**Michigan
Technological
University**

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2021

High-resolution molecular characterization of complex environmental mixtures: Aquatic dissolved organic matter and wildfire-influenced aerosol

Amna Ijaz
Michigan Technological University, aijaz@mtu.edu

Copyright 2021 Amna Ijaz

Recommended Citation

Ijaz, Amna, "High-resolution molecular characterization of complex environmental mixtures: Aquatic dissolved organic matter and wildfire-influenced aerosol", Open Access Dissertation, Michigan Technological University, 2021.
<https://doi.org/10.37099/mtu.dc.etdr/1235>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etdr>

Part of the [Analytical Chemistry Commons](#), and the [Environmental Chemistry Commons](#)

HIGH-RESOLUTION MOLECULAR CHARACTERIZATION OF COMPLEX
ENVIRONMENTAL MIXTURES: AQUATIC DISSOLVED ORGANIC MATTER AND
WILDFIRE-INFLUENCED AEROSOL

By

Amna Ijaz

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Chemistry

MICHIGAN TECHNOLOGICAL UNIVERSITY

2021

© 2021 Amna Ijaz

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Chemistry.

Department of Chemistry

Dissertation Advisor: *Lynn R Mazzoleni*

Committee Member: *Daisuke Minakata*

Committee Member: *Evan S Kane*

Committee Member: *Sarah A Green*

Department Chair: *Sarah A Green*

ابو اور امی جان کے لیے۔

Table of Contents

Author Contribution Statement	vii
Acknowledgements	viii
List of Abbreviations	x
Abstract	xii
1 Introduction.....	1
1.1 Complex environmental mixtures.....	1
1.1.1 Atmospheric aerosol.....	1
1.1.1.1 Wildfire-influenced organic aerosol: Chemical properties and atmospheric processing.....	2
1.1.2 Aquatic humic substances.....	6
1.1.2.1 Commercial organic matter standard/reference mixtures.....	7
1.2 High-resolution mass spectrometry	10
1.2.1 Principles of operation	10
1.2.2 Ionization methods.....	11
1.2.2.1 Electrospray ionization.....	11
1.2.2.2 Atmospheric pressure chemical ionization.....	12
1.2.2.3 Atmospheric pressure photoionization.....	12
1.2.2.4 Laser desorption ionization	13
1.2.3 Mass analysers and signal processing.....	14
1.2.3.1 Orbitrap Elite	14
1.2.3.2 Fourier transform - Ion Cyclotron Resonance	16
1.2.3.3 Fourier-transformation of image current signals	18
1.2.4 From measured mass to molecular formula	19
1.2.4.1 MFAssignR	21
1.2.4.2 Computation of estimated structural and physicochemical parameters	22
2 Expanding the Molecular Characterization of Natural Organic Matter with Complementary Ionization Methods	26
2.1 Introduction	26
2.2 Methods	28
2.2.1 Sample acquisition and preparation	28
2.2.2 Sample ionization and ultrahigh-resolution mass spectrometry	28
2.2.3 Data processing and formula assignment	29
2.2.4 Computation of chemical characteristics and statistical analysis	30
2.3 Results and Discussion.....	31
2.3.1 General characteristics of the mass spectra.....	31
2.3.2 Molecular composition: Difference in trends of C, O, and DBE.....	33
2.3.3 Ionization selectivity: Exclusive and common detection of ions	36
2.4 Supplementary material.....	42

2.4.1	Supplementary tables	42
2.4.2	Supplementary figures	45
3	Qualitative Molecular Characterization of Fulvic Acid mixtures from the International Humic Substances Society with High-Resolution Orbitrap Mass Spectrometry	50
3.1	Introduction	51
3.2	Materials and Methods.....	52
3.2.1	Sample acquisition and preparation	52
3.2.2	Electrospray ionization and ultrahigh-resolution mass spectrometry	53
3.2.3	Assignment of molecular formulae	54
3.2.4	Computation of chemical characteristics and inter-sample differences	55
3.3	Results and Discussion.....	56
3.3.1	General characteristics of the mass spectra.....	56
3.3.2	Molecular composition of the IHSS DOM mixtures.....	60
3.3.2.1	Pony Lake fulvic acids	60
3.3.2.2	Suwannee River fulvic acids and natural organic matter (NOM) and Mississippi River NOM.....	62
3.3.2.3	Nordic Lake fulvic acids	64
3.3.2.4	Elliot soil fulvic acids	65
3.3.3	A common DOM fraction amidst inter-sample differences	69
3.3.4	Ionization complementarity and selectivity	74
3.4	Acknowledgements.....	75
3.5	Data availability.....	75
3.6	Supplementary material	76
3.6.1	Manual data validation.....	76
3.6.2	Supplementary tables	77
3.6.3	Supplementary figures.....	86
4	High-Resolution Molecular Characterization of Tar Ball-Rich Biomass Burning Organic Aerosol with a 21-T FT-ICR Mass Spectrometer.....	93
4.1	Introduction	93
4.2	Methods	95
4.2.1	Sample collection and preparation	95
4.2.2	Electrospray ionization and mass spectrometry	96
4.2.2.1	15-T FT-ICR MS analysis	96
4.2.2.2	21-T FT-ICR MS analysis	96
4.2.3	Data processing and molecular formula assignment.....	97
4.3	Results and Discussion.....	97
4.3.1	Chemical and molecular composition.....	97
4.3.2	CHO species.....	99
4.3.3	CHNO species	104
4.3.4	CHOS species	106
4.3.5	CHOP and CHNOP species	107

4.3.6	Comparison of molecular composition delineated from 15-T and 21-T FT-ICR instruments	108
4.3.7	Instrumental considerations	110
4.4	Acknowledgments	114
4.5	Data availability	115
4.6	Supplementary material	115
4.6.1	Bibliometric analysis	115
4.6.2	Calculation of estimable physical parameters	115
4.6.3	Supplementary tables	117
4.6.4	Supplementary figures	117
5	Molecular Insights into the Anatomy of Ambient Tar Balls in Wildfire Smoke from the Pacific Northwest	123
5.1	Introduction	123
5.2	Methods	126
5.2.1	Sample collection and preparation	126
5.2.2	High-resolution mass spectrometry	126
5.3	Results and Discussion	127
5.3.1	Tar balls are more abundant in aged wildfire plumes	127
5.3.2	An oxygenated exterior with a less oxygenated amorphous core	128
5.3.3	Tar balls are extremely complex at the molecular level	133
5.4	Research implications	139
5.5	Data availability	139
5.6	Supplementary material	140
5.6.1	Ultrahigh-resolution mass spectrometry	140
5.6.1.1	Electrospray ionisation	140
5.6.1.2	Laser desorption ionisation	141
5.6.1.3	Data processing and formula assignment	141
5.6.2	Supplementary tables	142
5.6.3	Supplementary figures	147
6	Conclusions	157
6.1	Conclusions from Chapter 2	157
6.2	Conclusions from Chapter 3	158
6.3	Conclusion from Chapter 4	158
6.4	Conclusions from Chapter 5	159
6.5	Future recommendations	160
7	Works Cited	162

Author Contribution Statement

Chapter 2 presents findings from data that was acquired by Dr Maryam Khaksari and Tyler Leverton at the Chemical Advanced Resolution Methods Laboratory (ChARM) of Michigan Technological University (MTU), Houghton, MI. Data processing and analysis was performed by myself.

Chapter 3 presents research that was performed on the side of a collaborative project published by Hawkes et al., 2020 in *Limnology and Oceanography: Methods*. Mass spectrometric data was acquired by myself and Dr Maryam Khaksari at ChARM, MTU, Houghton, MI. It was processed and analyzed independent of the Hawkes' project by myself under the guidance of Dr Lynn Mazzoleni.

Chapters 4 and 5 are prepared with the intent of publication in the near future and are considered a work in progress. Sample collection was performed by Dr Swarup China at Environmental Molecular Sciences Laboratory (EMSL) of the Pacific Northwest National Laboratory, Richland, WA. Mass spectrometric data was acquired by Dr William Kew at EMSL and was processed and analyzed at MTU by myself under the supervision of Dr Lynn Mazzoleni. Both manuscripts presented in this dissertation were written by myself with significant contributions from all collaborators mentioned above.

Acknowledgements

Taaya Fayaaz: You drove all the way to the middle of nowhere on your dreaded motorcycle in the scorching heat of Lahore's *dopeher* to give me some pocket money to spend in the US – 200\$. I had laughed at how little it was, but you insisted that it was a gift from you till the day “I made someone of myself”...and then I was supposed to return it. You were insightful in the funniest way. On all the countless occasions during these past few years that I have wanted to give up, I have also always had the tiniest bit of willpower to keep going, to focus on making someone of myself, to be able to return your gift.

It was a day before I left Pakistan to do this PhD. That was the last time I saw you, spoke to you in person, received your prayers, saw that characteristic twinkle in your eyes, and that's how you will always be in my memories. Resilient, kind, hopeful, forgiving, selfless, and filled with infinite love. I cannot return your gift or repay the lifetime of strength and support you gave to me as a second father, but I will try my best to pass it on. Rest in peace, Taaya. The world has not been and will never be the same without you.

~~~~~

Dr Lynn Mazzoleni: I would not have this PhD if it weren't for you. This applies to all advisor-student situations, but it was an 'original' case here. You took me on as your student, when nobody else would. It seems like I joined your group just yesterday after a torturous process of department transfer. From my first group meeting on Aug 14, 2018, as *officially* your student till today (May 10, 2021), it has been exactly 1000 days - 2 years, 8 months, 26 days! I calculated! I had the advantage of volunteering for a few of months of summer too but finishing in such a short time would not have been possible without your patience, kindness, and honesty. Your commitment, most of all, made it possible because I would have given up on a student like me a long time ago. It wasn't your job to keep up with the emotional instability that I brought with me, but you knew when to listen and when to let me be, which I appreciate beyond words. I have you to thank not only for the science that I now know, but for the mentorship and guidance that (I think) I am prepared to provide to my own students in the future.

~~~~~

My Committee Members: Dr Daisuke Minakata (Department of Civil, Environmental, and Geospatial Engineering), Dr Evan Kane (College of Forest Resources and Environmental Science), and Dr Sarah Green (Department of Chemistry) for providing valuable comments to improve their dissertation as well as for sharing their insight about my research and its broader implications.

My collaborators at Pacific Northwest National Laboratory, Washington: Dr Swarup China for his humility, attentiveness, generosity and care. He has never left an e-mail unanswered, no matter how trivial!

Dr William Kew for acquiring *almost all* the data presented in this dissertation and entertaining my endless inane questions about it with the utmost patience.

My lab fellows at the Atmospheric Chemistry Research Group at MTU: Dr Maryam Khaksari for being my academic counsellor and an amazing tutor, for checking up on me even after she had left Michigan Tech. Dr Simeon Schum for being the answer to all my Frequently Asked Questions. He deserves to get an honorary PhD out of mine. Dr Matthew Brege, Thusitha Divisekara, Abigail Mikolitis, Elizabeth Rose, and Tyler Levertton for providing much needed guidance throughout my PhD and for being gracious and respectful colleagues.

