9H_7O_3$	
	163.076444	1458687	$C_{10}H_{11}O_2$	
	163.112902	1220176	$C_{11}H_{15}O$	
	164.011518	316041	$C_8H_4O_4$	

	177.019323	1542385	$C_9H_5O_4$	
	179.034977	6589323	$C_9H_7O_4$	
	179.071362	3357598	$C_{10}H_{11}O_3$	
	179.07138	532659	$C_{10}H_{11}O_3$	
	179.10776	2915761	$C_{11}H_{15}O_2$	
	190.998587	2883017	$C_9H_3O_5$	
	191.03498	2834395	$C_{10}H_7O_4$	
	191.071345	1759726	$C_{11}H_{11}O_3$	
	193.014245	1842238	$C_9H_5O_5$	
	193.050566	1163728	$C_{10}H_9O_4$	
	195.029898	3064224	$C_9H_7O_5$	
	205.014202	1109487	$C_{10}H_5O_5$	
	205.050612	1184126	$C_{11}H_9O_4$	
	205.087044	1357197	$C_{12}H_{13}O_3$	
	205.123411	1755548	$C_{13}H_{17}O_2$	
	223.024773	1324645	$C_{10}H_7O_6$	
	223.061211	2928750	$C_{11}H_{11}O_5$	
	223.097586	2609783	$C_{12}H_{15}O_4$	
	223.13396	2489983	$C_{13}H_{19}O_3$	
	229.108111	278543	$C_{11}H_{17}O_5$	
	229.144591	408368	$C_{12}H_{21}O_4$	
	231.066346	406144	$C_{13}H_{11}O_4$	
	231.10268	273952	$C_{14}H_{15}O_3$	
	231.123783	327984	$C_{11}H_{19}O_5$	
	234.988495	368776	$C_{10}H_3O_7$	
	235.024842	1239139	$C_{11}H_7O_6$	
	235.061205	1478248	$C_{12}H_{11}O_5$	
	249.004058	3009462	$C_{11}H_5O_7$	
	249.040425	1717177	$C_{12}H_9O_6$	
	249.076867	280973	$C_{13}H_{13}O_5$	

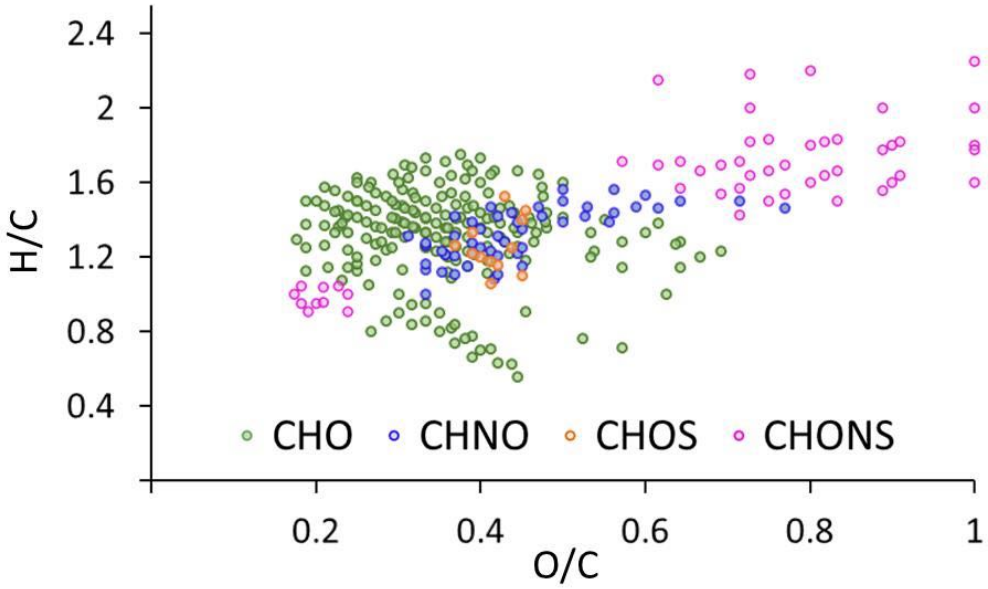
Appendix 5. 1.Van Krevelen plots highlighting the unsaturation level of common and unique chemical compounds found across the DOM samples at the Harney River. Notice that the DOM molecules at HR-1 are more unsaturated than the common ones detected along the transect.