Tahiyat and Mikhail: My sisters from other mothers, my friends, my confidantes, my counsellors. I will always be grateful for having you as my friends, but most of all for having known some of the most resilient scientists-in-the-making in the two of you. Thank you for always coming back and looking past my inability to apologise.

~~~~~

Phopo Anis, Phopo Attia, Khaala Ajji, Phopo Razia and Phopa Imtiaaz (and families): A mere acknowledgment of your support for this PhD would be an understatement. Thank you for making my family's present life and our education possible at all.

My entourage of five, my people: Abbu, Ammi, Akifa, Ayesha, and Subuktageen without whom I do not exist.

List of Abbreviations

AI: Aromaticity index

APCI: Atmospheric pressure chemical ionization

APPI: Atmospheric pressure photoionization

BBOA: Biomass burning organic aerosol

CCN: Cloud condensation nuclei

DBE: Double bond equivalence

DOC: Dissolved organic carbon

DOM: Dissolved organic matter

ESI: Electrospray ionization

FT: Fourier transformation

ICR: Ion cyclotron resonance

IHSS: International Humic Substances Society

IN: Ice nuclei

LDI: Laser desorption ionization

MS: Mass spectrometer

MSNOM: Mississippi River natural organic matter

NLFA: Nordic Lake fulvic acids

NOM: Natural organic matter

OM: Organic matter

OSc: Oxidation state of C

OTE: Orbitrap elite

PLFA: Pony Lake fulvic acids

POA: Primary organic aerosol

RH: Relative humidity

SOA: Secondary organic aerosol

SRFA: Suwannee River fulvic acids

SRNOM: Suwannee River natural organic matter

TB: Tar balls

Tg: Glass transition temperature

Abstract

Complex environmental mixtures – be it particulate material in the atmosphere, decades old organic matter trapped within glaciers, or biological debris flowing with rivers and streams – are exquisitely complex. It is impossible to elucidate their exact molecular-level composition, but high-resolution mass spectrometry allows us to have a peek at the elemental compositions of thousands of chemical species that constitute these mixtures. In this dissertation, the overarching aim was to explore the molecular diversity of complex mixtures from two sources: Surface water and atmospheric organic aerosol.

The first objective of this dissertation was to demonstrate ionization selectivity of three popular ionization methods so that the necessity of using more than one ionization method for untargeted qualitative analysis of complex mixtures could be validated. Electrospray ionization (ESI), atmospheric pressure photoionization (APPI), and atmospheric pressure chemical ionization (APCI) were tested on commercial humic substances in combination with the Fourier Transform - Orbitrap Elite Mass Spectrometer. Our findings provide evidence for the tendency of ESI to access polar, more oxygenated compounds that form a majority of humic substances. A minor fraction comprising relatively less polar, aromatic compounds, could be accessed with either APPI or APCI, highlighting the importance of employing complementary ionization methods to obtain representative molecular compositions of complex mixtures.

The second objective of this dissertation was to demonstrate the extreme molecular complexity of organic aerosol collected downwind of wildfires in the Pacific Northwest of the United States. The focus was particularly on biomass burning organic aerosol that had aged to develop an abundance of tar balls (TB), which are carbonaceous spherules of extremely variable optical properties. Tar balls contribute to direct and indirect radiative forcing in the atmosphere. Their detailed molecular composition is yet to be elucidated; therefore, we made an attempt to find a preliminary TB-specific molecular signature by comparing several TB-rich and non-TB aerosol mixtures. Using Fourier Transform – Ion Cyclotron Resonance Mass Spectrometers of two resolving powers, as well as complementary ionization techniques, ESI and laser desorption ionization, we present detailed molecular composition of TB, which indicates them to be a mixture of low-oxygen organic constituents enclosed in a more oxidatively aged shell.

1 Introduction

1.1 Complex environmental mixtures

The natural environment is replete with organic matter (OM) of varying complexities. Natural OM may contain up to tens of thousands of molecular species, whose complexity is further enhanced by the presence of isomers. Individual components of complex mixtures cannot be chromatographically separated, nor can nuclear magnetic resonance spectroscopy decipher singular molecular species¹. Untargeted ultrahigh-resolution mass spectrometry permits resolution of molecules whose mass differs by as little as 0.5 mDa, i.e., the mass of an electron², and thus, presents itself as an attractive choice for comprehensive analysis of complex mixtures. Although, a great deal is left to be desired in terms of understanding structural and isomeric complexity with mass spectrometry, it is nonetheless an irreplaceable choice for deep diving into the molecular nature of complex mixtures. The principle of mass spectrometry is deviously simple, where the analyte is first converted to gas-phase ions, either cations or anions, by an ionization source³. These ions are then separated in a mass analyser by their mass-to-charge (m/z) ratio and detected in terms of ion intensity. The electrical signal detected from the ions is processed computationally to acquire the primary result of mass spectrometric analysis: A mass spectrum that comprises measured m/z and ion intensity.

This dissertation is a compilation of experiments performed on two complex mixture types: (i) atmospheric aerosol, particularly that have been influenced by wildfires and are rich in biomass-burning organic aerosol (BBOA), and (ii) fulvic and humic acid mixtures commercially prepared as standards by the International Humic Substances Society (IHSS).

1.1.1 Atmospheric aerosol

Atmospheric aerosols range from small droplets to solid particles suspended ubiquitously in the atmosphere and are classified as primary or secondary particulate matter based on their direct emission or formation from gaseous precursors, respectively⁴. Once emitted, they typically last from a day to two weeks in the troposphere and may persist up to a year or longer in the

stratosphere⁴. A large majority of atmospheric aerosols comprise inorganic species, including SO_4^{2-} , NO_3^- , and NH_4^+ , and sea salt⁵. Other smaller, but still very substantial constituents of aerosol are organic species, carbonaceous material called black carbon (BC), minerals trapped in dust particles, and primary biological particles. Mineral dust, sea salt, BC, and biological particles are primary emissions, where all except BC are naturally produced. Black carbon may originate from both anthropogenic and natural sources⁶. The remaining types of aerosol, SO_4^{2-} (from non-sea sources), NO_3^- , and NH_4^+ are predominantly products of secondary aerosol formation. Organic aerosol exists both as primary organic aerosol (POA) and secondary organic aerosol (SOA).

The reasoning behind exploration of the physics, chemistry, and biology of atmospheric aerosols is to better understand the way they influence life as we know on it on Earth. Atmospheric aerosol affects the climate in many ways, regardless of their anthropogenic or natural origins⁴. Effect of aerosols can be direct, which is controlled by their optical properties, whereby they absorb or scatter incoming solar radiation, altering Earth's radiative balance. Scattering of light off aerosols generally leads to enhanced reflectance of the Earth's atmosphere, and thus, causes cooling, while absorbance of radiation is associated with energy retention inside the atmosphere, and thus, warming. Aerosol have indirect effects on the climate via their role as cloud condensation or ice nuclei (CCN and IN), i.e., sites on to which cloud droplets or ice particles can take root and form^{7, 8}. Only a very small subset of all aerosol present in the atmosphere are capable of acting as CCN and an even smaller subset functions as IN. While some contradictory evidence has been made available, the current models largely provide evidence of aerosols to cool the atmosphere via formation of clouds⁴. An aerosol particle makes a good CCN as long as it can absorb water, be activated, and turn into CCN under conditions of supersaturation – this process is dependent upon the particle size and composition⁴. Similarly, aerosol particles are expected be good IN if they exist as solids even under supersaturated conditions. The optical and hygroscopic properties of individual aerosol particles are dictated by their chemical composition.

1.1.1.1 Wildfire-influenced organic aerosol: Chemical properties and atmospheric processing

Smoke produced from incomplete combustion has long been the most obvious proof of the existence of atmospheric aerosols⁹. These dark particles have been termed soot or smoke and refers to thousands of organic compounds, BC, and

inorganic components released from combustion¹⁰. Wildfires have immense destructive power that not only threatens animal and human lives, but also substantially deteriorates air quality¹¹. Regional and local influence of wildfires on the atmosphere is apparent in the trace gases, such as CO, NO_x, O₃, as well as particulate matter and BC observed after wildfires¹². In the last 10 years (2011-2020), ~62,693 wildfires were recorded annually in the United States that burned over an average of 7.5 million acres¹³. In 2020 alone, 10.3 million acres were impacted by 58,250 wildfires. Although more wildfires burn in eastern and central United States, only small acreage is impacted. A smaller number of large-scale wildfires impact the western United States every year^{13, 14}. In the western United States, during summertime, wildfires are a major determinant of organic aerosol concentrations¹⁵. An increased occurrence of wildfires is predicted for the future, owing to a warmer, drier climate, and changes in fuel availability⁴. Changed vegetation patterns and dynamics may play a role as well.

Carbonaceous aerosols that are produced from large-scale biomass burning events include organic matter with inorganic inclusions, BC, brown carbon, and soot¹⁶⁻¹⁸. Black carbon is generally primary aerosol and has received tremendous attention due to its ubiquitous occurrence, refractory nature, insolubility, strong capacity to absorb light, and its tendency to intermix with other aerosol^{4, 19}. It generally comprises of elemental C. Once released into the atmosphere, POA are prone to extensive chemical evolution. Most POA has enough volatility to evaporate after emission²⁰ leaving behind semi-volatile species (SVOC) that react with oxidants, such as OH•, O₃, and NO₃•, in the gas phase to form low-volatility oxidized compounds with a tendency to recondense into particulate phase within hours or days⁹. It is this atmospheric processing that causes O-containing compounds to be more prevalent in *aged* particulate organic emissions as compared to freshly released POA. Brown carbon, which is largely made up of SOA, has wide-ranging light-absorbing and -scattering implications in the ultraviolet (UV) and visible spectrum^{21, 22}. Important constituents of BC include organonitrates and organic amines²³⁻²⁶, polycyclic aromatic hydrocarbons²⁷, and methoxy phenols^{28, 29}.

Two key properties that change during evolution are volatility and oxidation state⁹, and thus, form the basis of models that describe and predict SOA formation³⁰. Furthermore, two of the most fundamental ways in which SOA continually evolves in the atmosphere are functionalization and fragmentation⁹. Oxidative processing of an organic species manifests as functionalization, i.e., introduction of functional groups, including -OH (hydroxyl), -CO (carbonyl), -ONO₂ (nitrooxy), -OOH (hydroperoxy), -C(O)OH (carboxyl), and -C(O)OR (ester). A popular model of SOA evolution, the volatility basis set^{30, 31}, assumes that products formed by functionalization have lower volatilities (Elaborated in Section 1.3.2) and higher oxidation states than the parent compounds and have

a greater tendency to condense into the particle phase. For fragmentation, this model assumes that C-C bonds may break at any position throughout the molecule, and the presence of functional groups may aid the process. As the oxygenation, i.e., number of O atoms in a molecule and its O/C, increases so does the tendency of a molecule to fragment by reaction with an atmospheric oxidant like OH[•]³². Owing to the random nature of fragmentation, SOA products are distributed over a wide range of oxidation and volatility. They may even have lowered propensity to exist in the particulate phase and may reconvert to the gas-phase like their precursor POA.

On conversion of SOA to the particle phase, its chemical evolution does not stop. Although much slower than gas-phase chemistry, particle-phase reactions continue to occur among the species within an aerosol particle, as well as between the aerosol and atmospheric oxidants that impinge upon it⁹. In heterogenous gas-aerosol reactions, an incoming gaseous atmospheric oxidant (OH[•], O₃, NO₃[•]) diffuses into the aerosol, followed by reaction with either the surface or constituents in the bulk. The rate constant for such a gas-particle reaction is much lower than when both species exist in the gaseous phases, therefore, condensed-phase SOA are rather protected from atmospheric oxidation. In other words, the lifetime of a species in the atmosphere increases greatly as its volatility decreases and it moves out of the gas phase and into the particle phase, allowing them to travel great distances from their site of emission across the troposphere³³. Another type of particle-phase reactions is non-oxidative accretion of singular species into high-molecular weight and low-volatility products. Figure 1.1 shows some major accretion reactions that take place in the particle-phase SOA. The rates of these reactions may increase under acidic conditions, but are not completely dependent upon acidity and may occur when no acids are present⁹. Importantly, some of the products from the reactions in Figure 1.1, such as peroxyhemiacetals, organosulphates, and esters are formed irreversibly and persist in their low-volatility, particle-phase forms.

A product of atmospheric evolution that has been associated until now specifically with BBOA is the formation of heavily carbonaceous, amorphous, light-absorbing spherules called tar balls (TB). They were relatively recently discovered by electron microscopy experiments in 2003-4 on organic aerosol originating from African wildfires. They were recognized as entities that were morphologically different from inorganic constituents and BC^{16, 34}. Although ample evidence is available for organic aerosols to exist as extremely viscous or semi-solid entities, TB appear to be exclusively solid. They are spherules reported to be <500 nm on average in diameter^{16, 24, 35, 36}. They have gained much attention due to their highly variable optical properties^{16, 37} that permit them to directly influence air quality, radiation budget, and climate, while their hygroscopic nature has been associated with indirect effects on climate, such as

via their potential to function as CCN³⁸. Tar balls may make up even >95% of the particulate mass downwind of large-scale biomass burning events^{24, 39-41}). They have been denoted as ‘processed primary particles’ to distinguish them from conventional particulate SOA^{24, 39}, and their formation was attributed to a loss of volatile species from a maturing plume³⁹. The fact that a wide range of absorption wavelengths have been reported for a variety of studies on TB⁴²⁻⁴⁴, it can be reasonably assumed that their molecular composition may vary from one source to the next. Studies using spectroscopic and microscopic approaches have extensively been carried out on ambient TB to elucidate their physical, optical, and bulk elemental/chemical analyses, but an understanding of their molecular composition is still limited. The work presented in this dissertation is the first comprehensive analysis to the best of our knowledge on the molecular composition of ambient TB-rich aerosol collected downstream of wildfires in the Pacific Northwest using high-resolution mass spectrometry.

Figure 1.1 The different types of accretion reactions that secondary organic aerosol may undergo in particle phase causing it to continually age in the atmosphere⁹

1.1.2 Aquatic humic substances

Dissolved organic carbon (DOC) content varies greatly between different types of aquatic ecosystems and even within the same type of waterbody. Generally, average DOC in groundwater, precipitation, oligotrophic lakes, eutrophic lakes, rivers, marshes, and bogs has been reported to be 0.7, 1.0, 2.0, 10.0, 5.0, 15, and 30 mg/L^{45, 46}. Only about 20% of all dissolved organic matter (DOM) in

natural waters is present in the form of identifiable compounds, such as carboxylic acids, carbohydrates, amino acids, etc., while a much larger majority is trapped within a complex soup of ‘undefined’ chemical species that are mostly degradation products or residue from dying plants, bacteria, fungi, etc.⁴⁵. This largest fraction of DOM in aquatic systems is predominantly acidic and has hydrophobic or hydrophilic fractions⁴⁷, where the former are operationally termed and defined as dissolved humic substances⁴⁶. Aquatic humic substances are coloured, polyelectrolytic, non-volatile organic acids that can be isolated with XAD resins employing acid-base reactions⁴⁵. Generally, their elemental composition has been constrained within 50% C, 35-40% O, 4-5% H, 1-2% N, and <1% S with molecular weight ranges of 50-5000 Da. These ranges are highly variable from one source to the next depending upon major influences faced by the aquatic reservoir from which humic substances are collected. Major functional groups in this soup of organic acids are carboxyl, phenolic, carbonyl, and hydroxyl groups. Within humic substances, two classes are defined – again, operationally – based on solubility under different pH conditions. The fraction of humic substances that precipitates at pH <2.0 is termed humic acids and constitutes ~10% of DOM, while the fraction that remains soluble at all pH values is called fulvic acids and makes ~40% of DOM⁴⁵.

1.1.2.1 Commercial organic matter standard/reference mixtures

In his classic textbook, “Organic Geochemistry of Natural Waters,” Michael Thurman writes: “*No reagent labelled aquatic humic substances exists, and there are no simple methods of analysis for aquatic humic substances*”⁴⁵. Currently, commercial mixtures prepared with immense uniformity from the same source over years are the closest we can get to realizing such a ‘reagent’. In 1976-78, MacCarthy and Malcolm highlighted the increasing interest of researchers from varied scientific backgrounds, such as physics, chemistry, biology, geochemistry, and agriculture, in environmental organic matter mixtures. They emphasized the urgency of integrating the extensive knowledge that was being gathered on humic substances from different perspectives^{48, 49}. According to the authors⁴⁸, “It is virtually impossible to objectively and critically compare the laboratory results of different workers in this interdisciplinary area of research...These complications could be minimized to a considerable degree if many researchers were to work with identical samples of humic materials obtained from a single distributing source.” As a sequel to their proposal, the International Humic Substances Society (IHSS) was founded that eventually established a reference collection of humic and fulvic acids – later categorized into reference and standard mixtures – from soil, leonhardite, peat, and aquatic environments⁵⁰.

Since their introduction, the chemical, physical, and molecular nature of IHSS mixtures has been studied extensively, beginning in 1989 from the compilation of studies on humic substances in Suwannee River⁵¹. This was closely followed by detailed characterization of standard fulvic and humic acids from Suwannee River, soil, and peat; standard humic acids from Leonhardite; reference fulvic and humic acids from Lake Hellrudmyra; and reference humic acids from Summit Hill soil by ^{13}C and ^1H NMR analyses⁵².

Currently (as of June 2020), the IHSS offers eight standard fulvic and humic acid mixtures and two NOM reference mixtures from soil, riverine, peat, and Leonhardite sources⁵³, which have been crucial to the field of environmental mass spectrometry, IHSS mixtures have been crucial in understanding instrumental optimization⁵⁴⁻⁵⁶, method development⁵⁷⁻⁵⁹, ionization selectivity, and considerations for best analytical approaches⁶⁰⁻⁶². Notably, these mixtures have satisfied their intended purpose, in general, for analytical purposes by functioning as a proxy for positive control in studies focused on untargeted characterization of complex mixtures of organic matter⁶³⁻⁶⁶. In such reports, findings about the unknown complex mixtures from a variety of analytical tools were validated by comparing compositions delineated for IHSS references and standard mixtures against the ‘expected compositions.’ Inclusion of an IHSS mixture allows ‘apparent’ molecular characteristics observed for complex environmental mixtures (due to discrepancies in methodology and instrumental parameters) to be reliably differentiated from the ‘true’ characteristics, i.e., if observations for an IHSS reference or standard mixture sway radically from expected patterns, the conclusions that are drawn for an unknown sample of interest may also be rendered suspicious. In this dissertation, we have used six different IHSS reference mixtures enlisted with their properties in Table 1.1 with the aim of elucidating their detailed molecular composition with the Fourier-Transform Orbitrap Elite Mass Spectrometer (FT-OTE MS). We have also used them to demonstrate ionization selectivity of three atmospheric pressure ionization methods (Electrospray, chemical, and photoionization) in positive and negative polarities.

Table 1.1 Origin, product code, and %(W/W) elemental compositions of dry ash-free dissolved organic matter mixtures reported by the International Humic Substances Society (IHSS; <http://humic-substances.org/elemental-compositions-and-stable-isotopic-ratios-of-ihss-samples/>). Elemental ratios were calculated based on bulk content.

	IHSS category number	C	H	O	N	S	H/C	O/C
SRFA	2S101F	52.34	4.36	42.98	0.67	0.46	0.99	0.61
This fulvic acid mixture is collected from terrestrially-influenced blackwater of Suwannee River that originates from the Okefenokee Swamp in South Georgia, USA								
PLFA	1R109F	52.47	5.39	31.38	6.51	3.03	1.22	0.45
This microbially-influenced fulvic acid is collected from a hypereutrophic, saline coastal pond on Cape Royds, Antarctica.								
NLFA	1R105F	52.31	3.98	45.12	0.68	0.46	0.90	0.65
This fulvic acid mixture is collected from a small tarn with blackwater in a relatively isolated catchment near Oslo, Norway								
ESFA	5S102F	51.92	3.91	39.64	2.40	1.43	0.90	0.57
This fulvic acid mixture is collected from a fertile prairie soil from an undisturbed area on the Joliet Army Ammunition Plant near Joliet, Illinois.								
SRNOM	2R101N	50.70	3.97	41.48	1.27	1.78	0.93	0.61
This humic and fulvic acid mixture is collected from terrestrially-influenced blackwater of Suwannee River that originates from the Okefenokee Swamp in South Georgia, USA.								
MSNOM	1R110N	49.98	4.61	41.4	2.36	2.62	1.10	0.62
This humic and fulvic acid mixture is collected from Mississippi River in Minneapolis, Minnesota, USA.								

The second alphabet in the category number indicates whether the sample is an IHSS ‘reference’ or ‘standard;’ they are denoted by R and S, respectively. They differ in the extent of processing that they are subjected to. Further details are available at the IHSS website: <http://humic-substances.org/source-materials-for-ihss-samples/>

1.2 High-resolution mass spectrometry

1.2.1 Principles of operation

A mass spectrometer, regardless of its level of analytical sophistication, has the same fundamental components, albeit the principle of operation to resolve ions and measure their masses may differ incredibly between different types of mass spectrometers. As outlined by³ and depicted in Figure 1.2, there is (i) a sample inlet that can be a direct insertion probe, a gas chromatograph, or a liquid chromatograph, (ii) an ionization source that converts the sample to gaseous ions, (iii) one or more mass analysers that separate ions, (iv) a detector that counts the ion groups separated by the mass analyser, and (v) a data processor to yield the mass spectrum. Ions are separated in the mass analyser by their m/z ratios, where m is the relative mass of the ions and z denotes charge number³. The mechanism by which this is done depends upon the type of mass analyser used. The two types of mass analysers used in this dissertation are discussed in Section 1.2.3. Ion separation can take place only under low-pressure conditions or complete vacuum to ensure minimal collision of ions with one another, air molecules, and with the walls of the instrument on their way to the detector³. After signal processing, the abscissa of the mass spectrum represents the measured m/z of the ions and the ordinate shows ion intensities³. The most intense peak in the mass spectrum is called the base peak and is arbitrarily assigned an intensity of 100%. The intensities of all other peaks originating from the analyte are measured as percentages relative to the base peak.

In case of simpler analytes, such as pure compounds, ionization methods that break the parent ion into fragments are preferable. The fragment ions shed light on the identity of the analyte. In the case of complex mixtures, fragmentation leads to an unnecessary complication of an already complex mixture, therefore, soft ionization techniques that impart slight to no fragmentation of individual molecules are favoured. All ions identified in such mixtures are so-called parent ions and provide information on the bulk chemical composition, as well as molecular components, of the mixture. Although a repertoire of ionization methods are now available to choose from, for the purpose of this dissertation, four ionization methods were used: electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), atmospheric pressure photoionization (APPI), and laser desorption ionization (LDI), which are elaborated below.

Figure 1.2 A simplified schematic representation of the basic components of a mass spectrometer³

1.2.2 Ionization methods

1.2.2.1 Electrospray ionization

Electrospray ionization (ESI) is the most widely popular method of atmospheric pressure ionization⁶⁷. It has been extensively applied, particularly in negative ion mode to analyse ambient aerosol from a variety of sources, including rain water, fog, OA from the troposphere⁶⁸⁻⁷⁴, as well as humic substances⁷⁵. Electrospray is a soft ionization technique that is best suited for polar, non-volatile molecules of very small (10 Da) to very large sizes, such as proteins. Electrospray ionization works by application of a strong electric field on liquid-state sample as it passes through a capillary at a rate of 1-10 $\mu\text{L}/\text{min}$. The electric field is generated by placing the capillary at a potential difference of 3-6 kV from a counter-electrode³. Charge accumulates in the liquid located at opening of the capillary, which elongates the liquid into a cone like shape extending from the capillary toward the counter electrode; this is called the Taylor cone⁷⁶. Eventually, the surface tension is not enough to hold the liquid together, and it breaks it into smaller highly charged droplets. The droplets are made to pass through a region of heated inert gas, such as N_2 , or through a heated capillary until all solvent

evaporates. As each of the droplets loses solvent, the charge accumulated on it strengthens leading to further breakdown into even smaller droplets. This process continues until only monocharged ions remain. For molecules larger than a ~1000 Da, this principle of ionization allows a single molecule to gain multiple charges as well. Mass spectrometers measure ions in terms of their m/z ratios, therefore, ESI is able to access molecules that are otherwise too large (50,000 Da) to be detected by other ionization techniques.

1.2.2.2 Atmospheric pressure chemical ionization

Atmospheric pressure chemical ionization (APCI) is a variant of the conventional chemical ionization method, except that it is performed at atmospheric pressure. It has been used for organic aerosol analysis⁷⁷⁻⁷⁹. In this method, no spray voltage is applied. Instead, a pneumatic nebulizer is used⁶⁷. Analyte solution is introduced into the nebulizer at a rate of 200-2000 $\mu\text{L}/\text{min}$ to be transformed into a thin vapour by application of a high-speed N_2 beam³. The aerosol thus formed is passed through high temperature (500°C) by a heater cartridge (generally a quartz tube) for evaporation of solvent. Once desolvated, ions are carried through a corona discharge electrode for ionization. The corona discharge comprises of primary ions from electron ionization, i.e., $\text{N}_2^{+\bullet}$ and $\text{O}_2^{+\bullet}$. In most settings, the evaporated solvent is converted to secondary reactant ions under the effect of the corona discharge. Positive molecular ions are generated by proton transfer or adduction of reactant gas, while proton abstraction and adduction may form negative molecular ions. Like ESI, APCI is also a soft ionization technique that produces minimal fragmentation of molecular species³. The major advantage afforded by APCI over ESI is the ionization of neutral species; while ESI is limited to polar species, APCI can readily access low-polarity species as well⁶⁷. It is useful for species up to 1500 Da and produces monocharged ions only³.

1.2.2.3 Atmospheric pressure photoionization

Atmospheric pressure photoionization (APPI) is a soft ionization technique complementary to APCI and ESI. It has been used extensively for analysis of organic aerosol and humic substances on its own⁸⁰⁻⁸³ as well as in combination with other ionization techniques^{41, 70}. In place of corona discharge-powered plasma used in APCI, a UV light source is used in APPI⁶⁷. A similar arrangement of

the pneumatic nebulizer and heated vapourizer are used. Upon vapourization, the analyte is subjected to photons from a discharge lamp, which leads to a series of gas-phase reactions causing ionization of the analyte³. The most common photon source for APPI is a krypton discharge lamp that produces photons with energy in the UV range. The lamp is chosen based on the expected ionization potentials of the analyte and solvents. Ideally, the photons should have higher energy than the ionization potential of the analytes and lower energy than the ionization potential of the solvent molecules. Most analytes, such as organic molecules have ionization energies between 7 and 10 eV, which makes the krypton discharge lamp an ideal choice for APPI as it emits photons of 10-10.6 eV³. A caveat in APPI is that photons from the discharge lamp have more probability to be absorbed by the solvent molecules or by the nitrogen spraying gas rather than by the analyte molecules, reducing the likelihood of analyte ions to be ionized⁶⁷. Although the solvent may only be photoexcited and not ionized, it reduces the number of photons that are available to ionize the analyte. To improve photoionization, an additional chemical called dopant is added that can be more readily photoionized and ionize the analyte by charge transfer. Common APPI dopants are toluene and acetone. Common mechanism of ion formation in positive mode in APPI is proton adduction or formation of radical cation, while abstraction of proton from molecular ion is the most likely means of ion formation in negative mode³. As compared to APCI and ESI, APPI is very sensitive to experimental conditions, such as the solvents, additives, or buffers used. However, it has a much higher capability to ionize non-polar compounds that may be undetectable by ESI and APCI.

1.2.2.4 Laser desorption ionization

Laser desorption ionization (LDI) was introduced in the 1960s as a useful and simple technique for analysis of solid samples⁶⁷. Simply, laser light is subjected upon a solid layer of sample at ambient temperature transferring energy to the sample and, eventually, causing its evaporation and ionization. The laser pulses are generally 10^6 to 10^7 W/cm²³ and are focused on a very small region, 10^{-3} to 10^{-4} cm², of the sample to allow it to ablate. Diameter of laser beam at the site of ablation can be varied to obtain better spatial resolution and insight into sample composition. When an analyte is ablated off of a substrate surface, a microplasma containing neutrals and ions is generated in a dense cloud above the sample surface. The constituents of this microplasma react among themselves in the vapour phase forming analyte ions. The most common lasers used have ultraviolet (UV) or infrared (IR) wavelengths. Laser desorption ionization is especially suitable for analysis of organic compounds and molecules

containing large π -electron systems, polymers that absorb in the UV and IR wavelength ranges⁶⁷. It must be noted that although LDI is a soft ionization technique⁶⁰, it presents a much higher possibility of fragmentation as compared to matrix-assisted LDI, as well as ESI, APCI, and APPI⁶⁷, especially for molecules above 500 Da.

1.2.3 Mass analysers and signal processing

Once generated by the ion source, ions are transferred into the mass spectrometer. An interface is required to move ions from atmospheric pressure to the high-vacuum (10^{-5} torr) environment within the mass analyser. This is done via differential pumping systems to create a buffer during the move of ions across a huge pressure difference³. Usually, two or more intermediate vacuum compartments are present between the ionization source and the mass analyser (Figures 1.4 and 1.6). The compartments are arranged with the most high-pressure compartment toward the ionization source with gradually reducing pressure till the last compartment, which is before the mass analyser. These compartments are connected to one another by lenses containing orifices for ion path. The size of the lens orifices has to be carefully chosen; if the orifice is too wide, vacuum cannot be maintained effectively, and if it is too small, enough number of ions cannot pass through. Ion transfer optics systems, such as focusing lenses or focusing multipole lenses, are responsible for effective injection of ions through these orifices and finally into the mass analyser.

The projects presented in this dissertation made use of two types of high-resolution mass analysers: An Orbitrap Elite housed at the Chemical Advanced Resolution Methods (ChARM) Laboratory at Michigan Technological University and Fourier transform - Ion Cyclotron Resonance mass spectrometers located at the Pacific Northwest National Laboratory, Washington. Their basic principles of function are briefly reviewed here.

1.2.3.1 Orbitrap Elite

The orbitrap is a sophisticated Kingdon trap that employs an electrostatic field to trap moving ions. The original Kingdon trap developed in 1923 was a simple set up with a cylindrical electrode enclosing a wire running across its axis⁶⁷. The inner wire is set to an electric potential; then, ions that tangentially enter the

region between the cylindrical electrode and wire are trapped in a rotational motion around the wire under the influence of centrifugal and electrostatic forces. The Orbitrap uses the same principle for trapping ions but has an improved physical design. As depicted in Figure 1.3, the Orbitrap has a central spindle-shaped electrode and a barrel-shaped outer electrode that is split into two halves. Once injected into the Orbitrap, ions adopt a motion around the central electrode that resembles satellites in motion. Ions in stable trajectories rotate around the axis of the central electrode while oscillating along the axis. Amplitudes of ion packet rotation is the same regardless of their m/z , but their frequency of oscillation (ω ; unit of rad/s) is mathematically represented by the Equation 1.1. Axial oscillations are detected and recorded as current image by an amplifier connected to the outer electrode. The image current signal thus generated is subjected to Fourier transformation (briefly discussed in Section 1.2.3.3) to frequency domain signal that can be inputted in Equation 1.1 to obtain the m/z value.

$$\omega = \sqrt{k(z/m)} \dots \dots \dots \text{Equation 1.1}$$

Figure 1.3 A longitudinal view of an Orbitrap showing ions in motion around the z axis of the central electrode (a). An image current (c) is induced on the outer electrode halves (b) is detected via an amplifier. The current from different ion packets in deconvoluted and converted to their respective frequencies via Fourier-transformation⁸⁴

Orbitrap instruments were commercialized in 2005 by Thermo Fisher Scientific and since then have undergone many technical improvements⁸⁵. The mass spectrometer used for purposes of this dissertation houses an Orbitrap Elite,

which is a high-field version the Orbitrap, introduced in 2012. The rate of oscillations increases with increasing field strength, which leads to enhanced resolving power. The Orbitrap Elite's high field strength was achieved by reducing the trap dimensions and enhancing the Fourier-transformation⁸⁵. This new Orbitrap Elite mass analyser is combined with a Velos Pro dual cell differential pressure linear ion trap mass spectrometer. Instrumental schematics are depicted in Figure 1.4. A set of stainless-steel apertures called S-lens use resonance frequency voltage to focus the ion beam into a curved beam blocker⁸⁵. Neutral droplets and/or solvent clusters are not allowed to reach downstream ion optics. Ions move on to dual cell differential pressure linear ion trap, where nominal masses can be measured at faster scan speeds. Ions can also be directed into the OTE for measurement at resolution of 240,000 at 400 m/z via a C-shaped quadrupole trap called the C-trap⁸⁴. The HCD collision cell shown in Figure 1.4 is designed to perform MS/MS or MSⁿ experiments and are not relevant to the purposes of this dissertation.

Figure 1.4 Schematic representation of a Thermo Fisher Orbitrap Elite Mass Spectrometer, including ionization source at the front end, ion transfer optics, linear quadrupole ion trap, and the high-resolution Orbitrap Elite itself⁸⁵

1.2.3.2 Fourier transform - Ion Cyclotron Resonance

The Fourier transform - Ion Cyclotron Resonance (FT-ICR) mass analyser is based on the principle that the trajectory of ions is curved under the influence of a magnetic field perpendicular to their velocity, forcing ions to move in a circular path^{3, 67}. Provided that the ions are in slow motion and the magnetic field is strong, the radius of ion's trajectory around the axis of the magnetic field becomes small enough that the ions are trapped³. Stable ion trajectory is a result of balanced centripetal and centrifugal forces, whose mathematical

treatment yields Equation 1.2, where ω is the angular cyclotron frequency, ν is the cyclotron frequency, q is the charge (in Coulombs), m is mass (in kg), and B (in Tesla) denotes the magnetic field strength. As ν can be measured with the highest accuracy, FT-ICR-based mass spectrometers are powerful instruments to yield accurate m/z values of even the most complex mixtures.

$$\omega = 2\pi\nu = B(q/m) \dots \dots \dots \text{Equation 1.2}$$

As depicted in Figure 1.5, ions revolving around the central axis, which is the magnetic field, are excited with an RF wave (i.e., a transverse electric field alternating with respect to ν) that causes their kinetic energy, and thus, trajectory radius to increase. This electric field is applied using two electrode plates placed on opposite ends of the ion orbit (Figure 1.5)⁶⁷. As ions accelerate under the influence of this field, their orbital radius increases until it becomes a spiral. Ions of the same m/z maintain orbit as tight packets, i.e., remain in-phase with the same rotational frequency, and are detected as an image current over time. The time-domain signal thus obtained is subjected to deconvolution and conversion to frequency using Fourier transformation. Exact m/z values are obtained by plugging isolated frequencies in Equation 1.2.

Figure 1.5 A schematic representation of an FT-ICR cell showing its mechanism of action and the trajectory of ions with respect to electric and magnetic fields⁸⁶

For the purposes of this dissertation, two FT-ICR MS instruments at the Environmental Molecular Sciences Laboratory, part of the Pacific Northwest National Laboratory in Richland, WA, United States, were used. We employed a Bruker 15-T SolariX FT-ICR MS and a 21-T FT-ICR MS custom-built to analyse aerosol mixtures. Resolving powers and scan speeds are linearly proportional to the magnetic field strength, while the mass accuracy and dynamic range are quadratically related to it⁸⁷. The 21-T FT-ICR MS is connected to a Velos Pro dual linear quadrupole ion trap MS at the front end (Thermo Fisher Scientific, San Jose, CA, USA)(Figure 1.6)⁸⁸, which makes the instrument highly sensitive, efficient at ion isolation, and also provides MS/MS capabilities with collision induced dissociation. This part of the instrument allows modulation of the ion population size sent to the ICR cell via the automatic gain control function. The second major component of the 21-T FT-ICR is a linear quadrupole ion trap that functions as an intermediate storage device before ions are transferred from the Velos Pro to the ICR cell.

Figure 1.6 Schematic representation of a 21-T FT-ICR MS housed at the Environmental Molecular Sciences Laboratory of the Pacific Northwest National Laboratory depicting front end ion-trap, ion optics, and the FT-ICR cell⁸⁸

1.2.3.3 Fourier-transformation of image current signals

Jürgen H Gross⁶⁷ define Fourier transformation as “a mathematical operation that transforms one complex-valued function of a real variable into another.” In case of the Fourier transform – mass spectrometry, the complex function that needs to be transformed is ‘time,’ and thus, is termed time-domain function. The

function produced as a result is ‘frequency’ and is called frequency-domain function. It is essentially only a different representation of the original function. This can be visualized in Figure 1.7, where if a single chord of music were to be considered to be the original function that contains several frequencies, the application of Fourier-transform will separate all frequencies and present their amplitudes.

Figure 1.7 Fourier transformation: (A) The principle of Fourier transformation depicted in terms of musical notes. A sound signal measured over time constitutes many sound frequencies of a variety of intensities superimposed on one another. Fourier transform is designed to separate each frequency from the ‘sound mixture’ with the correct intensity, (B) signal intensity that is initially a function of time becomes signal intensity as a function of frequency, which is then converted to m/z as shown in (C)³

1.2.4 From measured mass to molecular formula

The meaning and usage of the term, “mass” (of an atom, molecule, or ion etc.), are manifold in chemistry and depend on the context. While *average mass* is the prevalent term in stoichiometric chemistry that is calculated from the weighted average of atomic masses of all atoms present in a molecular entity, mass spectrometry deals with the concept of mass with much more finesse. A clear distinction is made between nominal, exact, monoisotopic, and polyisotopic masses. *Nominal mass* is a whole number that is calculated by summing the mass of the predominant isotope rounded to the nearest integer, i.e., nucleon number, of each element in the molecule under consideration³. *Exact mass* is calculated by summing the masses of isotopes constituting the molecule under

consideration. It is further categorized as *monoisotopic mass* if masses of only the most abundant isotopes are summed and *polyisotopic mass* is calculated in the same way for polyisotopologues. Exact masses differ slightly from the summed masses of unbound electrons, neutrons, and protons. This difference is attributed to the binding energy holding these particles together and is called the *mass defect*³. However, in the context of mass spectrometry, mass defect is the mathematical difference between the exact mass and the nominal mass.

The inherent function of mass spectrometers is to separate or identify ions based on their masses, and the *type* of mass a mass spectrometer measures depends upon the resolving power of its mass analyser. Resolving power of a mass spectrometer is defined as the full width of a peak at half maximum height, where narrower peak shape indicates higher resolving power³. High-resolution mass analysers can differentiate and measure with high accuracy the monoisotopic and polyisotopic masses, which would generally remain unresolved and appear as a single broad peak with low-resolution mass analysers. High-resolution instruments produce very narrow peaks with their centre very close to the *actual mass* of the ions; they have been reported to separate molecular species as close to even 0.5 mDa in mass². Such analytical capabilities are invaluable for the analysis of complex mixtures. Complex mixtures, such as those presented in this dissertation, are composed of C, H, O, S, N, and P, all of which can occur in a multitude of combinations. Multiple exact masses can occur at the same nominal mass. Ions of different elemental compositions are present at characteristic mass differences enlisted in Table 1.4 and need specific resolving powers ($m_2 - m_1 / m_2$) to be separated from one another⁸⁹. The smaller the mass difference, the higher the resolving power required to separate those two ions.

Table 1.4 Characteristic mass differences at the same nominal mass originating from differences in elemental composition

Composition 1	Composition 2	Mass difference	Resolving power required at 499 <i>m/z</i>
CH ₄	O	36.39 mDa	14,000
C ₄	O ₃	15.26 mDa	33,000
S	O ₂	-17.76 mDa	28,000
N ₂	CO	11.23 mDa	45,000
SH ₄	C ₃	3.37 mDa	150,000
PH	O ₂	-8.24 mDa	61,000
N	¹³ CH	-8.15 mDa	61,000

In untargeted complex mixture analysis, where a mass spectrometer yields hundreds to thousands of exact masses, only attaining the mass is not beneficial. The fruit lies in the fact that these masses, if measured with sufficient accuracy, can be assigned an elemental composition. Numerous formula assignment tools are available, which differ in the type of complex mixture they are designed for, such as Composer by Sierra Analytics and PetroOrg are optimized for petroleum analysis. These tools differ in the way they assign formulae (brute force calculation, database matching, homologous series extension, etc) and test the equivocality of the formulae assigned⁹⁰⁻⁹³. Here, an in-house formula assignment tool, MFAssignR⁹³, written in R language was employed for all data analysis and processing.

1.2.4.1 MFAssignR

MFAssignR is publicly available as an R package comprising numerous customizable functions (represented with closed parentheses following their titles here) for noise removal, isotope filtering, mass recalibration, and formula assignment⁹³. Baseline noise can be estimated using KMDNoise(). This function computes the CH₂ Kendrick mass defect of all ions, then a user-selected slice of data with random mass defects is identified. The average intensity of peaks in the selected slice is returned as the estimated noise applicable to the raw spectrum. The second step in data analysis with MFAssignR is the application of IsoFiltR() to identify and separate mono- and poly-isotopic peaks. The function compares all masses with one another to identify expected mass differences between monoisotopic and polyisotopic peaks, as well as by looking for isotopic intensity patterns. MFAssignCHO() is then used to obtain preliminary assignments, which are used to select recalibrant ions. The recalibrant ions are used in Recal() function to correct the entire mass list from any systematic bias or mass shifts. The final step is to assign formulae using MFAssign() to all mono- and polyisotopic peaks after noise removal and mass recalibration. MFAssignR performs rigorous QA on the assigned formulae by taking into consideration fundamental rules of chemical feasibility, such as nitrogen, Senior, and large atom rules etc., as well as user-defined parameters, including isotope matching for sulphur assignments and CH₄ assignment instead of O, where appropriate, with the SulfCheck and NMScut options, respectively⁹³. The output is generated in .csv format and contains several important pieces of information, notably the unambiguous neutral formula assigned to each mass, any ¹³C₁₋₂³⁴S species identified, the number of C, H, O, N, S, and P in each formula, and double bond equivalence, etc.

1.2.4.2 Computation of estimated structural and physicochemical parameters

Many computational models have been developed that employ experimental data generated from high-resolution mass spectrometric analyses, specifically, numbers of atoms of different elements in molecular species identified. As detailed here, these models estimate physicochemical parameters, such as viscosity, saturation mass concentration, glass transition temperatures, etc. and even provide insight into the structural characteristics of complex organic mixtures, such as number of double bond equivalences or aromaticity. According to Seinfeld and Pandis⁹, “The oxidation of a VOC to yield products of low volatility is the hallmark of SOA generation.” The first of these important SOA properties, i.e., oxidation levels, can be computed as average oxidation state of C (OS_c) in a molecule using a model developed specifically for atmospheric organic aerosol by Kroll et al.⁹⁴. The original expression considers molecules containing H and O only, therefore, the average OS_c is calculated using Equation 1.3. However, in our studies, other elements (N, S, and P) are considered as well, therefore, a modified version (Equation 1.4) is used.