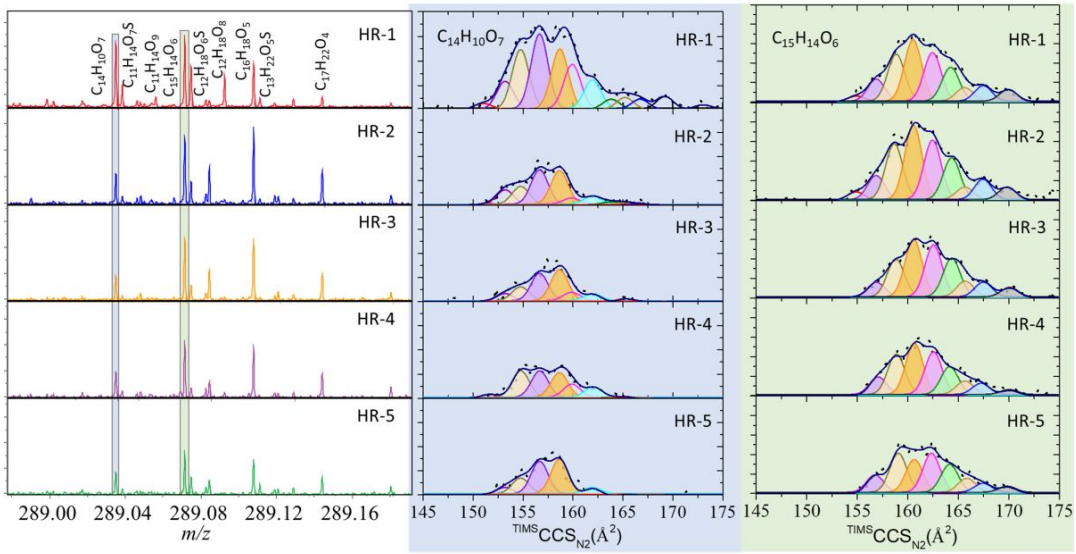
Appendix 5. 2.Van Krevelen plots highlighting the isomeric complexity of molecular formulas shared by all DOM samples (A) and filtered by heteroatom classes (B-D). Note that all horizontal plots are associated with the sample bar scales indicate the number of isomers estimated by molecular formula.



Appendix 5. 3.Van Krevelen plot of common compounds identified across HR-2, HR-3, HR-4, and HR-5 DOM samples.



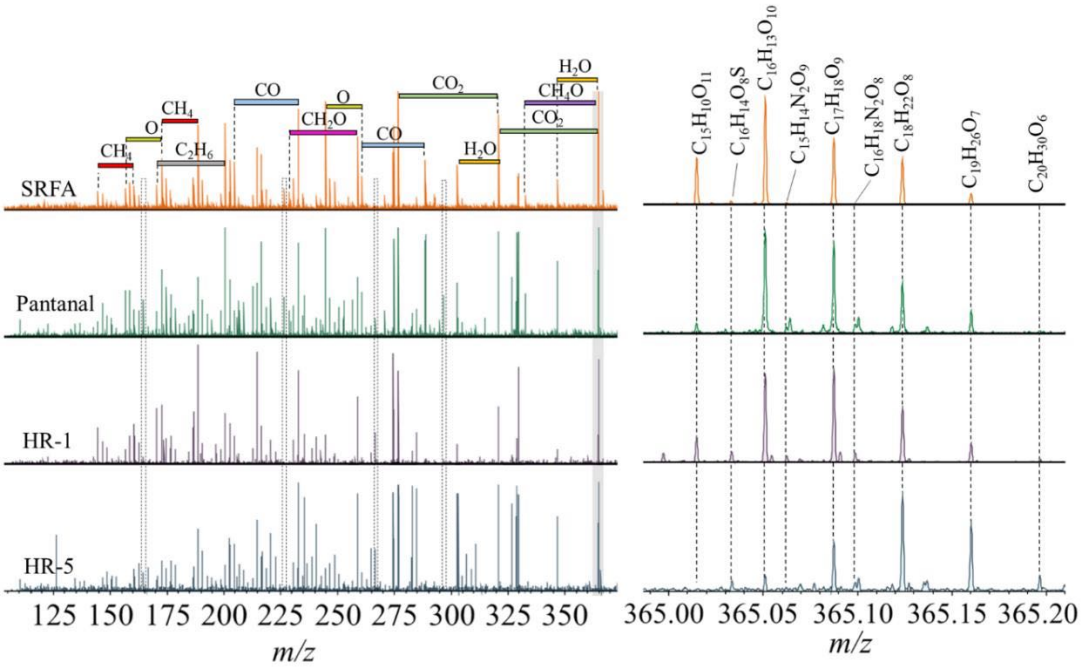
Appendix 5. 4. Expanded view (289 *m/z*) of the ESI-FT ICR MS spectra collected for the five DOM samples with isobaric molecular species annotated (left). Extracted TIMSCCSN2 profiles of the molecular formulas $C_{14}H_{10}O_7$ (center) and $C_{15}H_{14}O_6$ (right) obtained from ESI-TIMS-FT ICR MS along the salinity transect of the Harney River.



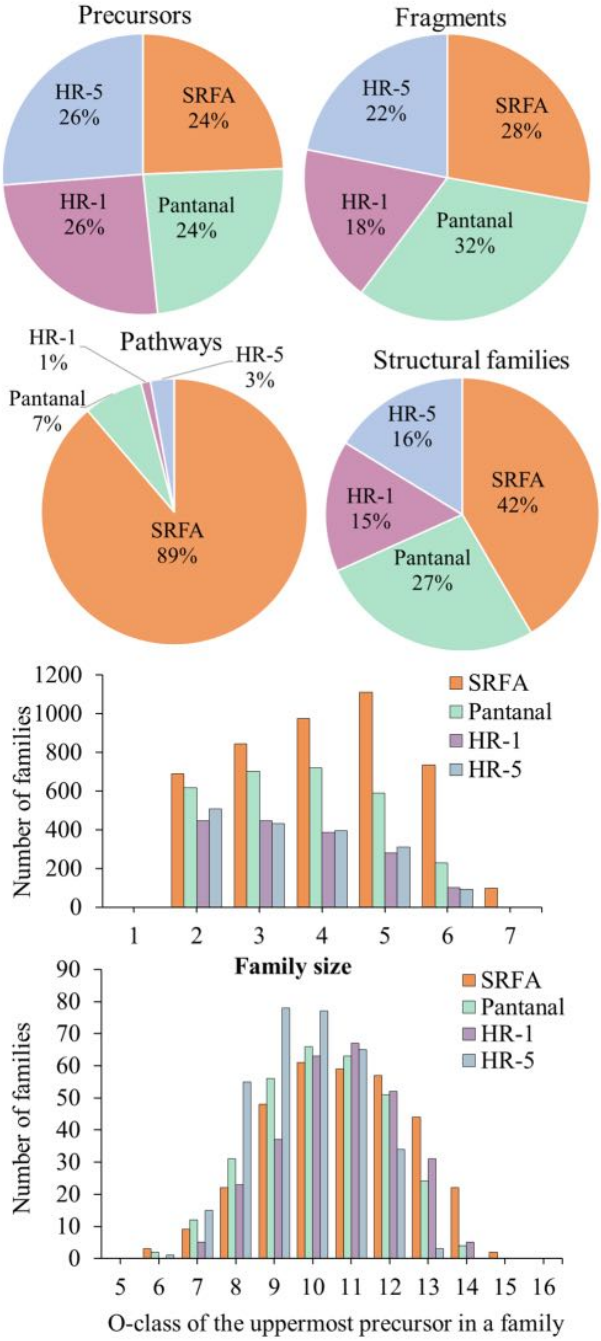
Appendix 6. 1.Summary of compositional information, including the CHO, CHON, CHOS, and CHONS heteroatom classes obtained for SRFA, Pantanal, HR-1, and HR-5 samples using ESI-FT-ICR MS.