The second important property of SOA highlighted by Spyros and Pandis, volatility, is estimated in terms of saturation mass concentration (C^0 ; units: $\mu\text{g}/\text{m}^3$), which is a key thermodynamic property dictating the gas-particle partitioning of compounds. A number of estimation methods are available with the most notable being the linear correlation developed by Donahue et al.³⁰ for organic aerosol, which was extended further by Li et al.⁹⁵ to include heteroatoms (N and S; Equation 1.5). This saturation mass concentration is estimated under conditions of ideal thermodynamic mixing⁹⁵. In this parameterization, n^0C is the reference C number. bC , bO , bH , bN , bS are coefficients denoting the contribution of each atom to the $\log_{10}C^0$. Lastly, bCO is the C-O nonideality. The element symbol denotes the number of atoms of that element in a given molecule. The values of these coefficients as computed by Li et al.⁹⁵ differ between molecular groups and are presented in Table 1.2. The computed values are fit into saturation mass concentration bins. Molecular formulae with $\log_{10}C^0 \leq -3.52$ are categorized as extremely low volatility (ELVOC), $-3.52 < \log_{10}C^0 \leq -0.52$ as low volatility (LVOC), $-0.52 < \log_{10}C^0 \leq 2.47$ as semi-volatile (SVOC), $2.47 < \log_{10}C^0 \leq 6.47$ as intermediate volatility (IVOC), and those with $6.47 < \log_{10}C^0$ as volatile organic compounds (VOC). These volatility bins allow one to make informed speculation about the physical existence of a given species in the atmosphere. Compounds classified as ELVOCs are expected to immediately partition into particle phase upon formation and act as good nuclei for new

particle formation in the atmosphere, while LVOCs typically exist in the particle phase at atmospherically relevant concentrations⁹. Compounds identified to be SVOCs exist in both the particle and gas phase under typical ambient conditions. The remaining two classes, IVOCs and VOCs, exist the gas phase, but the former may easily condense to particle phase on being oxidized.

$$OSc = 2\frac{O}{C} - \frac{H}{C} \dots \dots \dots \text{Equation 1.3}$$

$$OSc = 2\frac{O}{C} - \frac{H}{C} - 5\frac{N}{C} - 6\frac{S}{C} - 5\frac{P}{C} \dots \dots \dots \text{Equation 1.4}$$

$$\log_{10}C^0 = (n^0C - C)bC - (O \times bO) - (H \times bH) - 2\frac{(C \times O)}{(C + O)} - (N \times bN) - (S \times bS) \dots \dots \dots \text{Equation 1.5}$$

Table 1.2 Values of n⁰C and other coefficients proposed by Li et al.⁹⁵ for use in Equation 1.5 to calculate log₁₀ of saturation mass concentration (C⁰) using data from high-resolution mass spectrometric analyses

Molecular group	n ⁰ C	bC	bH	bO	bCO	bN	bS
CH	17.95	0.5742	-0.142				
CHN	23.01	0.4307	-0.021			0.9528	
CHO	15.77	0.6238	-0.139	1.735	-0.8592		
CHNO	21.12	0.4139	-0.038	0.8092	-0.1174	1.1010	
CHOS	16.07	0.5348	-0.151	1.354	-0.4175		0.899
CHNOS	19.20	0.5469	-0.137	1.183	0.07310	1.0289	1.323

Another important parameter of atmospheric organic aerosol is their phase state. The phase state and viscosity of particulate matter determine its readiness to undergo kinetic and physical transformations⁹⁶. For instance, a viscous glassy particle that does not readily undergo gas-phase partitioning may be more prone to long-range transport and will be resistant to atmospheric processing. In view of the implication of phase state, a parameterization was developed to predict glass transition temperatures (T_g) under dry⁹⁶ and humid⁹⁷ conditions. The T_g is a temperature that is characteristic of chemical species at which they transition from a glassy/solid state to a semi-solid state if the temperature is increased⁹⁸. Models for both T_{g,dry} and T_{g,RH} use data from high-

resolution mass spectrometry, i.e., number of C, H, and O assigned to a molecular species based on its measured mass. The T_g of a compound differs under dry and humid conditions as water is a plasticizer, which lowers the T_g of a compound under humid conditions. The parameterization used in this study for $T_{g,dry}$ ⁹⁶ and $T_{g,RH}$ ⁹⁷ are presented in Equations 1.6 and 1.7. Here, n^0C is the reference C number, bC , bH , and bO represent the numeric contribution of these elements to T_g . Also, bCH and bCO are coefficients that reflect contributions from C-H and C-O bonds, respectively. Their values are given in Table 1.3. In Equation 1.7, $T_{g,w}$ is the glass transition temperature of pure water, i.e., 136 K, and K_{gt} is the Gordon-Taylor constant that is equal to 2.5 ± 1.5 . Here, w_{org} is the mass fraction of organics and was calculated using Equation 1.8⁹⁶, where RH is the measured relative humidity on the day and site of sample collection. If the computed $T_{g,RH}$ is higher than the ambient temperature (T_{amb}) at the site and day of collection, a molecular species is estimated to be solid. Any species with $0.8 \times T_{amb} < T_{g,RH} < T_{amb}$ are expected to exist as semi-solids while those with $T_{g,RH} < 0.8 \times T_{amb}$ are expected to exist as liquids.

$$T_{g,dry} = (n^0C + \log C)bC + (\log H \times bH) + (\log C \times \log H \times bCH) + (\log O \times bO) + (\log C \times \log O \times bCO) \dots \text{Equation 1.6}$$

$$T_{g,RH} = \frac{(1 - w_{org})T_{g,w} + \left(\frac{1}{K_{gt}} \times w_{org} \times T_{g,dry}\right)}{(1 - w_{org}) + \left(\frac{1}{K_{gt}} \times w_{org}\right)} \dots \text{Equation 1.7}$$

$$w_{org} = \frac{1.4 - \left(\frac{1.4 \times RH}{100}\right)}{1.4 - \left(\frac{1.28 \times RH}{100}\right)} \dots \text{Equation 1.8}$$

Structural characteristics of species identified in a complex mixture can be estimated using two simple parameters, double bond equivalence (DBE) and the aromaticity index (AI_{mod})^{99, 100} presented in Equations 1.9 and 1.10. The first parameter calculates the number of double bonds and rings present in a molecule. A decreasing number of H atoms leads to higher DBE, but this parameter does not consider elements with multiple valences, such as S; it uses the lowest possible valence state. The AI_{mod} is a relatively newer measure that considers the possibility of heteroatoms forming double bonds that do not contribute to aromaticity, ringed structures, or the overall H-saturation of a

molecule^{99, 100}. This parameter is presented in Equation 1.10. The AI_{mod} takes into account that some heteroatoms, particularly O, may form double bonds with C and are considered in DBE calculation. This parameter can be used to classify molecules into four structural classes: Aliphatic ($AI_{mod} < 0$), olefinic ($0 < AI_{mod} \leq 0.5$), aromatic ($0.5 < AI_{mod} < 0.67$ and $C \leq 8$), and condensed aromatic ($AI_{mod} \leq 0.67$; and $C \leq 8$).

$$DBE = 1 + \frac{1}{2}(2C - H + N) \dots\dots\dots \text{Equation 1.9}$$

$$AI_{mod} = \frac{1+C-0.5(O)-S-0.5(H)-0.5(N)-0.5(P)}{C-0.5(O)-S-N-P} \dots\dots\dots \text{Equation 1.10}$$

Table 1.3 Values of n^0C and other coefficients proposed by DeRieux et al.⁹⁶ for use in Equation 1.6 to calculate glass transition temperatures under dry conditions ($T_{g,dry}$) using data from high-resolution mass spectrometric analyses. Values in brackets denote uncertainties.

Molecular group	n^0C	bC	bH	bO	bCH	bCO
CH	1.96 (1.81)	61.99 (53.65)	-113.33 (44.47)		28.74 (20.86)	
CHO	12.13 (2.66)	10.95 (13.60)	-41.82 (14.78)	118.96 (9.72)	21.61 (5.30)	-24.38 (4.21)

The estimations produced from these models are approximate; for instance, when the uncertainties in bonding coefficients are propagated for the calculation of $T_{g,dry}$, the uncertainties in the calculated value can be almost a 100% of the original value itself. Nonetheless, the significance of such parameterizations to predict, and therefore, understand the evolution of organic aerosol in the atmosphere is undeniable. This dissertation makes extensive use of all the aforementioned parameters to elaborate the molecular composition and physicochemical properties of organic aerosol.

2 Expanding the Molecular Characterization of Natural Organic Matter with Complementary Ionization Methods

Abstract

A major challenge in untargeted mass spectrometric analysis of complex mixtures is ionization selectivity. A single ionization method is generally insufficient to ionize all constituents of a complex mixture, so the ‘most suitable’ method is chosen that can access a representative fraction. For instance, while negative-mode electrospray ionization (ESI) is suitable for a sample rich in carboxylic acids, positive-mode atmospheric pressure photoionization (APPI) or chemical ionization (APCI) may be well suited for a mixture rich in non-polar aromatic moieties. Here, high-resolution mass spectrometric analysis of three natural organic matter mixtures was performed using six ionization methods: (+/-)ESI, APCI, and APPI in combination with the Fourier transform – Orbitrap Elite Mass Spectrometer. We detected thousands of chemical species of distinct chemical compositions in each of the IHSS mixtures. While common features were detected by different methods, each method selectively identified a fraction of complex mixtures as well whose composition was complementary to that accessible by other methods. Although the full breadth of molecular composition of complex mixtures eludes current analytical methods, our findings on ionization selectivity highlight the need to employ complementary ionization techniques in mass spectrometric analysis of such mixtures.

2.1 Introduction

Complex environmental mixtures, such as aerosol downwind of wildfires, decades old organic matter trapped within glaciers, or detritus flowing with rivers and streams, are exquisitely complex. Their exact molecular-level composition is impossible to elucidate, but high-resolution mass spectrometry allows us to see thousands of molecular compounds that constitute these mixtures. A major challenge in untargeted mass spectrometric analysis of complex mixtures is the inability of a single ionization method to ionize ‘everything’ in them. The extent and type of information obtained from different mass spectrometric analyses can be customized based on research needs. One popular way of doing so is by

exploiting the selectivity of ionization methods. Conventionally, electrospray ionization (ESI) has been used to study diverse types of natural organic matter (NOM), owing to the abundance of polar compounds in NOM^{57,Hawkes, 2020 #86, 101, 102}. However, the presence of non-polar or semi-polar fractions is now established with alternative ionization methods, such as atmospheric pressure photoionization (APPI)⁶¹, atmospheric pressure chemical ionization (APCI)^{60, 61}, and matrix-assisted/laser desorption ionization (MA/LDI)⁶⁰. A large fraction of molecular components ionized by different methods may overlap, but each ionization method ionizes certain molecules selectively as well^{60, 61}. It is this selectivity that has motivated the continuous development of novel ionization methods to explore previously unexplored fractions of NOM.

A pioneering work on high-resolution mass spectrometric analysis by Fievre et al., 1997 indicated that untargeted analysis of complex mixtures, such as humic or fulvic acids, assumes that all components have equal ionization efficiencies with a given ionization method¹⁰³. Furthermore, if a soft ionization technique is employed, fragmentation is considered to be negligible. A later study of humic substances in Suwannee River fulvic acids (SRFA) demonstrated 46.8% and 39.0% of the formulae assigned to SRFA by LDI and MALDI to be condensed aromatics, while ~99% of the formulae were non-aromatic with ESI⁶⁰. Given the complex nature of NOM and the chemical variety of its molecular components, no mass spectrum can be generated that is representative of the original, unionized mixture¹⁰³. In this context, it can be concluded that despite being a method of choice for ionization of NOM, ESI cannot provide a comprehensive picture of the molecular components.

Detailed mass spectrometric analysis of three commercial NOM mixtures from the International Humic Substances Society (IHSS) is reported here using high-resolution mass spectrometry by the Fourier-transform Orbitrap Elite mass spectrometer (FT-OTE MS). This instrument was used to generate six complementary datasets for each sample from three ionization methods: ESI, APPI, and APCI in either positive or negative polarity. Although this is not the first report of its kind, to the best of our knowledge, this is the first comprehensive demonstration of the selectivity of these popular ionization methods on NOM mixtures using the FT-OTE MS. Like many others, our findings highlight the importance of employing complementary ionization methods for analysis of NOM mixtures.

2.2 Methods

2.2.1 Sample acquisition and preparation

Samples were obtained from the IHSS and comprised of a fulvic acid mixture from Suwannee River (SRFA: 2S101F) and two NOM mixtures from Suwannee River (SRNOM: 2R101N) and Mississippi River (MSNOM: 1R110N). Information on their origin and elemental compositions is given in Table S2.1. Details on the collection and preparation of IHSS mixtures are available at <https://humic-substances.org/source-materials-for-ihss-samples/>. All mixtures were prepared identically by dissolving desalted, freeze-dried powder of the mixtures as obtained from IHSS in HPLC-grade 50:50 MeOH:H₂O solution. For positive mode analysis, a 100 ng/μL solution was prepared, and a 20 ng/μL solution was prepared for negative mode. A separate solution was prepared for (-)APPI with the same concentration with an additional 1% toluene.

2.2.2 Sample ionization and ultrahigh-resolution mass spectrometry

Spectra were acquired using ESI, APCI, and APPI in positive and negative modes on two separate days to subject all three samples to a total of six ionization methods. For APCI and APPI, scans were collected from m/z 250-800, as well as m/z 50-400. A single broadband measurement spanning m/z 50 to 800 caused the signal to diminish below m/z 250, hence we chose to collect two large segments that were merged into one during data processing. No ions were detected below m/z 150 with ESI, therefore, a scan range of m/z 150-800 was covered in both polarities. Thermo LTQ Tune Plus software (Thermo Fisher Scientific) was used to collect 200 scans for each acquisition. The XCalibur Qual Browser (Thermo Fisher Scientific) was used to export centroid data as mass lists with two columns: m/z and intensity, in comma-separated value (.csv) format. For each method and sample, the data was acquired in instrumental triplicates. Considering that APCI and APPI scans were segmented into two, a total of 90 mass lists ([high m/z segment] 3 samples $\times$ 6 methods $\times$ [low m/z segment] 3 replicates + 3 samples $\times$ 4 methods $\times$ 3 replicates) were generated, excluding the blank datasets.

2.2.3 Data processing and formula assignment

All mass lists were processed and formulae were assigned using the R package, MFAssignR⁹³ (version 0.0.4). The first step in the assignment of molecular formulae to the mass lists obtained from XCalibur Qual Browser (Thermo Fisher Scientific) was the estimation of baseline noise, which was performed using KMDNoise(). The average intensity of the noise peaks was multiplied by a factor of 6 and used as signal-to-noise (S/N) cut-off throughout data processing and analysis here. The second step was the application of IsoFiltR() to classify detected ions as monoisotopic or polyisotopic. Mass lists thus generated was subjected to internal recalibration to remove any systematic bias present in the measured masses. Compounds with C, H, and O constituents are the most abundant in NOM with relatively intense peaks as compared to the heteroatom-containing ones. They were thus preliminarily assigned molecular formulae using MFAssignCHO() without recalibration. At this stage, the assigned formula's theoretical mass was allowed to be within ≤ 3 ppm of the experimentally measured mass of the ion. Several CH₂ homologous series spanning the entire scan range were then selected as internal recalibrants from these preliminary assignments and were used to recalibrate mass measurements for all peaks in the raw dataset via Recal().

Following removal of noise and recalibration of measured masses, MFAssign() was used to assign molecular formulae to all mono and polyisotopic ions. Monoisotopic molecular formulae were allowed to have the following elements: ¹²C_c¹H_h¹⁶O_o¹⁴N₀₋₂³²S₀₋₂, while their polyisotopologues could have ¹³C₀₋₂³⁴S₀₋₁. However, all further discussion is based on monoisotopic ions only as isotope-containing formulae are only duplicates of a corresponding monoisotopic formula, and thus, do not contribute independently to the composition. Radical ions, i.e., odd-electron ions, are expected to form in APPI and APCI^{3, 67, 104}, therefore, we allowed their assignment for these methods in both polarities (POEx = 1 or NOEx = 1 for positive and negative assignments by MFAssign()). In most cases, radical ions had a more dominant (de)protonated ion, which rendered them duplicates. In all such cases, only a single ion type, i.e., distinct formulae only, were considered to reflect upon the composition of the sample.

Ion mode was chosen according to polarity, and up to one ²³Na (M_x = 1 in MFAssign()) was allowed in formulae assigned to mass lists obtained from (+)ESI analysis. In cases, where the same molecular formula was detected to exist with both H⁺ and Na⁺ adduct, the one with a taller peak was retained to avoid duplicates. Permissible error for final formula assignments was set at ≤ 1.5 ppm.

Values of double bond equivalences minus number of oxygen atoms (DBE-O) were allowed to be between -10 to +10. Elemental ratios were limited to $0.3 \leq \text{H/C} \leq 2.0$ and $0.0 \leq \text{O/C} \leq 1.0$. Detailed information on all the aforementioned functions (and others) and options for quality assurance (QA) available in MFAssignR can be found elsewhere⁹³.

Once formulae had been assigned, the segments for APCI and APPI in both polarities were merged. Also, triplicates were merged to obtain a single dataset for each sample and method, leaving us with 18 datasets (3 samples $\times$ 6 methods). For this purpose, species that were detected in at least two replicates were retained, while others were discarded. As intensities of ions vary across replicates of the same sample and acquisition, peak intensities were averaged to get the final intensity in merged datasets. Finally, peaks from contaminant ions were removed. All peaks with intensity $\geq 0.1\%$ of the tallest peak in the blank were considered to originate from contaminants and were removed from the sample datasets if the corresponding experimentally measured masses were within ± 1.5 ppm of each other and the assigned formulae were the same.

2.2.4 Computation of chemical characteristics and statistical analysis

For structural classification, modified aromaticity index (AI_{mod}) was calculated using: $[1 + \text{C} - (\text{O}/2) - \text{S} - \{(\text{N} + \text{H})/2\}] / [\text{C} - (\text{O}/2) - \text{N} - \text{S}]$ ^{99, 100}. Formulae were classified as aliphatic and olefinic if $\text{AI}_{\text{mod}} = 0$ and $0 < \text{AI}_{\text{mod}} < 0.5$, respectively¹⁰⁵. They were classified as aromatic and condensed aromatic if $\text{AI}_{\text{mod}} \geq 0.5$ and $\text{AI}_{\text{mod}} \geq 0.67$, respectively^{99, 100}. Olefinic species were further classified as high-O unsaturated and low-O unsaturated⁵⁵. Averages and standard deviations of such parameters related to chemical composition were calculated using the R Stats Package (version 3.5.1). Where required, all abundances were normalized to the sum of abundance of final formulae in the dataset.

Compositional differences between and across samples, and importantly, between the composition of the same samples as seen with the different ionization methods, were quantified as Bray-Curtis dissimilarity (BCD) with an R package, Vegan (version 2.5.5). All individual molecular formulae, such as $\text{C}_8\text{H}_6\text{O}_3\text{S}$, were treated as “species” that were quantified by their normalized abundances. The data was subjected to `vegdist()` for computation of BCD between all possible pairwise combinations of the ionization methods and samples included here. For ease of understanding, dissimilarity values were

subtracted from 1 and multiplied with a 100 to obtain BCD similarity percentages. For visualization purposes, variation in molecular composition elucidated by the different ionization methods was also measured by subjecting normalized intensities of identified species, i.e., distinct formulae, to non-metric multidimensional scaling (NMDS). Average values of compositional parameters, such as DBE_{avg} , H/C_{avg} , and O/C_{avg} , were fit to NMDS. This statistical analysis was performed using Vegan in R (version 2.5.5) with functions `metaMDS()` and `envfit()`. The aim of this analysis was not to isolate fractions that are selectively ionized by each ionization method, but only to identify the difference in characteristics observed in the fraction ionized and detected by each ionization method.

2.3 Results and Discussion

2.3.1 General characteristics of the mass spectra

Superficially, the 18 mass spectra seemed similar regardless of different methods, polarities, and samples. Mass spacing patterns of 14 and 2 Da that originate from a series of molecular species differing in CH_4 and H_2 groups were predominantly seen by all ionization methods. With respect to distribution of ion intensities across the scanned mass range, APCI and APPI were much similar to one another as compared to ESI in both positive and negative mode (Figures 2.1, S2.1-S2.2). The objective of this study was to demonstrate the selectivity of ESI, APCI, and APPI in positive and negative mode and the nature of molecules toward which they are selective. Although we used three IHSS mixtures for this purpose, considering that similar observations were recorded for all mixtures, most discussion in this section will be based on results for MSNOM. Table 2.1 presents a summary of molecular composition deduced from monoisotopic species detected in MSNOM. All corresponding observations for SRFA and SRNOM are presented in supplementary material (Tables S2.3 and S2.4). Quantitatively, a total of 9347 *distinct* formulae and 28,193 total formulae were assigned to MSNOM by the three ionization methods in the two polarities: 3929, 4067, and 4389 species were detected in MSNOM above a strict noise cut-off of $6\times$ the estimated noise from (-)ESI, (-)APCI, and (-)APPI and 2971, 6629, and 6208 species from (+)ESI, (+)APCI, and (+)APPI. The average formula error ranged from 0.15 ± 0.13 for (+)APPI to 0.38 ± 0.31 ppm for (-)APCI.

A notable difference between the ionization methods here was the detection of peaks with $m/z < 150$ by APCI and APPI, which remained undetected by ESI. This observation is in line with APCI and APPI being capable of ionizing less polar, small molecules, while ESI is well-established to readily access larger, more polar molecules³ (Figure 2.1). Early studies on humic substances considered them to be large ($m/z > 1000$) polymeric molecules¹⁰⁶, followed by later studies reporting the presence of smaller molecules ($< m/z 700$) as well¹⁰³.

Figure 2.1 Reconstructed mass spectra and elemental ratios from negative and positive mode analysis of Mississippi River natural organic matter using ESI, APPI, and APCI with FT-OTE MS. These plots were drawn with molecular formulae obtained after merging segments (where applicable), merging replicates, and subtracting the blank. Colours denote molecular groups: CHO, green; CHNO, blue; CHOS, red; CH, brown; CHN, cyan. The high normalized abundance of 6.0 in (+)ESI is due to one very tall analyte peak.

2.3.2 Molecular composition: Difference in trends of C, O, and DBE