Sample	SRFA	Pantanal	HR-1	HR-5
Total formulas	2919	2915	2854	1820
CHO	2117	2239	1399	961
CHON	497	541	729	486
CHOS	305	130	635	352
CHONS	0	5	91	21
CHO%	72.5	76.7	47.9	32.9
CHON%	17.0	18.5	25.0	16.6
CHOS%	10.4	4.5	21.8	12.1
CHONS%	0	0.2	3.1	0.7
Precursors CHO	989	916	1033	1064

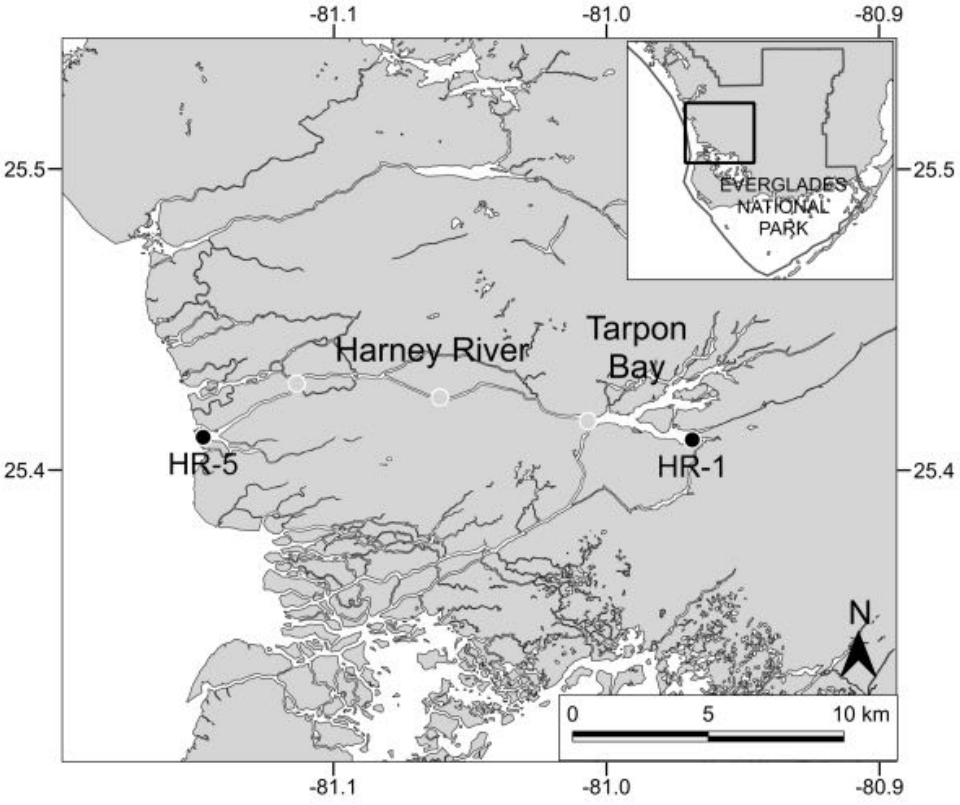
Appendix 6. 2. Expanded view of precursors molecules at m/z 365 isolated in the quadrupole (Δm/z=1) from the SRFA, Pantanal, HR-1, and HR-5 DOM samples respectively (left). Chemical formulas assigned to mass signals are also shown. ESI (-) FT ICR CID MS/MS spectra of isolated precursors at nominal mass 365 for the SRFA, Pantanal, HR-1, and HR-5 DOM samples, respectively. Typical neutral losses of H_2O, CH_4, O, C_2H_6, CH_2O, CH_4O, CO, and CO_2 are also annotated (right).



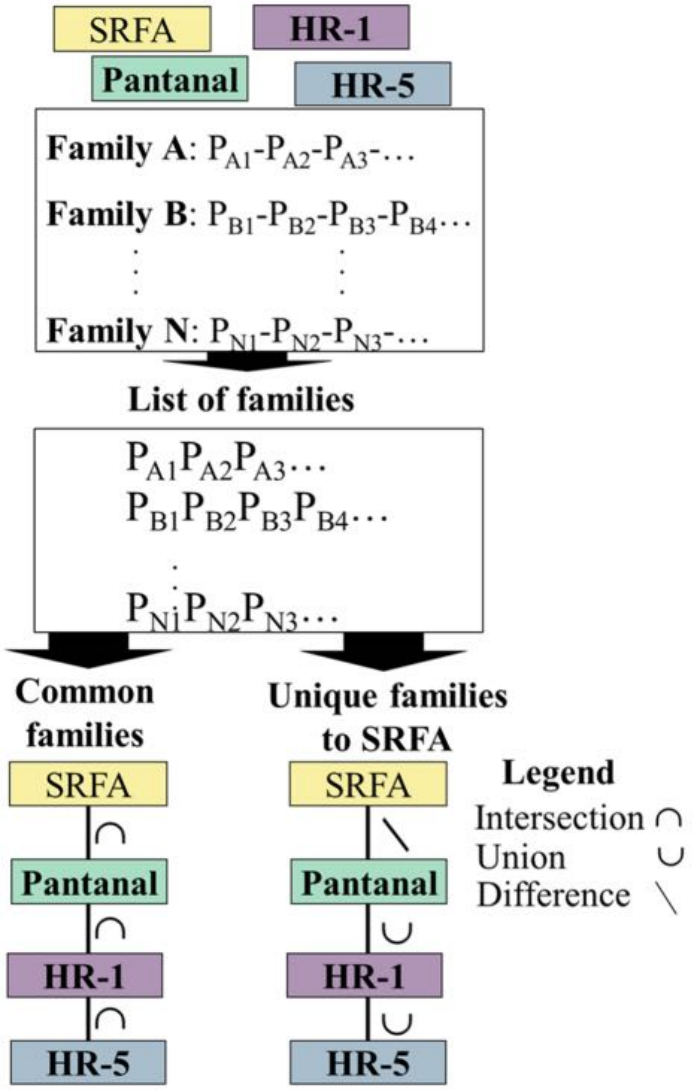
Appendix 6. 3. Comparison of the number of CHO precursors and fragments detected in each DOM sample by the ESI-FT-ICR CASI CID MS/MS analytical workflow. Number of fragmentation pathways and structural families computed across samples by Graph-DOM (top panel). Distribution of the number of families per family size (center panel) and number of families per oxygen class of the uppermost precursor for the CHO class of SRFA, Pantanal, HR-1, and HR-5 DOM samples (bottom panel).



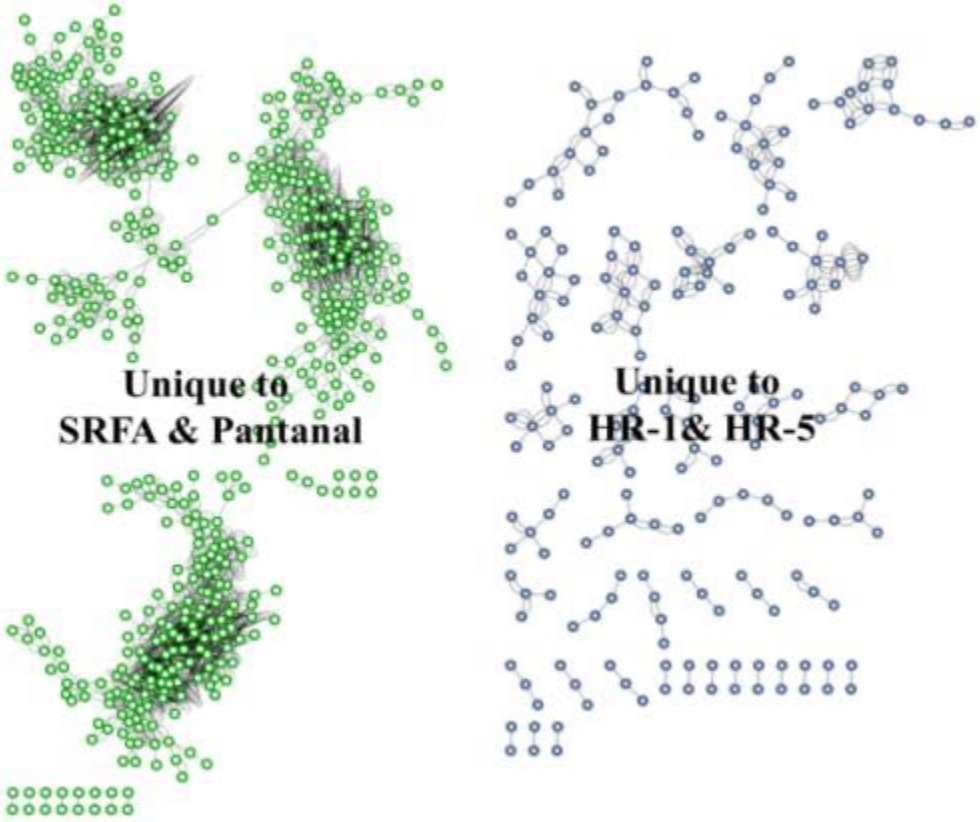
Appendix 6. 4. Map of the geographical location of HR-1 and HR-5 sampling points at the Harney River, Florida Everglades.