Using KMD analysis and van Krevelen visualization, Hertkorn et al.⁶¹ exhibited oxygenation to be a major discriminatory factor between positive and negative ions generated by the same ionization method, where more oxygenated molecules were selectively ionized in negative polarity. In negative ionization, APCI exhibited preference for small, O-deficient molecules, while ESI and APPI tended to ionize O-rich compounds⁶¹. In a comparison of MALDI/LDI with ESI in negative mode, the former preferentially ionized species of low O/C and H/C, while ESI was selective for high O/C and H/C⁶⁰. In our study, all ionization methods similarly showed clear-cut differences in the formula distributions and ranges of DBE, C, and O covered (Figure 2.3). In negative polarity, although the ranges of O covered were similar, the distributions were rather different, where the number of formulae decreased from low O numbers to higher O numbers in (-)ESI as compared to (-)APCI and (-)APPI. The lowered number of formulae with more O atoms can be explained by the mass distribution and C numbers. While (-)ESI detected species with $C \leq 36$, (-)APCI and (-)APPI had $C \leq 43$ and 41. As larger molecules can hold more O, (-)APCI and (-)APPI exhibited much higher number of formulae with greater O numbers. Regardless of this, the preference of (-)ESI for more oxygenated species is undeniable as can be seen in the O/C_{avg} of 0.51 ± 0.18 , which was the highest amongst all six methods applied to MSNOM (Table 2.1; Figure 2.4).

Figure 2.3. Distribution of formulae assigned to Mississippi River natural organic matter over C, O, and DBE ranges. All formulae drawn in the positive y-axis represent (+)APCI, APPI, or ESI, while all formulae in the negative y-axis were obtained from (-)APCI, APPI, or ESI.

Table 2.1. A summary of molecular species detected in Mississippi River natural organic matter. Numbers in brackets represent uncertainties.

Ionization method	Formulae assigned	Negative mode				Formulae assigned	Positive mode				AI _{mod}
		O/C	H/C	DBE	AI _{mod}		O/C	H/C	DBE	AI _{mod}	
ESI	3929	0.51 (0.18)	1.03 (0.28)	10.58 (4.02)	0.37 (0.21)	2971	0.37 (0.18)	1.25 (0.27)	8.92 (3.17)	0.29 (0.19)	
APCI	4067	0.48 (0.18)	0.95 (0.28)	12.70 (5.26)	0.46 (0.26)	6629	0.42 (0.19)	0.98 (0.28)	12.79 (4.26)	0.44 (0.19)	
APPI	4389	0.49 (0.19)	0.98 (0.26)	11.77 (4.24)	0.42 (0.20)	6208	0.41 (0.19)	0.98 (0.27)	12.72 (4.18)	0.44 (0.19)	
ESI only	600	0.55 (0.24)	1.24 (0.34)	6.87 (3.52)	--	792	0.37 (0.21)	1.48 (0.24)	7.35 (3.21)	--	
APCI only	343	0.52 (0.26)	0.99 (0.43)	13.92 (8.80)	--	403	0.24 (0.31)	0.36 (1.06)	5.14 (13.10)	--	
APPI only	133	0.68 (0.31)	1.20 (0.36)	9.24 (4.88)	--	203	0.31 (0.26)	1.06 (0.37)	13.10 (5.66)	--	

2.3.3 Ionization selectivity: Exclusive and common detection of ions

The UpSet plots in Figure 2.2 show the overlap between the three different ionization methods. In all three samples, the numbers of molecular constituents commonly identified by ≥ 5 ionization methods were much larger than the number of constituents identified exclusively by each method or by combinations of < 5 methods. Notable exceptions to these were species exclusively identified by (+)ESI and commonly identified by two ionization pairs: (+)APCI/(+)APPI and (-)APCI/(-)APPI. While this overlap between the different ionization methods is promising as it indicates the reliability of information acquired from different mass spectrometric methods, unique formulae were assigned by every single ionization method as well (Table 2.1). The number of unique species detected in MSNOM was in the order (-)APPI > (+)APPI > (-)APCI > (+)APCI > (-)ESI > (+)ESI. The order was slightly altered in SRNOM and SRFA, with one certainty: more species were identified in negative polarity as compared to positive polarity for MSNOM and SRNOM, while the inverse was true for SRFA. Both these samples are more complex as compared to SRFA and have a combination of humic and fulvic acids. For SRNOM, we recorded exclusive species in the order of (-)APPI > (+)APPI > (+)APCI > (-)APCI > (-)ESI > (+)ESI. In SRFA, we recorded (+)APPI > (+)APCI > (-)APPI > (-)ESI > (-)APCI > (+)ESI. Detection of more species in negative mode in SRNOM and MSNOM may be indicative of the prevalence of more O-dominated functional groups in these samples as compared to SRFA.

In a similar analysis on SRFA, Hertkorn et al.⁶¹ observed count of ions in the order of ESI < APCI < APPI in positive mode and APPI < ESI < APCI in negative mode. Furthermore, despite the presence of common species across the two polarities with the same ionization method, the authors reported a ‘profound variance in selectivity’ of the ionization methods. A comparison of ESI, LDI, APCI, and APPI in negative polarity on scotch whisky samples showed the least unique features to be detected by APPI, while the highest overlap occurred between LDI and APCI, along with a substantial number of formulae commonly detected by all ionization methods¹⁰⁴. As shown in Figure 2.2, our observations were undoubtedly in line with this observation, where clear-cut difference could be outlined in the chemical nature of species that were exclusively ionized by APCI, APPI, or ESI in the same polarity but also across the two polarities for the same ionization method. As demonstrated previously by Blackburn et al.⁶⁰, a larger number of these formulae exclusively assigned to one dataset originate because its corresponding ion was exclusively ionized by one method. However, there are cases, where some formulae may fall outside of the error threshold in some

datasets, while within these thresholds in others as instrumental calibration may slightly vary across ionization methods.

Figure 2.2 Number of formulae common or exclusive to individual ionization methods or possible combinations of methods. The height of the bar along the y-axis and number labelled on top indicates the number of formulae identified by given method(s) on the x-axis

In the positive mode, not only the ranges of DBE, O, and C covered by each method varied but distribution of formulae differed greatly as well. Similar to the negative mode, (+)APCI and (+)APPI provided evidence of larger C skeletons. A notable difference was observed in the oxygenation levels of species selectively ionized by (+)ESI. Positive ESI did not appear to be selective for more oxygenated species as compared to (+)APCI and (+)APPI; not only was the O/C_{avg} of (+)ESI much lower than it was for (-)ESI, but it was much lower than for (+)APCI and (+)APPI as well (Figures 2.4 and S2.3). Both (+/-)APCI and (+/-)APPI ionized much higher content of aromatics and condensed aromatics as compared to (+/-)ESI. The preference of APPI and APCI for aromatic species, i.e., species of higher DBE, as compared to ESI was much distinct in positive mode than it was in negative mode (Figures 2.5 and S2.4). The AI_{mod} assumes that only half of the O atoms in a molecule to contribute to unsaturation⁶⁰. Previously, APCI and APPI have been reported to have selectivity of slightly higher DBE as compared to ESI¹⁰⁴.

Figure 2.4. Distribution of O/C and AI_{mod} of formulae assigned to Mississippi River NOM by the six ionization methods. The violin plots are classified further as all ions assigned by a method (gray) and the formulae exclusively assigned (purple).

Figure 2.5 The number of formulae detected by each of the six ionization methods in MSNOM, SRNOM, and SRFA and their distribution across modified aromaticity index classes of Koch and Dittmar^{99, 100}

A high degree of similarity is expected between these samples, especially SRFA and SRNOM, as that is essentially the same sample processed to a different extent⁶¹. Mississippi River NOM also has heavy terrestrial influence like Suwannee

River. Figure S2.5 shows the heteroatomic classes belonging to all molecular groups, CH, CHN, CHO, CHNO, and CHOS, found in the three samples. Interestingly, with very slight differences, these IHSS mixtures cover O-O₂₂, NO-NO₂₁, N₂O-N₂O₂₁, and O₂S-O₁₂S. For all three IHSS mixtures analyzed here, largest overlap of species identified by all six methods was in CHO species. Positive ESI showed prominent selectivity for NO_x species, while N₂O_x species were selectively ionized by (-)ESI in all samples. These patterns of ionization selectivity are exactly the same, which corroborates a possible compositional similarity between the samples. This is depicted in Figure 2.7A, where >40% similarity is exhibited between most pairwise combinations of ionization method/sample. Furthermore, Figure S2.6 shows the percent distribution of formulae assigned to each sample by all methods in different molecular groups. The same molecular groups are present in all samples. The most prominent difference is, however, the increment of heteroatomic content in the order of SRFA > SRNOM > MSNOM. It must be emphasized at this point that only slight differences were observed between the samples by a single ionization method. Figure 2.7B highlights the difference in composition of the same sample when analyzed with different ionization methods. Oxygenation parameters, OS_c, number of O atoms, and O/C, are determinative factors separating composition elucidated in negative polarity from that in positive polarity. Aromaticity (AI_{mod} and DBE) separate (+)APCI and (+)APPI from the rest of the ionization methods, while (+)ESI is distinguished from others largely because of H-saturation and number of H atoms. The clustering of samples by ionization method is because of their similar chemical compositions. The immense separation of sample composition as observed with the six methods provides very strong basis for ionization selectivity and highlights the necessity of employing ≥2 methods for comprehensive complex mixture analysis.

2.4 Supplementary material

2.4.1 Supplementary tables

Table S2.1 Origin, product code, and %(W/W) elemental compositions of dry ash-free dissolved organic matter mixtures reported by the International Humic Substances Society (IHSS; <http://humic-substances.org/elemental-compositions-and-stable-isotopic-ratios-of-ihss-samples/>). Elemental ratios were calculated based on bulk content.

	Origin	IHSS category number	C	H	O	N	S	H/C	O/C
SRFA	Terrestrially-influenced blackwater from Suwannee River that originates from the Okefenokee Swamp in South Georgia, USA	2S101F	52.34	4.36	42.98	0.67	0.46	0.99	0.61
SRNOM	Terrestrially-influenced blackwater from Suwannee River that originates from the Okefenokee Swamp in South Georgia, USA	2R101N	50.70	3.97	41.48	1.27	1.78	0.93	0.61
MSNOM	Mississippi River in Minneapolis, Minnesota, USA	1R110N	49.98	4.61	41.4	2.36	2.62	1.10	0.62

The second alphabet in the category number indicates whether the sample is an IHSS 'reference' or 'standard;' they are denoted by R and S, respectively. They differ in the extent of processing that they are subjected to. Further details are available at the IHSS website: <http://humic-substances.org/source-materials-for-ihss-samples/>

Table S2.3 A summary of molecular species detected in Suwannee River natural organic matter (SRNOM). There were 8356 distinct species identified for SRNOM in total. Numbers in brackets denote uncertainties.

Ionization method	Formulae assigned	O/C	H/C	DBE	AI _{mod}	Formulae assigned	O/C	H/C	DBE	AI _{mod}
	Negative mode					Positive mode				
ESI	3286	0.51 (0.18)	0.98 (0.26)	11.34 (4.14)	0.40 (0.18)	2856	0.39 (0.19)	1.24 (0.27)	9.49 (3.35)	0.29 (0.18)
APCI	3817	0.48 (0.17)	0.91 (0.26)	13.59 (5.56)	0.47 (0.19)	5749	0.41 (0.19)	0.96 (0.27)	13.31 (4.39)	0.45 (0.18)
APPI	3869	0.49 (0.18)	0.94 (0.25)	12.60 (4.54)	0.44 (0.18)	5591	0.41 (0.18)	0.95 (0.27)	13.28 (4.36)	0.46 (0.17)
ESI only	484	0.52 (0.24)	1.09 (0.35)	8.52 (4.33)	--	797	0.38 (0.23)	1.46 (0.26)	7.50 (3.46)	--
APCI only	360	0.49 (0.25)	0.92 (0.38)	14.36 (9.25)	--	302	0.40 (0.26)	1.10 (0.33)	13.15 (5.31)	--
APPI only	114	0.58 (0.29)	1.07 (0.34)	9.80 (5.00)	--	170	0.30 (0.19)	0.98 (0.38)	13.25 (5.49)	--

Table S2.4 A summary of molecular species detected in Suwannee River fulvic acids. There were a total of 7338 distinct species. Numbers in brackets denote uncertainties.

Ionization method	Formulae assigned	Negative mode				Formulae assigned	Positive mode				DBE	AI _{mod}
		O/C	H/C	DBE	AI _{mod}		O/C	H/C	DBE	AI _{mod}		
ESI	3000	0.48 (0.16)	0.99 (0.26)	11.73 (4.20)	0.41 (0.18)	3624	0.46 (0.17)	1.16 (0.25)	10.80 (3.67)	0.30 (0.18)		
APCI	3801	0.48 (0.16)	0.89 (0.28)	14.28 (5.98)	0.48 (0.19)	3768	0.40 (0.14)	1.03 (0.26)	12.60 (4.00)	0.41 (0.17)		
APPI	3986	0.49 (0.16)	0.93 (0.26)	13.39 (5.08)	0.44 (0.20)	3267	0.40 (0.14)	1.02 (0.25)	12.68 (3.91)	0.42 (0.16)		
ESI only	424	0.45 (0.20)	0.96 (0.29)	10.41 (4.24)	--	838	0.50 (0.22)	1.41 (0.19)	8.30 (3.16)	--		
APCI only	487	0.48 (0.22)	0.79 (0.38)	16.27 (9.07)	--	195	0.29 (0.17)	1.06 (0.31)	12.53 (4.31)	--		
APPI only	271	0.63 (0.23)	0.95 (0.26)	11.97 (4.80)	--	66	0.32 (0.18)	1.11 (0.28)	10.71 (5.31)	--		

2.4.2 Supplementary figures

Figure S2.1 Reconstructed mass spectra and elemental ratios from negative and positive mode analysis of Suwannee River natural organic matter using ESI, APPI, and APCI with FT-OTE MS. These plots were drawn with molecular formulae obtained after merging segments (where applicable), merging replicates, and subtracting the blank. Colours denote molecular groups: CHO, green; CHNO, blue; CHOS, red; CH, brown; CHN, cyan.

Figure S2.2 Reconstructed mass spectra and elemental ratios from negative and positive mode analysis of Suwannee River fulvic acids using ESI, APPI, and APCI with FT-OTE MS. These plots were drawn with molecular formulae obtained after merging segments (where applicable), merging replicates, and subtracting the blank. Colours denote molecular groups: CHO, green; CHNO, blue; CHOS, red; CH, brown; CHN, cyan.

Figure S2.3 Distribution of O/C and AI_{mod} of formulae assigned to SRFA and SRNOM by the six ionization methods. The violin plots are classified further as all ions assigned by a method (gray) and the formulae exclusively assigned (purple).

Figure S2.4 The number of formulae detected by each of the six ionization methods in MSNOM, SRNOM, and SRFA and their distribution across aromaticity index classes of Koch and Dittmar^{99, 100}

Figure S2.5 Heteroatom classes observed in IHSS dissolved organic matter mixtures using the six ionization methods.

Figure S2.6 Percent distribution of formulae assigned to IHSS dissolved organic matter mixtures by the six ionization methods across different heteroatomic molecular groups

3 Qualitative Molecular Characterization of Fulvic Acid mixtures from the International Humic Substances Society with High-Resolution Orbitrap Mass Spectrometry

Abstract

Detailed molecular analyses provide a chemical perspective of complex organic matter mixtures that improves our understanding of their origin, transformation, and fate. Ultrahigh-resolution mass spectrometry can unfold the composition of complex environmental mixtures down to their individual molecular entities. Reference and standard mixtures from the International Humic Substances Society (IHSS) have long been an integral component of analytical studies on complex organic matter to ensure objectivity as they allow researchers to distinguish perceived molecular characteristics that may be a figment of procedural or instrumental errors from the true composition of an unknown sample. While some IHSS DOM mixtures, such as fulvic acids from Suwannee River and Pony Lake are rather well-studied, others, such as natural organic matter from Mississippi River or fulvic acids from Lake Hellrudmyra, need more attention. Here, we present the detailed chemical characterization of six reference and standard mixtures from the IHSS, comprising fulvic acids from Suwannee River (Georgia, US), Pony Lake (Cape Royds, Antarctica), Elliot soil (Illinois, US), and Lake Hellrudmyra (Oslo, Norway), and two natural organic matter mixtures collected from Suwannee River and Mississippi River (Minnesota, US). This information is compiled in terms of distinct molecular formulae assigned to each m/z in mass spectra acquired from a Fourier Transform - Orbitrap Elite mass spectrometer and is useable for assessing the quality of ultrahigh-resolution mass spectrometric data by direct comparison. We use the information gained from molecular-level analyses, including heteroatom content, estimated aromaticity, and hydrogen saturation, to help guide the choice of electrospray ionization polarity and IHSS reference or standard mixture(s) for inclusion in analytical studies depending on their research objectives and the nature of target complex mixture(s) of interest.

3.1 Introduction

Environmental organic matter is a complex mixture with thousands of molecules that are formed from diverse sources and processes. In aquatic and soil systems, organic matter comprises a range of humic and non-humic substances from autochthonous sources, such as macrophytic algae and bacteria, and allochthonous or pedogenic sources, such as leaf litter^{107, 108}. In these ecosystems, OM participates in crucial processes, including transport, degradation, photolysis, and volatilization of organic pollutants¹⁰⁹; chelation of metals to alter their availability and toxicity in the environment^{110, 111}; photochemical reactions¹¹²; and nutrient cycling¹¹³, etc. Despite its ubiquity, OM is far from simple; it exists as complex mixtures of discrete compounds, whose concentration, composition, and chemistry are highly variable within and across ecosystems. A multitude of analytical techniques, including nuclear magnetic resonance, Fourier-transform infrared, and fluorescence spectroscopy¹¹⁴⁻¹¹⁷, are used to probe the coarse chemical nature of OM, whereas other tools, such as high-resolution mass spectrometers, have the capability to extract detailed molecular-level information^{102, 105, 118}. Constituents of OM range from small (<300 Da) to large (>600 Da) polyfunctional compounds with significant contribution of heteroatoms, especially O, N, S, and P. In high-resolution mass spectrometry, these mixtures present as multiple isobaric molecular species at single nominal masses, each of which further translates to numerous structural isomers. Studies utilizing Fourier transform - ion cyclotron resonance and Fourier transform - Orbitrap mass spectrometers (FT-ICR MS and FT-OTC MS, respectively) have provided invaluable insight into molecular compositions of OM from many different sources, ranging from the deep sea¹¹⁸, agricultural soils¹¹⁹, peatlands¹²⁰, the atmosphere^{73, 105, 121}, glacial catchments¹²², permafrost¹²³, and foodstuffs¹⁰⁴.

In the field of mass spectrometry, commercially-available references of complex mixtures from the International Humic Substances Society (IHSS) have been crucial in understanding instrumental optimization⁵⁴⁻⁵⁶, method development⁵⁷⁻⁵⁹, ionization selectivity, and considerations for best analytical approaches⁶⁰⁻⁶². Notably, these mixtures have satisfied their intended purpose, in general, for analytical purposes by functioning as a proxy for positive control in studies focused on untargeted characterization of complex mixtures of OM⁶³⁻⁶⁶. In such reports, findings about the unknown complex mixtures from a variety of analytical tools were validated by comparing compositions delineated for IHSS references and standard mixtures against the 'expected compositions.' Inclusion of an IHSS mixture allows 'apparent' molecular characteristics observed for complex environmental mixtures (due to discrepancies in methodology and

instrumental parameters) to be reliably differentiated from the ‘true’ characteristics, i.e., if observations for an IHSS reference or standard mixture sway radically from expected patterns, the conclusions that are drawn for an unknown sample of interest may also be rendered suspicious. Unless laboratories and research groups have well-established protocols, in which case they may forgo the inclusion of a reference mixture or not report it, in the absence of a standard or reference mixture, crucial errors may go unnoticed in the analysis of complex mixtures.

Currently (as of June 2020), the IHSS offers eight standard fulvic and humic acid mixtures and two NOM reference mixtures from soil, riverine, peat, and Leonhardtite sources⁵³. We present detailed characterization of six IHSS reference and standard mixtures enlisted in Table S1 in terms of molecular formula assignments to ions produced by either negative or positive electrospray ionization (ESI) and detected using an FT-OTE MS. We have used our findings on molecular composition to help guide the choice of the most chemically and procedurally relevant IHSS reference or standard mixture(s) for inclusion in a study, depending on the nature of target complex mixture of interest and objective of study. Furthermore, in light of previously known elemental content and hydrogen-saturation, we speculate that either negative or positive ESI may provide an adequate depiction of relative inter-sample differences between DOM mixtures, but a combination of two or more complementary ionization methods is preferable if the end goal is to delineate the maximal breadth of the composition of individual complex mixtures.

3.2 Materials and Methods

3.2.1 Sample acquisition and preparation

Samples were obtained from the IHSS and comprised DOM mixtures, including Suwannee River, Pony Lake, Nordic Lake, and Elliot soil fulvic acids (abbreviated SRFA, PLFA, NLFA, and ESFA, respectively), and two unfractionated NOM mixtures from Suwannee river and the upper Mississippi River (abbreviated SRNOM and MSNOM, respectively). Information on their origin and elemental compositions is given in Table S3.1. Details on the collection and preparation of IHSS mixtures are available at <https://humic-substances.org/source-materials-for-ihss-samples/>. Please note that SRFA, PLFA, ESFA, and SRNOM were received as part of the Inter-Laboratory Comparison Project reported elsewhere⁵⁵. All

data processing and findings presented here were performed independently for this report and are not a product of or related to the Inter-Laboratory Comparison Project. All DOM mixtures were prepared identically. Desalted, freeze-dried powder of the mixtures was dissolved in HPLC-grade 50:50 MeOH:H₂O solution and diluted to achieve a final concentration of ~20 mg/L of dissolved C. The mixtures were directly infused into the FT-OTE MS at a rate of 3-5 μ L/min using a 250-mL syringe for ESI.

3.2.2 Electrospray ionization and ultrahigh-resolution mass spectrometry

Before initiating our experiments, we baked the ESI source and sonicated the capillary tube in HPLC-grade MeOH and water for 30 minutes each to ensure that contaminants were removed. Charged analyte ions were produced by negative ([M-H]⁻) and positive ([M+H]⁺ and/or [M+Na]⁺) ESI and were subjected to mass analysis in two separate experiments. In both cases, the OTE was manually calibrated for mass accuracy before data acquisition with a Na-formate solution containing 13 known compounds (m/z 180-996) on the day that the analysis was performed on SRFA, PLFA, ESFA, and SRNOM, while NLFA and MSNOM were analyzed on a different day with the instrument being calibrated semi-automatically using manufacturer's instructions (Thermo Fisher Scientific). The reason being that the initial research design included only the first four DOM mixtures, while MSNOM and NLFA were added to the set later. As the aim of this study was to independently characterise all six DOM mixtures, and data of adequate quality were obtained from either instrumental calibration method, we decided to combine the two experiments in one report. Instrumental tune settings for both negative and positive ESI are presented in Table S3.2. For every mixture, once 1×10^6 ions were accumulated in the OTE, broadband spectra were recorded from m/z 150 to 1000 at a resolving power of 240,000 (at m/z 400) with the Thermo LTQ Tune Plus software (Thermo Fisher Scientific), where 200 scans were co-added by running average of three scans for each sample. Based on our observations, we speculate that averaging of mass spectra obtained from the FT-OTE MS reduces noise, and thus, enhances the signal-to-noise ratio considerably. The XCalibur Qual Browser (Thermo Fisher Scientific) was used to export centroid data as mass lists with two columns: m/z and intensity, in comma-separated value (.csv) format.

3.2.3 Assignment of molecular formulae

We processed the mass lists using R package, MFAssignR⁵⁹ (version 0.0.4). The first step in the assignment of molecular formulae to the mass lists obtained from XCalibur Qual Browser (Thermo Fisher Scientific) was the estimation of baseline noise, which was performed using KMDNoise() in MFAssignR. This function computes Kendrick mass defects to designate peaks with random mass defects in the entire dataset as noise and those with consistent mass defects as analyte peaks. The average intensity of the noise peaks in a user-selected portion of data is returned. KMDNoise() was applied here with default settings: KMDlower intercept= $0.1132x + 0.05$ and KMDupper intercept= $0.1132x + 0.2$, where x is the mass of the ion (or m/z). The second step was the application of IsoFiltR() on peaks with intensity six times higher than the estimated noise to decide if m/z values originate from monoisotopic or polyisotopic ions. In both ESI polarities, approximately 10,000 peaks were successfully identified to belong to either one of the ion types (Tables 3.1 and 3.2). Mass list thus generated was subjected to internal recalibration to remove any systematic bias present in the measured masses. Compounds with C, H, and O constituents are the most abundant in complex DOM mixtures with relatively intense peaks as compared to the heteroatom-containing ones. They were thus preliminarily assigned molecular formulae using MFAssignCHO() without recalibration. At this stage, the assigned formula's theoretical mass was allowed to be within 5 ppm of the experimentally measured mass of the ion (after addition of the mass of an H). Several CH₂ homologous series (enlisted in Table S3.7) spanning the entire scan range were then selected as internal recalibrants from these preliminary assignments and were used to recalibrate mass measurements for all peaks in the raw dataset via RecalX().