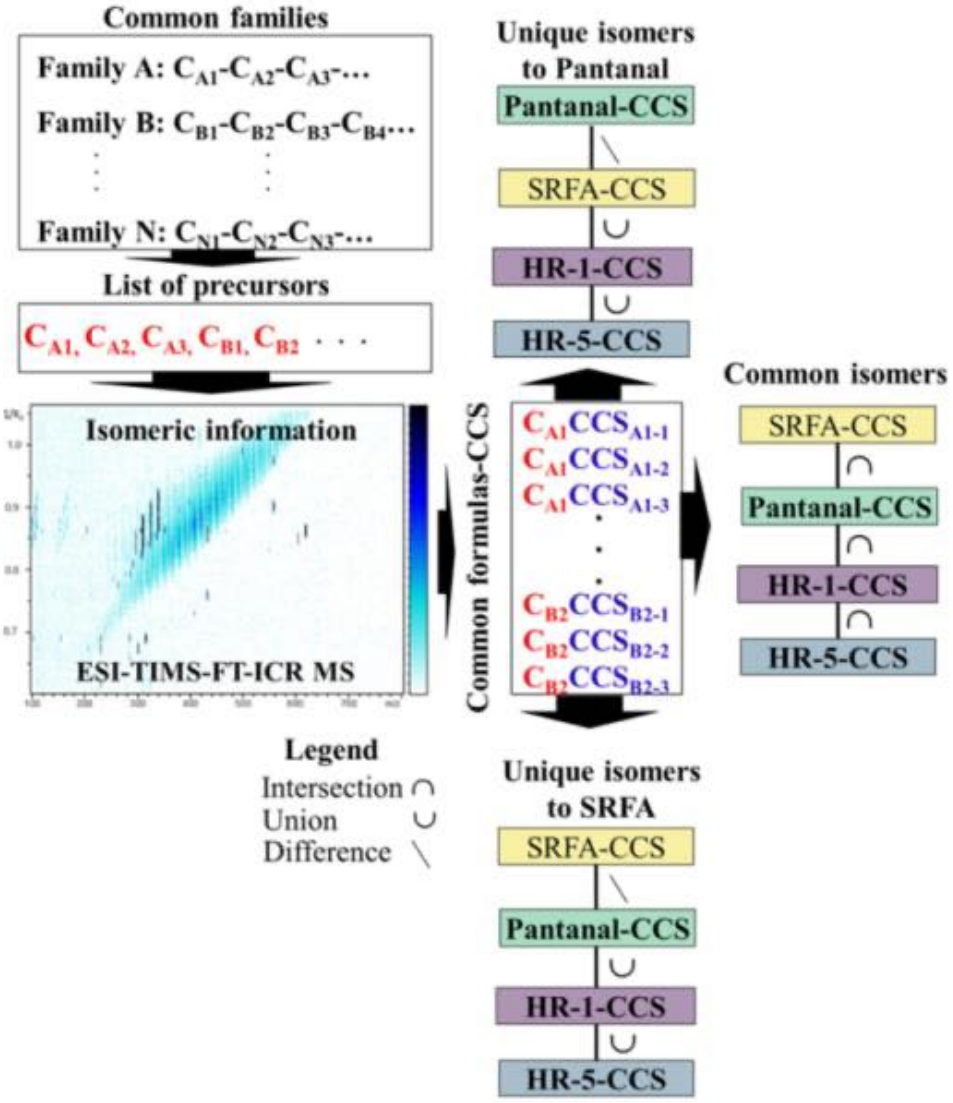
Appendix 6. 5.Conceptual model based on mathematical sets developed for the comparison of DOM structural families obtained from Graph-DOM code.