Following removal of noise and recalibration of measured masses, MFAssign() was used to assign molecular formulae to all mono and polyisotopic ions. Monoisotopic molecular formulae were allowed to have the following elements: ¹²C_{1-c}¹H_{0-h}¹⁶O_{1-o}¹⁴N₀₋₃³²S₀₋₁, while their polyisotopologues could have ¹³C₀₋₂³⁴S₀₋₁. However, all further discussion is based on monoisotopic ions only as isotope-containing formulae are only duplicates of a corresponding monoisotopic formula, and thus, do not contribute independently to the composition. Permissible error for final formula assignments was set at ≤ 3 ppm. Values of double bond equivalence minus number of oxygen atoms (DBE-O) were allowed to be between -13 to +13. Elemental ratios were limited to $0.3 \leq H/C \leq 2.0$, $0.0 \leq O/C \leq 1.0$. Ion mode was chosen according to polarity, and up to one ²³Na was allowed in formulae assigned to mass lists obtained from (+)ESI analysis. In cases where the same molecular formula was detected to exist with both H⁺ and

Na⁺ adduct, the one with a taller corresponding peak was retained to avoid duplicates. It should be noted that in some cases, where heteroatoms and high m/z ranges are considered in formula assignment, multiple formulae may pass formula extension parameters for a single m/z value and are returned by MFAssign() as ambiguous formulae. It is up to the user's discretion to choose the correct formula assigned to an m/z value. Such ambiguous formulae were not observed in mass lists from (-)ESI analysis. In (+)ESI, however, the total intensity of ions that were assigned an ambiguous formula was ~11 and 7% for NLFA and MSNOM; it was <3% for the rest of the mixtures. These formulae were excluded from further analysis.

Finally, peaks from contaminant ions were removed. All peaks with intensity $\geq 0.1\%$ of the tallest peak in the blank were considered to originate from contaminants and were removed from the datasets if the corresponding experimentally measured masses were within 3 ppm of each other and the assigned formulae were the same. Manual data validation, if required, was performed as discussed in supplementary material, Section 3.6.1.

3.2.4 Computation of chemical characteristics and inter-sample differences

Many parameters useful for studies on molecular composition and characterization, including O/C and H/C ratios, neutral mass, Kendrick mass defect, and double bond equivalence (DBE), etc. are returned in the MFAssignR output. To further molecular categorization, the modified aromaticity index (AI_{mod}) was calculated for each molecule using: $[1 + C - (O/2) - S - \{(N+H)/2\}] / [C - (O/2) - N - S]$ ^{99, 100}. Formulae were classified as aliphatic and olefinic if $AI_{mod} = 0$ and $0 < AI_{mod} < 0.5$, respectively¹⁰⁵; they were classified as aromatic and condensed aromatic if $AI_{mod} \geq 0.5$ and $AI_{mod} \geq 0.67$, respectively^{99, 100}. Olefinic species were further classified as high-O unsaturated and low-O unsaturated⁵⁵. Averages and standard deviations of such parameters related to chemical composition were calculated using the R Stats Package (version 3.5.1).

Generally, studies on complex DOM mixtures demarcate compound classes, such as lignin-, carbohydrate-, protein-, or lipid-like molecules, etc. on van Krevelen diagrams that are a plot of H/C against O/C ratios of individual molecular formulae¹²⁴. However, we remained wary of relying extensively on this method of classification owing to the ambiguity in the boundaries that define these classes¹²⁵. Compositional differences between our samples were quantified as Jaccard dissimilarity with an R package, Vegan (version 2.5.5) using vegdist().

All individual molecular formulae, such as $C_8H_6O_3S$, were treated as “species.” Their presence was indicated with a “1” and absence was indicated with a “0” to obtain a binomial data frame, which was subjected to `vegdist()` for computation of Jaccard dissimilarity values between all possible pairwise combinations of the DOM mixtures included here. For ease of understanding, dissimilarity values were subtracted from 1 and multiplied with a 100 to obtain Jaccard similarity percentages.

3.3 Results and Discussion

3.3.1 General characteristics of the mass spectra

Approximately 10,000 assignable peaks, originating from either mono or polyisotopic ions, were recorded above the S/N threshold for each IHSS mixture. In both negative and positive ESI, >80% of the total ion current was attributable to ions whose respective m/z values could be matched with an unambiguous molecular formula. The only exceptions to this were NLFA and MSNOM in (+)ESI, where only ~64 and 73% of the ion current was accounted for with a formula. As drawn in Figures S3.1-3.2, a small proportion of ions in all datasets in both polarities could not be assigned a formula. Corresponding peaks accounted for 3-13% of the total ion current in (-)ESI and 4-30% in (+)ESI. It is possible that such ions contained elements, isotopic signatures, or adducts lying outside of the formula assignment constraints applied in this study. The exact number of peaks that were (or were not) assigned an ambiguous or unambiguous formula are presented in Tables 3.1 and 3.2 for negative and positive ESI analysis, respectively.

Complete lists of ions that were assigned a molecular formula in each of the six DOM mixtures were aligned, where m/z values common in one or more samples were considered once, to obtain a total of 11,865 from (-)ESI and 11,559 formulae from (+)ESI analyses (Datasets S3.1 and S3.2). To gauge the reliability of molecular information thus acquired, we aligned our molecular formulae with the list of ions made available by Hawkes et al.⁵⁵ for SRFA, PLFA, ESFA, and SRNOM from several mass spectrometric instruments. A presence of >95% of these ions in the respective DOM mixture was recommended by the authors as an indication of reasonable instrumental performance. In (-)ESI analysis, 100, 98.5, 97.0, and 99.5% of these ions by Hawkes et al.⁵⁵ were detected in our mass lists with the same formula assigned by MFAssignR in SRFA, PLFA, ESFA,

and SRNOM, respectively. In (+)ESI, 99.7, 99.2, 97.6, and 99.3% of the ions recommended by Hawkes et al., 2020⁵⁵ were observed. Please note that these ‘reference’ ions for each of the four IHSS DOM mixtures were selected through very stringent criteria, i.e., they were detected and identified by 16 out of the 17 versatile MS instruments included in their study. Therefore, they represent a very small portion of the DOM mixtures that was commonly identified by all instruments, regardless of the tuning and optimization settings used for each. In our datasets, they formed only 19-23% of the full list of formulae assigned to SRFA, PLFA, ESFA, and SRNOM in (-)ESI and 12-18% in (+)ESI. Nonetheless, the presence of nearly all ions recommended by Hawkes et al.⁵⁵ and assignment of the correct formulae to them provides a reasonable estimate of the reliability of our findings from ESI FT-OTE MS analyses.

Figure 3.1 Mass spacing patterns in the spectrum of Pony Lake fulvic acids obtained from (-)ESI FT-OTE MS, (A) full mass spectrum of PLFA coloured by molecular groups: CHO, green; CHNO, blue; and CHOS, red, (B) an inset of the former from m/z 400 to 425 showing $\Delta m/z$ of 14.01565 and 1.0034 Da, (C) an inset of the former at nominal m/z 415 showing five molecular series with

peaks at $\Delta m/z$ of 0.0364 Da, and (D) the molecular formulae assigned to members of the five series in (C). Intensity of the single peak with an asterisk above it was quite low; it is possible that other members of its molecular series were below the signal-to-noise threshold used for formula assignment.

Peaks that were assigned an unambiguous molecular formula were present from m/z 150 to 800 in (-)ESI (Figure 3.2) and m/z 150 to 900 in (+)ESI (Figure S3.3). We found no evidence for the presence of multiply charged ions (i.e., no ^{13}C isotopic peaks were found at ~ 0.5 units after monoisotopic peaks), therefore, m/z values represent the actual molecular weight of OM components – considering a loss of proton in (-)ESI and addition of a proton or sodium ion in (+)ESI. Mass spacing patterns that are typical of mass spectra of fulvic acid mixtures^{102, 125} were present in all twelve of our mass spectra from both negative and positive ESI analyses, indicating that the resolving power and mass accuracy were sufficient for identification of N and S-containing components of these complex mixtures. Therefore, molecular formulae fell in three major molecular groups: CHO, CHNO, and CHOS, that were defined by their elemental constituents. Formulae in the CHO molecular group contained only C, H, and O, while CHNO and CHOS contained N and S along with C, H, and O, respectively. In (+)ESI, formulae containing only C and H (molecular group = CH) and C, H, and N (molecular group = CHN) were observed as well. However, these are excluded from further discussion due to their very small numbers. Using (-)ESI mass spectrum of PLFA – the most heterogeneous of all mixtures analyzed here – resolution of $(\text{CH}_2)_n$ and $(\text{CH}_2)_{n+1}$ homologous series ($\Delta m/z = 14.01565$), ^{12}C and ^{13}C ($\Delta m/z = 1.0034$), and CH_4 and ^{16}O ($\Delta m/z = 0.0364$) is shown in Figure 3.1. There was a total of 22 isobaric ions that were assigned formulae at m/z 451 above the S/N threshold, out of which 21 formed part of five distinct molecular series, including two CHNO and CHO series each and one CHOS series (Figure 3.1). Only 16 peaks were assigned formulae at the same nominal mass in SRFA, including two CHO and one CHNO series (data not shown).

Figure 3.2 Reconstructed mass spectra and elemental ratios from (-)ESI FT-OTE MS analysis of dissolved organic matter mixtures from the International Humic Substances Society. These plots were drawn with molecular formulae obtained after blank subtraction and manual data validation. Colours denote molecular groups: CHO, green; CHNO, blue; CHOS, red. Pie charts represent the proportion of each molecular group in a given mixture as summarized in Table S3.3.

3.3.2 Molecular composition of the IHSS DOM mixtures

3.3.2.1 *Pony Lake fulvic acids*

The DOM collected from Pony Lake is largely autochthonous, i.e., derived from local sources in the lake, such as extracellular release and leachate from algal cells or bacteria as terrestrial vegetation is sparse and only lichens, mosses, and algae thrive around it^{114, 126}. In our analyses, the largest number of distinct molecular formulae were assigned to PLFA (n = 6594 by (-)ESI) – more than half of which belonged to CHNO and CHOS groups (Table 3.2). Pony Lake FA was the only mixture, where the CHO molecular group did not make the largest fraction of all formulae assigned. Instead, >50% of the formulae belonged to CHNO. The ionic species with the greatest intensity in the mass spectrum of PLFA obtained from (-)ESI was assigned $C_{19}H_{26}O_6$ while that best ionized by (+)ESI differed by only two H atoms, i.e., $C_{19}H_{28}O_6$. Both molecular species are close to the previously reported compound of highest ionizability, $C_{19}H_{26}O_8$, in a (-)ESI FT-ICR MS analysis of PLFA by D’Andrilli et al.¹⁰².

Figure 3.3 Distribution of the complete list of molecular formulae assigned to dissolved organic matter mixtures from the International Humic Substances Society with respect to modified aromaticity index and other broad chemical classes described by Koch and Dittmar (2006; 2016) and Hawkes et al. (2020).

A DBE_{avg} of 9.60 was observed using both negative and positive ESI with an H/C_{avg} of 1.25 and 1.33, respectively. Pony Lake FA mixture was comparatively more H-saturated than other DOM mixtures studied here. The fraction of aromatic and condensed aromatic content in PLFA was starkly low as well (Figures 3.3 and S3.6), along with the lowest $\text{AI}_{\text{mod,avg}}$ in (-)ESI analysis (Figure 4). Experimental research with other analytical techniques has informed us of PLFA's low aromaticity. For instance, ^{13}C NMR analysis of PLFA has shown it to have much lower aromaticity than DOM from several terrestrially-influenced rivers¹¹⁴. Previous reports on FT-ICR MS analysis of PLFA associated it with lower aromatic content than SRFA^{102, 127}. In terms of O/C_{avg} , PLFA exhibited the lowest oxygenation of 0.43 and 0.37 in negative and positive ESI analyses, respectively (Tables 3.1 and 3.2). These trends were strengthened when the formulae that were uniquely assigned to ions generated by PLFA were considered. Such molecular formulae that may be considered as the unique chemical fingerprint of PLFA were predominantly N-containing: Among the 3150 formulae observed solely in PLFA using (-)ESI, ~73% belonged to the CHNO group (Table S3.2), while ~82% of the 1618 formulae observed uniquely using (+)ESI contained N.

3.3.2.2 Suwannee River fulvic acids and natural organic matter (NOM) and Mississippi River NOM

The molecular species that were most ionizable by (-)ESI in the three riverine IHSS mixtures had similar oxygenation and hydrogenation levels: H/C of ~ 0.5 and O/C of ~ 1.2 . In both NOM samples, $C_{18}H_{22}O_9$ was the most intensely ionized, while another member of the same CH_2 -based series, $C_{19}H_{24}O_9$, was the most intense in SRFA. In (+)ESI, $C_{17}H_{22}O_7$, $C_{16}H_{20}O_8$, and $C_{18}H_{22}O_9$, respectively, were the most intensely ionized in SRFA, SRNOM, and MSNOM. Overall, the DBE_{avg} , H/C_{avg} , O/C_{avg} , and proportion of aromatic compounds in these three mixtures appeared to be very similar when their molecular compositions were studied using (-)ESI (Table 3.1 and Figure 3.3). Some compositional similarity is expected between SRNOM and SRFA as they are essentially the same DOM mixture processed to a different extent. Reference NOM samples from the IHSS comprise not only the organic acids but also other organic solutes commonly present in natural waters⁵⁰. Among the 6485 molecular formulae assigned to SRFA and SRNOM altogether (common formulae counted once) with (-)ESI, a large proportion of 64.75% formulae were present in both samples, while 12.88% and 22.37% formulae were unique to SRFA and SRNOM, respectively (Table 3.3). On looking at Figure 3.5 and the numerical data compiled for species unique to either mixture in Table 3.1, the chemical fingerprint of SRNOM, relative to SRFA, from (-)ESI analysis becomes obvious, where SRNOM is more oxygenated and H-saturated with O/C_{avg} of 0.63 as compared to 0.46 for SRFA and H/C_{avg} of 0.92 as compared to 0.89. Positive ESI analysis did not reveal any difference between the H/C_{avg} of the two mixtures but showed SRNOM to be relatively more oxygenated with O/C_{avg} of 0.60 as compared to 0.45 (Table 3.2). Suwannee River FA had higher $Al_{mod,avg}$ (Figure 3.4) and larger fractions of aromatic and condensed aromatic content than SRNOM (Figures 3.4 and S3.6), which translated to it having higher DBE_{avg} of 12.70 as compared to 12.50 in (-)ESI and a DBE_{avg} of 12.00 as compared to 10.9 in (+)ESI (Tables 3.1 and 3.2).

Table 3.1 Summary of molecular formula assignments from (-)ESI FT-OTE MS analysis of dissolved organic matter mixtures from the International Humic Substances Society. Standard deviations are given in brackets. The same information is provided separately for each molecular group (CHO, CHNO, and CHOS) in Table S3.3.

	SRFA	PLFA	NLFA	ESFA	SRNOM	MSNOM
Monoisotopic peaks	8671	8415	8308	7947	8539	7787
Polyisotopic peaks	2807	2224	2882	2114	2943	2692
Number of unassigned peaks	3740	1679	3342	2980	3056	2189
Number of formulae^a	5034	6594	4942	5165	5650	6033
Averages (standard deviations)						
H/C_{avg}	0.98 (0.3)	1.25 (0.2)	0.94 (0.3)	0.96 (0.32)	0.97 (0.3)	1.02 (0.29)
O/C_{avg}	0.46 (0.2)	0.43 (0.6)	0.45 (0.2)	0.48 (0.15)	0.50 (0.2)	0.47 (0.18)
DBE_{avg}	12.70 (4.7)	9.60 (3.4)	13.20 (5.2)	12.2 (4.6)	12.5 (4.6)	11.4 (4.3)
C_{avg} atoms^b	22.60 (7.7)	22.10 (7.7)	22.7 (7.6)	21.4 (7.5)	21.9 (7.7)	20.6 (7.1)
O_{avg} atoms^b	10 (5)	9.3 (4.1)	10.10 (4.9)	10.2 (4.6)	11.0 (5.1)	9.6 (4.5)
N_{avg} atoms^b	0.37 (0.7)	1.0 (1.1)	0.34 (0.74)	0.45 (0.8)	0.45 (0.8)	0.55 (0.8)
Averages (standard deviations) for IONS EXCLUSIVELY PRESENT in each DOM mixture						
Number of formulae^a	187	3150	419	370	321	371
H/C_{avg}	0.89 (0.3)	1.33 (0.2)	0.75 (0.3)	0.80 (0.4)	0.92 (0.2)	1.04 (0.4)
O/C_{avg}	0.46 (0.2)	0.42 (0.1)	0.43 (0.2)	0.52 (0.2)	0.63 (0.2)	0.53 (0.3)
DBE_{avg}	16.58 (5.5)	10.20 (3.4)	18.29 (5.8)	12.81 (5.2)	15.76 (4.8)	11.03 (4.9)
C_{avg} atoms^b	27.84 (9.9)	25.26 (7.2)	27.03 (7.8)	19.30 (7.4)	26.72 (8.6)	19.53 (7.1)
O_{avg} atoms^b	12.93 (6.4)	10.48 (4.1)	11.54 (5.2)	9.70 (4.3)	16.45 (5.6)	9.60 (4.6)

N_{avg} atoms^b	0.91 (1.2)	1.38 (1.1)	0.61 (1.0)	1.07 (1.1)	0.71 (0.9)	1.07 (0.9)

^aThis is the number of formulae that were finalized after blank subtraction and manual data validation and include ¹²C, ¹H, ¹⁶O, ¹⁴N, and ³²S. Polyisotopologues are not included in this number; ^bNumber of atoms of the said elements per molecule

The Mississippi River meanders through 10 states and drains 41% of the continental United States, as well as two Canadian provinces. Through the course of history, it has been exposed to not only terrestrial influence but has also been subjected to industrial influx¹²⁸. A combination of these influences was reflected in slightly higher S (0.04%) and N (0.37%) content in the molecular components identified in it as compared to those in SRFA and SRNOM (data not shown), which is in line with the elemental compositions of all these samples (Table S3.1). Mississippi river NOM has been reported to have aromaticity lower than that of other IHSS DOM mixtures, except for PLFA⁵², which was corroborated by our (-)ESI analysis. Similarly, bulk elemental content reported by the IHSS indicate MSNOM to be more H-saturated, which could indicate lower aromaticity and DBE than both SRFA and SRNOM, with similar oxygenation (Figure 3.4). In negative and positive ESI analysis, based on the full list of molecular formulae assigned to MSNOM, its DBE_{avg} was computed to be 11.4 and 10.5, respectively, which was lower than the DBE_{avg} for both SRFA and SRNOM. As can be seen in Table 1, the difference in DBE_{avg} values for these mixtures was greater in (-)ESI. These trends of aromaticity, DBE_{avg}, and H/C_{avg} of MSNOM relative to SRFA and SRNOM became very distinct when we isolated the molecular formulae uniquely detected in MSNOM. In negative and positive ESI, formulae unique to MSNOM had DBE_{avg} of 11.03 (as compared to 16.58 for SRFA and 15.76 for SRNOM) and 10.41 (as compared to 15.90 for SRFA and 13.85 for SRNOM), respectively.

3.3.2.3 Nordic Lake fulvic acids

The IHSS obtains its NLFA mixture from a small glacial lake called Lake Hellrudmyra located near Oslo, Norway. The overall molecular composition of NLFA, when visualized as a van Krevelen diagram, appeared to be very similar to that of the terrestrially-influenced riverine samples (Figures 3.2 and S3.3). Here, the most ionizable molecular species by (-)ESI was assigned C₁₅H₂₀O₆, while species assigned C₂₀H₂₆O₈ was the most well-ionized in (+)ESI. The H/C_{avg} of

NLFA in both negative and positive ESI was lower than the rest of the samples (0.94 and 1.16, respectively). This is in concordance with the bulk elemental content reported by the IHSS (Table S3.1; Figure 3.4), which presents the possibility of this sample being relatively more aromatic with higher DBE than other IHSS DOM mixtures. Previous analytical studies have also indicated NLFA to be rich in aromatic content with little aliphatic matter⁵². Furthermore, elemental and NMR analysis has demonstrated NLFA to have lower H/C and higher O/C ratios as compared to PLFA, but similar to those for SRFA, with significantly higher aromaticity¹¹⁵. Indeed, in our study, NLFA had the highest proportion of aromatic and condensed aromatic formulae (43% and 25% in negative and positive ESI, respectively) in both polarities (Figures 3.3 and S3.6), as well as the highest $AI_{mod,avg}$ (Figure 3.4). Its DBE_{avg} was the highest in (-)ESI (13.20, increased to 18.29 for unique species), but was only 11 (11.76 for unique species) in (+)ESI analysis (Tables 3.1 and 3.2).

3.3.2.4 Elliot soil fulvic acids

Elliot soil FA was the only IHSS mixture, whose most ionizable species belonged to the CHOS molecular group. Here, the most intense peak in (-)ESI was assigned $C_7H_6O_5S$, while $C_{16}H_{28}O_6$ corresponded to the highest-intensity peak in (+)ESI. High-intensity CHOS species with C and O numbers <10 in (-)ESI were the most prominent aspect of ESFA. It should be noted that the high abundance of S-containing compounds should not be construed as a high proportion of the element in the DOM mixture, because S is much more abundant in SRNOM and MSNOM (Table S3.1). This only indicates the presence of highly ionizable S-containing compounds exclusively in ESFA. Elliot soil FA showed a very distinct dominance of aromatic and condensed aromatic species in (-)ESI analysis that was reinforced in the formulae uniquely assigned to it (Figures 3.3 and S3.6). This observation is supported by the fact that the bulk elemental composition of ESFA reported by the IHSS indicates it to have some of the lowest H-saturation among the DOM mixtures studied here (Figure 3.4). According to previous literature, the DOM from Elliot soil has slightly higher aromaticity than SRFA and much higher than PLFA, while it is similar to NLFA⁵², which is in line with our FT-OTE MS analyses. This trend was, however, not observed in (+)ESI analysis. Although the DBE_{avg} of the overall ESFA dataset (12.2) in (-)ESI was comparable to that of the riverine samples, it was 8.3, which is even lower than that for the least aromatic sample, PLFA ($DBE_{avg} = 9.6$), in (+)ESI analysis (Table 3.2).

Table 3.2 Summary of molecular formula assignments from (+)ESI FT-OTE MS analysis of dissolved organic matter mixtures from the International Humic Substances Society. Standard deviations are given in brackets. The same information is provided separately for each molecular group (CHO, CHNO, and CHOS) in Table S3.4.

	SRFA	PLFA	NLFA	ESFA	SRNOM	MSNOM
Monoisotopic peaks	8219	7533	7952	7342	7732	8170
Polyisotopic peaks	3398	2263	2391	2492	3058	2483
Number of unassigned peaks	1952	1748	2362	1414	1565	1592
Number of ambiguous assignments	520	7	661	244	31	416
Number of formulae^a	5628	5528	4591	5393	5358	5258
Averages (standard deviations)						
H/C_{avg}	1.2 (0.3)	1.33 (0.3)	1.16 (0.3)	1.37 (0.3)	1.2 (0.3)	1.2 (0.3)
O/C_{avg}	0.44 (0.2)	0.37 (0.1)	0.42 (0.2)	0.45 (0.2)	0.46 (0.2)	0.43 (0.2)
DBE_{avg}	12 (5)	9.6 (3.7)	11 (5)	8.3 (4.3)	10.9 (4.9)	10.5 (4.3)
C_{avg} atoms^b	26.1 (8.8)	24.6 (8.3)	23.9 (7.3)	22.0 (7.8)	24.0 (8.7)	22.9 (7.9)
O_{avg} atoms^b	11.2 (4.7)	8.9 (3.6)	9.5 (4.3)	9.7 (4.0)	10.9 (4.9)	9.5 (4.4)
N_{avg} atoms^b	0.35 (0.6)	0.78 (0.92)	0.52 (0.8)	0.40 (0.6)	0.31 (0.5)	0.52 (0.8)
Averages (standard deviations) for IONS EXCLUSIVELY PRESENT in each DOM mixture						
Number of formulae^a	621	1688	679	549	392	515
H/C_{avg}	1.06 (0.3)	1.23 (0.2)	1.21 (0.4)	1.62 (0.3)	1.06 (0.3)	1.19 (0.4)
O/C_{avg}	0.45 (0.2)	0.34 (0.1)	0.38 (0.3)	0.46 (0.2)	0.60 (0.2)	0.38 (0.2)
DBE_{avg}	15.90 (5.4)	11.45 (3.1)	11.76 (6.1)	5.40 (3.1)	13.85 (5.3)	10.41 (5.1)
C_{avg} atoms^b	32.22 (10.6)	28.50 (7.6)	24.80 (8.0)	21.76 (7.7)	26.54 (9.3)	21.82 (8.6)

O_{avg} atoms^b	13.99 (5.8)	9.57 (3.4)	8.37 (5.1)	9.92 (4.3)	15.21 (5.8)	8.12 (5.5)
N_{avg} atoms^b	0.70 (0.8)	1.49 (1.0)	1.40 (1.0)	0.94 (1.0)	0.56 (0.8)	1.23 (1.0)

^aThis is the number of formulae that were finalized after blank subtraction and manual data validation and include ¹²C, ¹H, ¹⁶O, ¹⁴N, and ³²S. Polyisotopologues are not included in this number; ^bNumber of atoms of the said elements per molecule

Figure 3.4 Distribution of elemental ratios (O/C and H/C) and modified aromaticity index of the complete list of molecular formulae assigned to each of the six dissolved organic matter mixtures analyzed by negative and positive ESI FT-OTE MS, as well as the ions that were found to be exclusively present in each of them. The former (all formulae) are shown in grey, and the exclusively present formulae are shown in purple. Corresponding averages of H/C, O/C, and DBE are given in Tables 3.1 and 3.2. Green diamonds denote the ratios calculated from experimentally determined bulk elemental content reported by IHSS on their website. Orange diamonds represent the values reported by Hawkes et al. 2020 for formulae that were commonly assigned to these mixtures by different MS instruments (16 out of the 17 instruments studied) in

an instrumental comparison study; NLFA and MSNOM were not included in this comparison, and thus, do not have an orange diamond drawn for them.

3.3.3 A common DOM fraction amidst inter-sample differences

The main motivation behind the selection of samples for this study lay in their differences, but our analyses revealed some very clear similarities – the most basic of these was the presence of the same three heteroatomic groups in all six DOM mixtures: CHNO, CHO, and CHOS, as well as CH and CHN in (+)ESI analysis. Despite the geographical separation and exclusivity of the ecosystems from which these DOM mixtures were derived, such as soil, fluvial, or lentic sources, distinct similarities in compositional parameters were recorded (Table S3.5 and Figure S3.4). Van Krevelen diagrams drawn in Figures 3.2 and S3.3 show the distribution of elemental ratios, O/C and H/C, to be rather similar across all samples, except PLFA, whose elemental ratios were concentrated toward relatively higher H/C and lower O/C. To confirm that this similarity between samples was not superficial (note that different molecular formulae with same O/C and H/C ratios will occur at the same spot on the van Krevelen diagram), we isolated formulae that were common to all six samples. In (-)ESI, among the total 11,865 molecular formulae, 1385 were common and were distributed across m/z 151 to 749 for CHO and 160 to 520 for CHNO, and lay in the confines of $0.5 \leq \text{H/C} \leq 1.7$ and $0.1 \leq \text{O/C} \leq 0.8$ that is characteristic of lignin-like and/or carboxylic-rich alicyclic (CRAM) molecules on the van Krevelen diagram¹²⁴ (Dataset S3.3). In (+)ESI, 1338 molecular formulae were common in a total of 11,559 formulae. These were spread across m/z 165 to 799 for CHO and 224 to 602 for CHNO within a van Krevelen space of $0.14 \leq \text{O/C} \leq 1.0$ and $0.6 \leq \text{H/C} \leq 2.0$ that also corresponds to lignin-like and/or CRAM, as well as tannin-like molecules¹²⁴ (Figure S3.4; Dataset S3.4). In both polarities, the largest overlap occurred in CHO components, which is corroborated by previous research. For instance, in a comparison between two IHSS DOM mixtures, PLFA and SRFA, D’Andrilli and colleagues¹⁰² reported heteroatoms to contribute only 16.3% of the overlap between the two with the rest being CHO formulae. It must be noted that a molecular formula being common in two or more complex mixtures does not imply that the actual compound is the same as well. Not only is a single measured mass in a mass spectrum a product of several isomeric ions within a DOM mixture, but the composition of these isomeric mixtures varies across samples as well. This can be somewhat visualized in Figure S3.4, where although the common ions certainly form the most intense peaks for all DOM

mixtures, their intensities vary across mixtures, which is likely due to the different ionization potentials of these compounds, as well as the different organic matter matrices, i.e., the rest of the compounds (not common in all samples) present in these complex mixtures.

An important aspect pertaining to this discussion on similarities between diverse DOM mixtures is the ambiguous nature of molecular categorizations from van Krevelen diagrams. A majority of compounds that lie well below the molecular lability boundary ($H/C < 1.5$) in a so-called “lignin-like” region may have originated from either autochthonous or allochthonous sources¹⁰¹, and thus, should be considered more in terms of their recalcitrance than their plant-based origin¹⁰². Zark and Dittmar¹¹⁸ used FT-ICR MS and NMR analyses to report the presence of a “universal background” in DOM from five spatially and environmentally diverse aquatic sources, including the deep sea, coastal ocean, and a eutrophic peat lake. Their shared molecular features comprised mainly of CRAM and were attributed to the source-independent microbial and photo-degradative transformation of DOM, which produces refractory DOM fractions that persist over millennia and travel great distances across land and sea. Therefore, in PLFA, which is devoid of terrestrial influence, the presence of this chemical signature shared with other DOM mixtures may be an illustration of the presence of end-products of microbial degradation.

Suwannee River FA and NOM mixtures exhibited the highest similarity to each other in both negative and positive ESI analysis with Jaccard similarity percentages of 64.75 and 59.31%. These percentages computed by the Jaccard method are the fraction of formulae that are common to two mixtures when their mass lists are aligned together to get a master list of both mixtures; the formulae that are not common are evidently unique to one or the other of the two mixtures as detailed in Tables 3.3 and S3.6. In the same manner, 64.53% of the total formulae determined using (-)ESI were common between SRNOM and MSNOM. Pony Lake FA mixture was consistently the most disparate in both negative and positive ESI analysis with similarity percentages <30% in (-)ESI (Table 3) and <40% in (+)ESI with all samples (Table S3.6). Pony Lake is rendered different from all other freshwater sources, even NLFA, which is also obtained from a remote cold-temperature lake, simply because of exposure to different factors that drive the influx and transformation of organic matter. For instance, the two NOM mixtures were the most heterogeneous with their N and S content originating from terrestrial and anthropogenic sources, while PLFA exhibited similar chemical heterogeneity but with only microbial factors driving its composition whatsoever.

The specific molecular composition and characteristics of IHSS mixtures should be considered in choosing the best reference mixture that can help evaluate the

reliability of analytical procedure applied on an unknown complex sample. In the simplest of cases, where samples do not belong to aquatic systems, such as organic matter mixtures from aerosol or soil, SRFA may be the most suitable choice owing to its well-studied nature and simplicity as it will be meant to perform no other function but to ensure instrumental performance and optimize data processing. In 2010, 75% of the overall sales of IHSS mixtures constituted Suwannee River humic and fulvic acid mixtures, as well as NOM, while other aquatic and terrestrial samples constituted only a smaller portion⁵⁰. The reason for the widespread use of SRFA could be its relative simplicity as compared to other IHSS mixtures while having sufficient complexity for being of interest in analytical studies¹²⁹. However, in cases, where aquatic samples are under study, a more chemically relevant choice can be made from among the many available IHSS mixtures for the sake of analytical validation, as well as for the inclusion of organic matter from a similar ecosystem. In studies, where heteroatom-rich mixtures are of concern, either MSNOM or SRNOM would be the most suitable proxies for a positive control. Nordic Lake FA will be a valuable and interesting reference material in studies involving organic matter samples from remote regions, such as glacial catchments, that are expected to have little or moderate terrestrial influence due to the abundance of aromatic species of terrestrial origin in it, which is rather mismatched with its remote cold-temperature lentic origin. It can be used in combination with PLFA that comes from a similar climate and environment but is high in aliphatic content, instead of aromatic species. Both NLFA and ESFA are not as well-studied as SRFA or SRNOM, and thus, it would be advisable to use them in combination with another, preferably well-studied and/or simpler, reference or standard mixture from the IHSS.

Figure 3.5 All possible pairwise combinations of the six dissolved organic matter mixtures from the International Humic Substances Society included in our study are visualized as a matrix of van Krevelen diagrams after being analyzed by (-)ESI FT-OTE MS. Molecular formulae that were present in both members of a pair are represented in gray in the background, and those unique to each mixture are shown by specific colours used in the column/row headers. This information is presented numerically in terms of Jaccard similarity percent in Table 3.3.

Table 3.3 Jaccard similarity percent between all possible pairs of dissolved organic matter mixtures included in this study. Similarity was calculated using presence and absence of molecular formulae identified by (-)ESI FT-OTE MS analysis. Percentage of molecular formulae unique to each mixture is also given. This information is presented visually as a matrix of van Krevelen diagrams in Figure 3.5.

SRFA	SRFA				MSNOM
PLFA	Common = 23.83% SRFA = 29.78% PLFA = 46.39%	PLFA			
NLFA	Common = 63.42% SRFA = 19.05% NLFA = 17.53%	Common = 19.22% PLFA = 48.90% NLFA = 31.81%	NLFA		
ESFA	Common = 58.33% SRFA = 19.82% ESFA = 21.86%	Common = 24.24% PLFA = 45.43% ESFA = 30.33%	Common = 53.28% NLFA = 21.66% ESFA = 25.06%	ESFA	
SRNOM	Common = 64.75% SRNOM = 22.37% SRFA = 12.88%	Common = 24.7% PLFA = 42.46% SRNOM = 32.84%	Common = 54.84% NLFA = 17.40% SRNOM = 27.76%	Common = 57.34% ESFA = 17.80% SRNOM = 24.86%	SRNOM
MSNOM	Common = 59.35% SRFA = 13.13% MSNOM = 27.52%	Common = 30.44% PLFA = 37.68% MSNOM = 31.89%	Common = 52.46% NLFA = 16.18% MSNOM = 31.35%	Common = 53.89% ESFA = 17.09% MSNOM = 29.02%	Common = 64.53% SRNOM = 20.43% MSNOM = 15.04%

3.3.4 Ionization complementarity and selectivity

Electrospray ionization is a soft-ionization technique that creates positively or negatively charged ions from polarizable species in a sample. Traditionally, (+)ESI analysis was performed on environmental organic matter mixtures, but the more recent application has shifted toward (-)ESI, owing to research that indicated DOM mixtures to be dominated by functional groups that are more prone to acquiring a negative charge, such as carboxylic, hydroxyl, or phenolic groups¹³⁰. If a study aims to maximally understand the intricate composition of a complex mixture using mass spectrometry, selection of the most suitable means of ionization is a key challenge because of the tendency of each ionization method to favour different types of molecules. On the one hand, it is impossible to access all molecular species present in a complex mixture by a single ionization method, and thus, the extent to which it is elucidated can be improved by using complementary ionization methods. For instance, in the case of ESFA, for whom very different molecular compositions were described by negative and positive ESI, it may be essential to use both polarities. Similarly, information on NLFA's aromatic nature can be better retrieved by combining ESI with another ionization method that can probe hydrogen-deficient and aromatic mixtures, such as atmospheric pressure photo or chemical ionization. On the other hand, in the simplest case scenarios, where the molecular characterization of multiple samples with a single ionization method is the goal, it is only feasible to rely on a single technique to measure as many target mixtures as possible, but at the expense of losing some information about them. In such situations, a method that determines composition-related parameters, such as H/C_{avg} , O/C_{avg} , aromaticity, etc. and inter-sample differences reliably may be the best choice.

Here, we used the simplest complementary pair of ionization methods: Negative and positive ESI, to demonstrate the extent of selectivity. Both ionization methods yielded some molecules that were exclusive to each with the rest being common to both. The highest overlap in molecular formulae obtained from the two polarities was recorded for PLFA, where ~43% of the formulae were common (Figure S3.5), while the least similarity of ~38% was observed for NLFA. Regardless of the nature of the DOM mixture, plotting molecular formulae on van Krevelen diagrams showed that molecules that were exclusively ionized by (+)ESI in any given DOM mixture were relatively more H-saturated and aliphatic, and those detected by (-)ESI were located in a more O-rich zone (Figure S3.5). Numerical values in Tables 3.1 and 3.2 provide the same information: All DOM mixtures had higher H/C_{avg} , lower DBE_{avg} , and lower O/C_{avg} when analyzed using (+)ESI. Regardless, the relative inter-sample differences delineated here were similar in either polarity (Table 3.2). For instance, NLFA

and SRFA were relatively less H-saturated and more aromatic as compared to other DOM mixtures in both positive and negative ESI. Similarly, MSNOM was less oxygenated with more N and S content than SRNOM in both polarities. In a study on mass spectrometric analysis of SRFA, PLFA, ESFA, and SRNOM⁵⁵, it was noted that saturation levels of DOM mixtures, in terms of AI_{mod} and H/C, were promising parameters for estimating the reliability of an instrument and tuning as they were rather consistently determined by mass spectrometric instruments of wide-ranging capabilities and resolving powers. Indeed, the H/C_{avg} recorded here for these DOM mixtures using (-)ESI were closer to the bulk elemental content reported by the IHSS. The $AI_{mod,avg}$ was also corroborated by that computed by Hawkes et al., 2020, especially for the relatively less complex FA mixtures as compared to the NOM in both negative and positive ESI (Figure 4). On the contrary, O/C_{avg} deviated from known elemental content reported by the IHSS in both polarities. Based on this information – together with the fact that mass lists acquired from (+)ESI analysis were more complicated to process and assign formulae, largely due to multiple adduct formation – (-)ESI is a more suitable ionization method for analytical studies aiming to qualitatively characterize multiple complex DOM mixtures with a single ionization method.

3.4 Acknowledgements

We would like to thank Dr Matthew Brege and Tyler Levertton for the many constructive discussions and advice, help with manual data validation, and guidance in the use of R and MFAssignR. Dr Jeffrey Hawkes, Dr Sarah Green, Dr Evan Kane, Dr Karl Meingaast, and Nastaran Khademimosheghani provided valuable feedback on the improvement of this manuscript. This work was supported in part through a Fulbright Foreign Student Award to AI.

3.5 Data availability

Datasets S3.1-S3.4 present formulae assigned to all six IHSS samples from the analysis presented in this chapter.

3.6 Supplementary material

3.6.1 Manual data validation

MFAssignR performs quality assurance on the assigned formulae by taking into consideration fundamental rules of chemical feasibility, such as N, Senior, and large atom rules etc., as well as user-defined parameters, including isotope matching for sulphur assignments and CH₄ assignment instead of O, where appropriate, with the SulfCheck and NMScut options, respectively. However, owing to the higher number of possible solutions (i.e., formulae) for every m/z value with increasing mass, especially when non-O heteroatom content is allowed, a key challenge with any automated system of formula assignments is to identify and remove unlikely formula assignments. Herzsprung et al.^{120, 131, 132} developed two assessment strategies to differentiate formulae obtained from high-resolution mass spectrometric data as ‘reliable’ and ‘less reliable’ based on (i) DBE versus number of oxygen atoms and (ii) DBE-O versus number of formulae. We used constraints outlined by these authors to manually sift through formulae left after blank subtraction and removed any outliers. Furthermore, C and DBE continuums were expected in all organic matter mixtures here, therefore, any isolated assignments at the lower or higher carbon number and DBE ends that were not continuous with the rest of the data were removed to avoid false positives. The number of such assignments was, however, very small. The only exceptions to this were NLFA and MSNOM, when analyzed using (+)ESI. As shown in Figure S3.2, both these mixtures had many tall peaks in (+)ESI that stood out from the rest of the spectrum. Many of them were assigned ambiguous formulae and were removed, while some peaks that were assigned unambiguous formulae exhibited unusual values of DBE-O, C, and O (especially in the CHNO molecular group) that were discontinuous from the rest of the formula assignments, and thus, were removed.

We observed multiple assignments containing C, H, N, O, and S together in single formulae, i.e., CHNOS molecular group, for all samples ($n = 3, 62, 157, 20, 53,$ and 261 in negative polarity and $n = 50, 0, 16, 0, 0, 0$ in positive polarity for SRFA, PLFA, NLFA, ESFA, SRNOM, and MSNOM, respectively), but they were removed from further analysis because such a combination of elements is difficult to resolve and assign confidently⁵⁵. Furthermore, CHOS formulae observed in (+)ESI ($n = 446, 1, 510, 212, 0, 243$ for SRFA, PLFA, NLFA, ESFA, SRNOM, and MSNOM) were removed from consideration as well but will be made available to the reader upon request. Assigned mass lists thus generated were

considered finalized and saved with raw intensities for further analyses (Datasets S3.1 and S3.2).

3.6.2 Supplementary tables

Table S3.1 Origin, product code, and %(W/W) elemental compositions of dry ash-free dissolved organic matter mixtures reported by the International Humic Substances Society (IHSS; <http://humic-substances.org/elemental-compositions-and-stable-isotopic-ratios-of-ihss-samples/>). Elemental ratios were calculated based on bulk content.

	IHSS category number	C	H	O	N	S	H/C	O/C
SRFA	2S101F	52.34	4.36	42.98	0.67	0.46	0.99	0.61
This fulvic acid mixture is collected from terrestrially-influenced blackwater of Suwannee River that originates from the Okefenokee Swamp in South Georgia, USA								
PLFA	1R109F	52.47	5.39	31.38	6.51	3.03	1.22	0.45
This microbially-influenced fulvic acid is collected from a hypereutrophic, saline coastal pond on Cape Royds, Antarctica.								
NLFA	1R105F	52.31	3.98	45.12	0.68	0.46	0.90	0.65
This fulvic acid mixture is collected from a small tarn with blackwater in a relatively isolated catchment near Oslo, Norway								
ESFA	5S102F	51.92	3.91	39.64	2.40	1.43	0.90	0.57
This fulvic acid mixture is collected from a fertile prairie soil from an undisturbed area on the Joliet Army Ammunition Plant near Joliet, Illinois.								
SRNOM	2R101N	50.70	3.97	41.48	1.27	1.78	0.93	0.61
This humic and fulvic acid mixture is collected from terrestrially-influenced blackwater of Suwannee River that originates from the Okefenokee Swamp in South Georgia, USA.								
MSNOM	1R110N	49.98	4.61	41.4	2.36	2.62	1.10	0.62
This humic and fulvic acid mixture is collected from Mississippi River in Minneapolis, Minnesota, USA.								

The second alphabet in the category number indicates whether the sample is an IHSS ‘reference’ or ‘standard;’ they are denoted by R and S, respectively. They differ in the extent of processing that they are subjected to. Further details are available at the IHSS website: <http://humic-substances.org/source-materials-for-ihss-samples/>

Table S3.2 Instrumental tune settings for data acquisition from the FT-OTE MS with positive and negative electrospray ionization.

	Negative polarity	Positive polarity
Heater temperature	100°C	120°C
Spray voltage	2800 V	4000 V
Capillary inlet temperature	265°C	285°C
S-lens RF level	69	69

Table S3.3 Summary of molecular formula assignments belonging to the three major molecular groups (CHO, CHNO, and CHOS) from (-)ESI FT-OTE MS analysis of dissolved organic matter mixtures from the International Humic Substances Society. Standard deviations are given in brackets. The same information is provided for all molecular formula assignments together in Table 3.1.

	SRFA			PLFA			NLFA			ESFA			SRNOM			MSNOM		
	CHO	CHN O	CHO S	CHO	CHN O	CHO S	CHO	CHN O	CHO S	CHO	CHN O	CHO S	CHO	CHN O	CHO S	CHO	CHNO	CHO S
Number of formulae ^a	359 4	118 4	256	232 3	371 0	561	371 2	968	261	335 2	149 2	321	358 3	172 4	343	340 9	2211	413
H/C _{avg}	1.01 (0.3)	0.88 (0.2)	1.03 (0.3)	1.30 (0.3)	1.20 (0.2)	1.37 (0.2)	0.97 (0.3)	0.81 (0.2)	1.00 (0.3)	1.02 (0.3)	0.83 (0.2)	0.98 (0.3)	0.99 (0.3)	0.90 (0.2)	1.09 (0.3)	1.04 (0.3)	0.96 (0.3)	1.14 (0.3)
O/C _{avg}	0.47 (0.2)	0.47 (0.1)	0.34 (0.1)	0.43 (0.2)	0.42 (0.1)	0.47 (0.1)	0.45 (0.2)	0.48 (0.2)	0.34 (0.1)	0.48 (0.2)	0.49 (0.1)	0.43 (0.2)	0.50 (0.2)	0.51 (0.2)	0.47 (0.2)	0.46 (0.2)	0.48 (0.2)	0.50 (0.2)
DBE _{avg}	13.1 (5.1)	12.3 (3.1)	8.0 (2.7)	8.9 (3.4)	10.5 (3.1)	6.7 (2.5)	13.7 (5.5)	12.9 (3.5)	8.2 (2.7)	12.6 (5.1)	12.2 (3.3)	7.8 (2.6)	13.2 (4.9)	12.1 (3.3)	7.1 (2.2)	12.0 (4.7)	11.3 (3.4)	6.8 (2.3)
C _{avg} atoms ^b	24.4 (7.9)	18.9 (4.7)	14.4 (3.3)	23.1 (8.2)	22.1 (7.4)	18.4 (6.6)	24.3 (7.7)	18.6 (4.6)	14.4 (3.2)	23.6 (7.6)	17.9 (4.9)	13.5 (3.8)	24.2 (7.9)	18.8 (5.3)	13.4 (3.1)	22.9 (7.4)	18.3 (5.3)	13.4 (3.4)
O _{avg} atoms ^b	11.4 (5.1)	9.0 (3.6)	5.0 (2.3)	9.8 (4.7)	9.1 (3.8)	8.4 (3.3)	10.8 (5.0)	9.0 (3.9)	4.9 (2.1)	11.3 (4.8)	8.8 (3.4)	5.7 (2.0)	12.0 (5.4)	9.7 (4.0)	6.1 (2.5)	10.5 (4.8)	8.7 (3.8)	6.5 (2.7)
N _{avg} atoms ^b	--	1.57 (0.6)	--	--	1.77 (0.8)	--	--	1.71 (0.7)	--	--	1.55 (0.6)	--	--	1.46 (0.6)	--	--	1.51 (0.7)	--
Averages (standard deviations) for IONS EXCLUSIVELY PRESENT in each DOM mixture																		
Number of formulae ^a	83	79	25	479	232 2	349	268	125	26	109	217	44	161	142	18	49	263	59
H/C _{wa}	0.98 (0.3)	0.76 (0.2)	1.03 (0.3)	1.47 (0.1)	1.28 (0.2)	1.43 (0.2)	0.73 (0.2)	0.74 (0.3)	0.98 (0.4)	1.01 (0.6)	0.70 (0.3)	0.82 (0.3)	0.96 (0.2)	0.85 (0.2)	1.01 (0.3)	1.34 (0.4)	0.96 (0.3)	1.18 (0.3)
O/C _{wa}	0.48 (0.1)	0.46 (0.2)	0.35 (0.1)	0.46 (0.2)	0.40 (0.1)	0.46 (0.1)	0.40 (0.1)	0.53 (0.2)	0.28 (0.1)	0.56 (0.2)	0.52 (0.2)	0.43 (0.2)	0.66 (0.1)	0.60 (0.2)	0.53 (0.3)	0.37 (0.3)	0.53 (0.2)	0.70 (0.3)
DBE _{wa}	19.6 6 (5.1)	15.4 5 (3.9)	9.88 (3.8)	8.92 (3.3)	10.8 9 (3.3)	7.32 (2.5)	20.6 4 (4.9)	15.0 8 (4.5)	9.50 (3.4)	12.4 4 (7.0)	13.7 6 (4.1)	9.04 (2.5)	17.3 9 (4.6)	14.8 9 (3.8)	8.06 (3.1)	9.59 (5.5)	12.27 (4.6)	6.74 (3.0)

C_{wa} atoms^b	36.7 3 (6.2)	21.6 4 (5.8)	17.8 8 (3.8)	29.3 7 (7.1)	24.9 5 (7.0)	21.5 9 (5.7)	30.9 8 (6.1)	20.6 2 (5.3)	17.2 0 (3.4)	23.1 7 (8.7)	18.3 7 (6.0)	14.3 2 (5.2)	31.6 3 (7.6)	22.7 7 (6.2)	14.0 5 (4.4)	24.5 9 (9.4)	19.90 (6.2)	13.6 8 (4.1)
O_{wa} atoms^b	17.8 2 (5.1)	9.78 (4.2)	6.64 (3.3)	12.9 4 (4.6)	10.0 6 (3.9)	9.84 (3.1)	12.6 8 (5.0)	10.4 7 (4.6)	4.92 (2.8)	12.4 0 (4.9)	9.10 (3.5)	5.95 (2.4)	20.1 9 (3.5)	13.3 5 (4.3)	7.33 (4.2)	8.04 (5.1)	10.07 6 (4.8)	8.80 (2.9)
N_{wa} atoms^b	--	2.16 (0.7)	--	--	1.87 (0.8)	--	--	2.04 (0.8)	--	--	1.82 (0.8)	--	--	1.61 (0.7)	--	--	1.51 (0.7)	--

^aThis is the number of formulae that were finalized after blank subtraction and manual data validation and include ¹²C, ¹H, ¹⁶O, ¹⁴N, and ³²S. Polysotopologues are not included in this number; ^bNumber of atoms of the said elements per molecule.

Table S3.4 Summary of molecular formula assignments belonging to the three major molecular groups (CHO, CHNO, and CHOS) from (+)ESI FT-OTE MS analysis of dissolved organic matter mixtures from the International Humic Substances Society. Standard deviations are given in brackets. The same information is provided for all molecular formula assignments together in Table 3.2.

	SRFA		PLFA		NLFA		ESFA		SRNOM		MSNOM	
	CHO	CHNO	CHO	CHNO	CHO	CHNO	CHO	CHNO	CHO	CHNO	CHO	CHNO
Number of formulae^a	4030	1586	2735	2786	2999	1565	3613	1770	3933	1422	3202	2025
H/C_{avg}	1.19 (0.3)	1.06 (0.3)	1.37 (0.3)	1.29 (0.2)	1.18 (0.3)	1.13 (0.4)	1.32 (0.3)	1.45 (0.3)	1.19 (0.3)	1.25 (0.3)	1.19 (0.3)	1.19 (0.3)
O/C_{avg}	0.44 (0.2)	0.45 (0.2)	0.38 (0.1)	0.37 (0.1)	0.42 (0.2)	0.40 (0.2)	0.46 (0.2)	0.44 (0.1)	0.46 (0.2)	0.46 (0.2)	0.43 (0.2)	0.42 (0.2)
DBE_{avg}	12.1 (5.1)	12.4 (4.9)	9.2 (4.0)	9.9 (3.4)	11.1 (4.6)	11.9 (5.6)	9.1 (4.5)	6.6 (3.2)	11.7 (5.0)	8.7 (3.9)	11.0 (4.3)	9.7 (4.3)
C_{avg} atoms^b	27.5 (9.4)	22.5 (5.8)	26.0 (8.8)	23.2 (7.6)	24.7 (7.2)	22.3 (7.1)	24.0 (7.9)	17.8 (5.9)	26.1 (8.6)	18.3 (5.9)	25.0 (7.9)	19.4 (6.6)
O_{avg} atoms^b	11.8 (4.9)	9.9 (4.1)	9.5 (3.8)	8.4 (3.3)	10.2 (4.2)	8.2 (4.2)	10.7 (4.0)	7.7 (3.4)	11.8 (4.9)	8.4 (3.9)	10.4 (4.4)	8.0 (4.0)
N_{avg} atoms^b	--	1.26 (0.4)	--	1.54 (0.7)	--	1.51 (0.6)	--	1.22 (0.5)	--	1.15 (0.4)	--	1.35 (0.6)
Averages (standard deviations) for IONS EXCLUSIVELY PRESENT in each DOM mixture												
Number of formulae^a	306	311	290	1396	154	517	220	324	238	153	140	359
H/C_{wa}	1.16 (0.3)	0.98 (0.3)	1.34 (0.3)	1.29 (0.2)	1.31 (0.5)	1.18 (0.4)	1.60 (0.4)	1.65 (0.3)	1.04 (0.2)	1.08 (0.4)	1.24 (0.3)	1.17 (0.4)

O/C_{wa}	0.42 (0.1)	0.48 (0.2)	0.31 (0.1)	0.35 (0.1)	0.45 (0.3)	0.37 (0.3)	0.52 (0.1)	0.44 (0.2)	0.60 (0.2)	0.60 (0.2)	0.35 (0.3)	0.40 (0.2)
DBE_{wa}	17.36 (5.2)	14.54 (5.2)	11.96 (3.8)	11.35 (2.9)	9.93 (6.6)	12.37 (5.9)	5.56 (3.3)	5.22 (2.9)	15.78 (4.8)	10.88 (4.5)	10.67 (5.6)	10.45 (5.0)