Appendix 6. 6.Cytoscape networks of structural families resulting from the intersection between SRFA and Pantanal, excluding both HR-1 and HR-5 families and the intersection between HR-1 and HR-5 excluding SRFA and Pantanal samples.



Appendix 6. 7.Conceptual model based on mathematical sets developed for the isomeric comparison of precursors molecules of the structural families shared by SRFA, Pantanal, HR-1 and HR-5 DOM samples.



VITAE

VITA

DENNYS LEYVA BOMBUSE

Born: Matanzas, Cuba

1997 - 2002	B.S. Chemistry Concentration: Radiochemistry Higher Institute of Sciences and Nuclear Technology Havana, Cuba.
2004 - 2007	M.S. Radiochemistry Higher Institute of Technologies and Applied Sciences Havana, Cuba.
2018 - 2021	M.S. Chemistry Florida International University Miami, Florida
	Doctoral Candidate Florida International University Miami, Florida
	Graduate Research Assistant Florida International University Miami, Florida

SELECTED PEER REVIEWED PUBLICATIONS AND PRESENTATIONS

- Leyva, D.; Tariq, U.; Saeed, F.; Jaffé, R.; Fernandez-Lima, F. Unsupervised structural classification of Dissolved Organic Matter based on fragmentation pathways. *Environ. Sci. Technol.* 2022, 56 (2), 1458-1468.
- Tariq, U.; Leyva, D.; Fernandez-Lima, F.; Saeed, F. Graph theoretic approach for the analysis of comprehensive mass spectrometry (MS/MS) data of Dissolved Organic Matter. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2021, 3742-3746.
- Leyva, D.; Jaffé, R.; Fernandez-Lima, F. Structural Characterization of Dissolved Organic Matter at the Chemical Formula Level Using TIMS-FT-ICR MS/MS. *Anal. Chem.* 2020, 92 (17), 11960-11966.
- Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the structural complexity of dissolved organic matter: isomeric diversity. *Faraday Discuss.* 2019, 218 (0), 431-440.

- Tose, L. V.; Benigni, P.; Leyva, D.; Sundberg, A.; Ramírez, C. E.; Ridgeway, M. E.; Park, M. A.; Romão, W.; Jaffé, R.; Fernandez-Lima, F. Coupling trapped ion mobility spectrometry to mass spectrometry: trapped ion mobility spectrometry–time-of-flight mass spectrometry versus trapped ion mobility spectrometry–Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom*. 2018, 32 (15), 1287-1295.
- Leyva, D.; Tariq, U.; Saeed, F.; Jaffé, R.; Fernandez-Lima, F. Unsupervised Structural Classification of Dissolved Organic Matter based on fragmentation pathways. ASMS (American Society of Mass Spectrometry) annual conference, Philadelphia, Pennsylvania, 2021 (Oral presentation).
- Leyva, D.; Jaffé, R.; Fernandez-Lima, F. Chemical formula and structural level characterization of Dissolved Organic Matter by TIMS-FT-ICR MS/MS. ASMS annual conference, 2020. (Online Poster)
- Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Unraveling the Structural Complexity and Diversity of Dissolved Organic Matter using TIMS-FT-ICR MS. 15th Annual Workshop on LC/MS, Miami, Florida, 2019 (Oral presentation).
- Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the Structural Complexity of Dissolved Organic Matter: isomeric diversity. ASMS annual conference, Atlanta, Georgia, 2019 (Poster).
- Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Unraveling the Structural Complexity and Diversity of Dissolved Organic Matter using TIMS-FT-ICR MS.12th North American FTMS Conference, Key West, Florida, 2019 (Poster).
- Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Unraveling the Structural Complexity and Diversity of Dissolved Organic Matter using TIMS-FT-ICR MS. ACS Florida annual meeting and exposition (FAME), Tampa, Florida, 2019 (Oral presentation)