C_{wa} atoms^b	39.33 (9.2)	25.48 (6.3)	34.80 (9.0)	27.20 (6.5)	26.14 (7.6)	24.58 (8.0)	25.50 (7.8)	19.37 (6.5)	31.01 (8.0)	19.67 (6.6)	25.68 (9.4)	20.65 (8.0)
O_{wa} atoms^b	16.36 (4.9)	11.85 (5.9)	10.68 (4.3)	9.36 (3.2)	10.71 (5.5)	7.80 (4.7)	12.58 (3.5)	8.26 (3.6)	17.86 (5.0)	11.18 (4.3)	8.64 (6.2)	8.28 (5.0)
N_{wa} atoms^b	--	1.40 (0.5)	--	1.80 (0.8)	--	1.81 (0.7)	--	1.56 (0.8)	--	1.43 (0.5)	--	1.729 (0.8)

^aThis is the number of formulae that were finalized after blank subtraction and manual data validation and include ¹²C, ¹H, ¹⁶O, ¹⁴N, and ³²S. Polysotopologues are not included in this number; ^bNumber of atoms of the said elements per molecule.

Table S3.5 Summary of molecular formulae that were common to all six dissolved organic mixtures from the International Humic Substances Society using negative and positive ESI FT-OTE MS analysis. Standard deviations are given in brackets. This information is visualized in Figure S3.4

	All formulae	CHO	CHNO	CHOS
(-)ESI				
Number of formulae^a	1385	1145	226	14
H/C_{avg}	1.13 (0.2)	1.17 (0.2)	0.91 (0.1)	1.14 (0.1)
O/C_{avg}	0.44 (0.1)	0.44 (0.1)	0.44 (0.1)	0.42 (0.1)
DBE_{avg}	9.38 (2.8)	9.12 (2.9)	10.90 (1.6)	6.00 (1.4)
C_{avg} atoms^b	19.48 (6.7)	20.07 (7.0)	16.82 (3.2)	11.64 (2.1)
O_{avg} atoms^b	8.48 (3.6)	8.65 (3.8)	7.40 (1.9)	5.00 (1.4)
N_{avg} atoms^b	0.26 (0.6)	--	1.61 (0.5)	--
(+)ESI				
Number of formulae^a	1338	1163	175	--
H/C_{avg}	1.27 (0.2)	1.27 (0.2)	1.30 (0.2)	--
O/C_{avg}	0.41 (0.1s)	0.41 (0.1)	0.44 (0.1)	--
DBE_{avg}	9.38 (2.9)	9.59 (2.9)	7.95 (2.4)	--
C_{avg} atoms^b	23.06 (6.5)	23.79 (6.5)	18.20 (4.6)	--
O_{avg} atoms^b	9.21 (3.0)	9.39 (3.0)	8.01 (2.4)	--
N_{avg} atoms^b	0.13 (0.3)	--	1.01 (0.1)	--

^aThis is the number of formulae that were finalized after blank subtraction and manual data validation and include ¹²C, ¹H, ¹⁶O, ¹⁴N, and ³²S. Polyisotopologues are not included in this number; ^bNumber of atoms of the said elements per molecule.

Table S3.6 Jaccard similarity percent between all possible pairs of dissolved organic matter mixtures from the International Humic Substances Society. Similarity was calculated using presence and absence of molecular formulae identified by (+)ESI FT-OTE MS analysis. Percentage of molecular formulae unique to each mixture is also given. This information is presented visually as a matrix of van Krevelen diagrams in Figure S3.7.

SRFA	SRFA					
PLFA	Common = 30.79% SRFA = 35.193 PLFA = 34.021	PLFA				
NLFA	Common = 44.85% SRFA = 34.925 NLFA = 20.226	Common = 23.19% PLFA = 44.107 NLFA = 32.700	NLFA			
ESFA	Common = 44.35% SRFA = 32.495 ESFA = 23.156	Common = 37.65% PLFA = 35.015 ESFA = 27.339	Common = 34.36% NLFA = 30.3367 ESFA = 35.310	ESFA		
SRNOM	Common = 59.31% SRFA = 22.302 SRNOM = 18.387	Common = 34.67% PLFA = 33.720 SRNOM = 31.618	Common = 39.07% NLFA = 25.104 SRNOM = 35.826	Common = 54.25% ESFA = 19.778 SRNOM = 25.976	SRNOM	
MSNOM	Common = 46.5% SRFA = 29.242 MSNOM = 24.263	Common = 32.11% PLFA = 35.603 MSNOM = 32.296	Common = 48.18% NLFA = 20.896 MSNOM = 30.931	Common = 42.45% ESFA = 26.584 MSNOM = 30.9690	Common = 50.48% SRNOM = 25.471 MSNOM = 24.053	MSNOM

Table S3.7 Recalibrant series used to correct systematic bias in the exact masses of ions measured by negative and positive ESI FT-OTC MS analysis

Recalibrant series used			Sample	Average formula error of preliminary C, H, and O-containing formulae (ppm)	Average formula error of preliminary C, H, and O-containing formulae after correction of mass (ppm)	Average formula error of final formula assignments, including C, H, O, N- and/or S-containing formulae (ppm)
Adduct	Oxygen class	DBE				
(-)ESI						
H	O4 O18 O9 O20	7 22 9 21	SRFA	0.663	0.457	0.613
			PLFA	1.497	1.428	0.827
			NLFA	2.105	0.540	0.629
			ESFA	0.901	0.628	0.644
			SRNOM	0.575	0.449	0.585
			MSNOM	1.229	0.612	0.612
(+)ESI						
Na	O5 O8 O14 O16	4 8 12 15	SRFA	1.006	0.609	0.638
			PLFA	1.467	1.222	0.753
			NLFA	1.561	0.805	0.913
			ESFA	1.106	0.642	0.650
			SRNOM	1.058	0.654	0.631
			MSNOM	1.514	0.806	0.822

3.6.3 Supplementary figures

Figure S3.1 Reconstructed mass spectra of dissolved organic mixtures from the International Humic Substances Society analyzed by (-)ESI FT-OTE MS. Peaks in black and green on the positive ordinate represent the raw mass list and the peaks therein that were assigned a formula, respectively. Peaks in red on the

negative ordinate are those that could not be assigned a formula. Insets show the same information in terms of cumulative total ion abundance, which is basically total ion current. These plots were drawn using molecular formulae before blank subtraction and manual data validation.

Figure S3.2 Reconstructed mass spectra of dissolved organic mixtures from the International Humic Substances Society analyzed by (+)ESI FT-OTE MS. Peaks in black and green on the positive ordinate represent the raw mass list and the peaks therein that were assigned a formula, respectively. Peaks in red on the negative ordinate are those that could not be assigned a formula. Insets show

the same information in terms of cumulative total ion abundance, which is basically total ion current. These plots were drawn using molecular formulae before blank subtraction and manual data validation.

Figure S3.3 Reconstructed mass spectra and elemental ratios from (+)ESI FT-OTE MS analysis of dissolved organic matter mixtures from the International Humic Substances Society. These plots were drawn with molecular formulae obtained after blank subtraction and manual data validation. Colours denote molecular groups: CHO, green; CHNO, blue; CHOS, red, CH, brown; CHN, cyan. Pie charts represent the proportion of each molecular group in a given mixture as summarized in Table S3.4.

Figure S3.4 Reconstructed mass spectra of dissolved organic matter mixtures from the International Humic Substances Society cropped at the same normalized abundance of 0.03. Coloured peaks were common to all six mixtures included in this study ($n = 1385$ and 1338 for negative and positive ESI, respectively). Note that they have different peak intensities in each mixture. These common peaks are overlaid on other (either unique to the sample or

shared with ≤ 5 mixtures) peaks drawn in black. Numerical details of this set of ions is provided in Table S5.

Figure S3.5 van Krevelen diagrams with molecular formulae assigned to dissolved organic matter mixtures from the International Humic Substances Society by negative and positive ESI FT-OTE MS analysis. Colours represent polarities: Exclusively ionized by (-)ESI, red; exclusively ionized by (+)ESI, green; ionized by ESI in both polarities, black. For each sample, formulae from both ionizations were aligned; percentage of formulae that were unique to either polarity in this list or were common to both is mentioned.

Figure S3.6 Distribution of molecular formulae exclusively assigned to each of the six dissolved organic matter mixtures from the International Humic Substances Society with respect to modified aromaticity index and other broad chemical classes described by Koch and Dittmar (2006; 2016) and Hawkes et al. (2020).

Figure S3.7 All possible pairwise combinations of the six dissolved organic matter mixtures from the International Humic Substances Society included in our study are visualized as a matrix of van Krevelen diagrams after being analyzed by (+)ESI FT-OTE MS. Molecular formulae that were present in both members of a pair are represented in gray in the background, and those unique to each mixture are shown by specific colours used in the column/row headers. This information is presented numerically in terms of Jaccard similarity percent in Table S3.6.

4 High-Resolution Molecular Characterization of Tar Ball-Rich Biomass Burning Organic Aerosol with a 21-T FT-ICR Mass Spectrometer

Abstract

Aerosols that originate from or have encountered large-scale biomass burning events, such as wildfires, are loaded with organic particulate matter of immense environmental relevance that plays a key role in global radiative balance, air quality, biogeochemical cycling, and human health. We analyzed a biomass burning organic aerosol (BBOA) mixture collected downwind of wildfires in Richland, Washington, with a 21-Tesla (21-T) Fourier-transform - Ion Cyclotron Resonance mass spectrometer (FT-ICR MS). This mixture was formerly established to be extremely chemically rich as demonstrated by identification of 5453 distinct molecular formulae (m/z 100-800) by an FT - Orbitrap Elite MS in conjunction with negative-mode electrospray ionization ((-)ESI). With a 5-fold increase in resolving power at m/z 400, 10,533 distinct formulae were assigned to this aerosol mixture using (-)ESI/21-T FT-ICR MS, while only a 1.1-fold increase was recorded with the 15-T FT-ICR MS. Despite extensive differences in mass and intensity distributions, as well as detection of ~5000 molecular species exclusively by the 21-T FT-ICR MS, average elemental and structural information, in terms of O/C_{avg} , H/C_{avg} , and DBE_{avg} , retrieved from the three instruments was similar, although not identical. Molecular-level characterization of this BBOA mixture provided evidence of extensive oxidative atmospheric processing and oligomerization, as well as the presence of sulphate and nitrate/nitro functionalities. Notably, P-containing species detected exclusively by 21-T FT-ICR MS analysis have been described, exemplifying the need of ultrahigh-resolution tools for a more detailed and accurate depiction of the molecular composition of organic aerosol mixtures.

4.1 Introduction

The true molecular composition of complex environmental mixtures is a mystery that eludes even the most sophisticated analytical instrumentation. The extent of molecular-level information that can be obtained from complex mixtures has

increased greatly with advancement in high-resolution mass spectrometers. A heavy distillate fraction (bitumen) of crude oil was reportedly fine combed to identify ~50,000 distinct molecular species with a 21-T Fourier transform - Ion Cyclotron Resonance mass spectrometer (FT-ICR MS)¹³³. Since the first reports on the research application of a 21-T FT-ICR MS in 2015-16^{88, 134}, it has been used to characterise large biomolecules, such as proteins¹³⁵⁻¹⁴¹, bacterial siderophores¹⁴², complex environmental mixtures, including petroleum¹⁴³, various types of crude oil, and dissolved organic matter¹³³ (Figure S4.1). These mass spectrometers have also been coupled with liquid-chromatography systems^{133, 144} and different ionization methods^{145, 146} to widen the scope of their application. Researchers are limited not by a want of advancement in mass spectrometric instrumentation, but in meaningful interpretation and utilization of the fine details made available by these instruments.

The application of high-resolution mass spectrometers is especially important for compositional analysis of atmospheric organic aerosols. Due to their central role in controlling our Planet's radiative balance, biogeochemical cycling, air quality, and human health, aerosols constitute one of the nine variables that must be regulated globally to prevent transgression of planetary boundaries¹⁴⁷. Atmospheric plumes that encounter large-scale biomass burning events, such as wildfires, are expected to be loaded with particulate matter comprising mostly organic compounds¹⁴⁸⁻¹⁵¹. The chemical, physical, and microphysical properties of biomass burning organic aerosol (BBOA) drive the profound influence that it has on climate and the biotic/abiotic entities that it interacts with. Exact elemental compositions of molecular components of organic aerosol have informed parameterizations that predict the volatility of organic aerosol mixtures⁹⁵, glass transition temperatures^{96, 152}, phase states⁹⁷, and average carbon oxidation states in molecules⁹⁴. Therefore, detailed molecular profiling of organic aerosol can not only inform such computational modelling beneficial to climatic studies, but can also provide valuable insight into the atmospheric processing, reactivity, and fate of organic aerosols, yet we have not explored the depths of BBOA molecular composition that may be achieved with a 21-T FT-ICR MS.

Recently, we reported molecular characterization of a BBOA mixture of unprecedented molecular complexity that was collected downwind of wildfires in Richland, Washington, USA, in September 2017⁴¹. This tar ball (TB)-rich aerosol mixture was assigned 5453 distinct unambiguous molecular formulae by FT - Orbitrap Elite (FT-OTE) MS analysis, in conjunction with negative-mode electrospray ionization ((-)ESI). Here, we analyzed this mixture with (-)ESI/15-T and (-)ESI/21-T FT-ICR mass spectrometers to demonstrate a 1.1-fold and 1.9-fold increase in the number of distinct molecular formulae assigned. Despite the improvement in information acquired on the mixture's molecular components with the 21-T FT-ICR MS, average composition determined for the entire mixture

in terms of O/C_{avg} , H/C_{avg} , and DBE_{avg} deviated only slightly across the three instruments. The apparent enhancement in molecular complexity manifested most noticeably in the detection of species that were present in near-negligible amounts in this complex mixture, such as P-containing compounds. This was largely attributable to enhanced sensitivity of the 21-T FT-ICR MS. Our findings exemplify the need of ultrahigh-resolution mass spectrometry for a fine-scale understanding of the molecular constituents of BBOA mixtures. The primary focus of this report is to present chemical and molecular characterization of a wildfire-influenced BBOA complex mixture using a 21-T FT-ICR MS. Where necessary, we compare our observations with those from the other two mass spectrometers. Nonetheless, we recognise that a direct comparison across the three instruments is not possible, considering that they differed not only in their resolving powers, but also in basic instrumental designs, ESI sources, and data acquisition and processing.

4.2 Methods

4.2.1 Sample collection and preparation

Aerosol samples were collected on 90-mm PTFE filters by pulling 80 L of air per minute for 6 hours on September 05, 2017 at the Pacific Northwest National Laboratory downwind of wildfires that burned in central Washington State as described before⁴¹. A small piece (1/4th) of the filter was immersed in analytical-grade acetonitrile and shaken at 60 rpm for 90 minutes. The filter was then discarded, and liquid ACN-aerosol mixture was passed through a 0.2- μ m PTFE syringe to remove suspended solid material. The extract thus obtained was maintained at 4°C during storage and shipment until analysis.

4.2.2 Electrospray ionization and mass spectrometry

4.2.2.1 15-T FT-ICR MS analysis

Immediately prior to analysis, the sample was diluted 10-fold in 50:50 H₂O:acetonitrile. Spectra were acquired on a Bruker Solarix XR FT-ICR MS equipped with a para-cell and an actively shielded 15-T superconducting magnet. Samples were infused at 5 μ L/min into an Apollo II ESI source operating in negative mode with a capillary voltage of 4 kV, dry gas flow of 4 L/min, temperature of 180°C, and nebulizer gas pressure of 1 bar. The Q1 mass was set to m/z 100. Ions were accumulated for 150 ms prior to transfer to the analyser cell with a time-of-flight of 0.75 ms. All spectra were acquired in a scan range of m/z 153.6 and 1200.0 into an 8 MWord time domain with transient length of 2.8 s. Three hundred transients were coadded prior to zero filling and Fourier transformation. The measured resolving power was about 1,000,000 at m/z 400. Spectra were peak picked in DataAnalysis 5.0 (Bruker Daltonics, Billerica MA) with an S/N threshold of 4 and absolute intensity threshold of 1,000,000. The resulting mass list was exported to MFAssignR for further processing.

4.2.2.2 21-T FT-ICR MS analysis

Samples were diluted 10-fold in acetonitrile prior to analysis. Spectra were acquired on a custom built 21-T FT-ICR MS as previously described⁸⁸. Briefly, a Velos Pro linear ion trap front end is used for ionization and system control. It is coupled to a custom build ICR flight tube and ICR analyser cell. Samples were infused at 4 μ L/min into a heated ESI source (45°C) with a spray voltage of 3.4 kV in negative polarity with 8 arbitrary units of sheath gas to aid in nebulization and an inlet capillary temperature of 300°C. An AGC target of 5×10^5 was set with a typical ion injection time of 5 ms. Spectra were recorded between m/z 150 and 800 with a resolution setting of 200,000 – equivalent to a transient length of 1.5 s and theoretical resolution of 600,000 at m/z 400. One thousand scans were recorded with single microscan events.

Data were recorded with the FTMS Booster from SpectroSwiss allowing for co-addition of raw transients post-acquisition and absorption-mode Fourier-transform processing, yielding improved resolution and sensitivity. For this,

PeakByPeak (SpectroSwiss) was used to co-add, Fourier transform, and phase correct all scans. The mass spectrum was then peak picked based on logarithmic noise thresholding¹⁵³ with an S/N threshold of 6. The resulting mass list was exported to MFAssignR for further processing.

4.2.3 Data processing and molecular formula assignment

Mass lists comprising peaks with $S/N \geq 6$ (≥ 4 for 15-T FT-ICR MS) were subjected to further noise estimation using the `KMDNoise()` function in MFAssignR⁹³ to remove residual noise, if any. Monoisotopic and polyisotopic peaks were identified using `IsoFiltR()`. CH_2 -homologous series were chosen from preliminary formulae containing only C, H, and O as internal recalibrants using `Recal()` for correction of systematic bias in ion masses. Molecular formulae were assigned to recalibrated m/z values using `MFAssign()` within the following constraints: $\text{C}_c\text{H}_h\text{O}_o\text{N}_{0-3}\text{S}_{0-1}\text{P}_{0-1}^{13}\text{C}_{0-2}^{34}\text{S}_{0-1}$ (numbers of C, H, and O were unrestricted); $-13 \leq \text{DBE-O} \leq 20$; $0 \leq \text{O/C} \leq 2.0$; $0.3 \leq \text{H/C} \leq 2.5$, and maximum permissible error of 0.5 ppm. Unless otherwise indicated, assignments containing $^{13}\text{C}_{1-2}$ and/or ^{34}S atoms are not included in our discussion, considering that they are duplicates of corresponding monoisotopic formulae. Peak intensities were normalized to the intensity of the tallest peak. Structural information was inferred by calculating double bond equivalence (DBE) and modified aromaticity index (AI_{mod})^{99, 100}. Chemical and physical information: Saturation mass concentration (C^0) and associated volatility classification^{30, 95}, average oxidation state of C (OS_c)⁹⁴, dry⁹⁶ and relative humidity-dependent glass transition temperatures ($T_{\text{g,dry}}$ and $T_{\text{g,RH}}$), and phase state ratios (PSR)⁹⁷ were computed as described in their respective reports, as well the supplementary material of this report.

4.3 Results and Discussion

4.3.1 Chemical and molecular composition

Biomass burning organic aerosol is polydisperse and multifunctional with C, H, O, N, S, and P constituents present as combinations of hydroxyls, carboxyl, carbonyl, and ester functionalities and/or nitrates, sulphates, or

nitroxysulphates^{41, 154-156}. The immense chemical richness of BBOA results from its primary organic components that are directly produced from the combustion source and secondary organic substances that are formed as aerosol travel and transform along atmospheric currents. The BBOA mixture analyzed here was reported to have 5453 distinct molecular formulae from m/z 100-800 belonging to CHO, CHNO, and CHOS molecular groups using FT-OTE MS analysis⁴¹. When analyzed with a 15-T FT-ICR MS, the number of monoisotopic formulae assigned increased to 6053 within the same scan range with an additional 1383 formulae containing $^{13}\text{C}_{1-2}$ and/or ^{34}S . The 21-T FT-ICR MS detected almost twice the number of monoisotopic species, i.e., $n = 10,533$, and 3420 polyisotopic species from m/z 150-850 with the addition of two molecular groups, CHOP and CHNOP (Figure 4.1, Table 4.1). Overall, 63% of the total ion current, including noise, could be assigned a formula in the 15-T FT-ICR MS and 75% in the 21-T FT-ICR MS (Figure S4.2). Formula error was centred at zero for all three mass spectrometers, but both FT-ICR instruments offered a much narrower error range of ± 0.2 ppm as compared to ± 3.0 ppm for the FT-OTE MS (Figure S4.3).

Figure 4.1 Reconstructed mass spectra for BBOA mixture acquired from 15-T and 21-T FT-ICR mass spectrometers, where the ordinate has been normalized to the intensity of the tallest peak. In case of 15-T FT-ICR mass spectrum, the ordinate is not equal to 1.0 as most tall peaks were removed as sulfonate contaminant peaks or remained unidentified. van Krevelen diagrams (on the right) show continuous elemental distributions, and the pie-chart insets represent percent fraction of formulae belonging to each of the five molecular

groups (CHO, CHNO, CHOS, CHOP, and CHNOP). These diagrams exhibit only the finalized formulae, while contaminant or outlier/suspicious formulae are not depicted.

The largest fraction of assigned formulae belonged to CHNO species (~50%) in both FT-ICR instruments with CHO species being a close second. Table 4.1 presents the number of molecular formulae identified by the two instruments and the average values of compositional parameters. Previously, for this aerosol mixture, Brege et al.⁴¹ observed a continuum of C that extended from C₃ to C₄₅ and encompassed aliphatic, aromatic, and condensed aromatic species with a continuous range of O atoms and DBE values using (+/-)APPI and (+/-)ESI and an FT-OTE MS. Based on AI_{mod} values, a majority of species were identified to be olefinic in nature by Brege et al.⁴¹, which was corroborated by both FT-ICR instruments employed here. In 21-T FT-ICR analysis, 14.68% of the species were aromatic or condensed aromatic, 18.41% were aliphatic, while ~47.56 and 19.33% (66.89% altogether) were low- and high-O unsaturated, i.e., olefinic species. Similar distributions were seen with 15-T FT-ICR analysis: 16.68% species were aromatic or condensed aromatic, 23.01% were aliphatic, while ~30% each (60.3% altogether) were either low- or high-O unsaturated species. Generally, organic emissions from anthropogenic and biogenic sources are rich in aliphatic or olefinic species, including alkanes, alkenes, alkanoic and alkenoic acids, etc.^{154, 155} that are effective precursors for formation of secondary organic aerosol. The substantial fraction of aromatic species observed for the BBOA mixture analyzed here is because of a strong interaction of this aerosol with wildfires in the 48-hours preceding its collection⁴¹.

Brege et al.⁴¹ reported average O/C, H/C, and DBE of 0.51, 1.18, and 9.19 from (-)ESI/FT-OTE MS analysis. Average values of these parameters were 0.56 ± 0.2 , 1.19 ± 0.3 , and 9.38 ± 4.4 from 15-T (-)ESI/FT-ICR MS analysis and 0.46 ± 0.2 , 1.20 ± 0.3 , and 10.38 ± 4.8 from 21-T (-)ESI/FT-ICR MS analysis. Despite the immense difference in numbers and types of molecular species identified by the three instruments, bulk chemical information provided by them for this BBOA mixture was thus very much alike. As described in the following sections, the benefits of exceptional resolving power and sensitivity afforded by the 21-T FT-ICR MS are obvious only when species exclusively detected by it above the S/N ratio were considered. Due to the similarity of molecular compositions recorded using the two FT-ICR instruments, the following commentary is based on observations from the 21-T FT-ICR MS, unless noted otherwise.

4.3.2 CHO species

Species constituting only C, H, and O formed the second largest fraction of all molecular formulae assigned to this TB-rich BBOA mixture, but it accounted for the most ion current as the tallest peaks throughout the scan range belonged to this group (Figure 4.1). Formulae assigned exhibited a continuous range of oxygenation: O-O₁₉ with O_{avg} of 9.46 ± 3.9 (Figure S4.4). The highest C_{avg} was noted for this molecular group at 25.15 ± 7.6 with number of C atoms per molecule ranging from 6 to 41 (Figure S4.4). The O/C_{avg} and H/C_{avg} of all CHO formulae were 0.40 ± 0.2 and 1.22 ± 0.3 . Their DBE also covered a wide range from 1 to 23 with the most ion intensity originating from species with DBE of 6-12 (Figure S4.4) and DBE_{avg} of 10.80 ± 5.0 . The intermediate DBE values were supported by Al_{mod,avg} of 0.30 ± 0.2 that indicated most species to be olefinic.

Most direct products of cellulose pyrolysis are CHO compounds, such as levoglucosan (C₆H₁₀O₅, 162.052824 Da)^{157, 158} and cellobiosan (C₁₂H₂₀O₁₀, 324.105649 Da)¹⁵⁹. Levoglucosan is a particle-phase marker of biomass burning and was observed only in 15-T FT-ICR MS analysis with 3% relative intensity. It was outside the *m/z* range covered by the 21-T FT-ICR MS where the smallest analyte ion was observed at 182 Da. Levoglucosan's dimer, cellobiosan, was detected only by the 21-T FT-ICR MS with 4.0% relative intensity. Other pyrolysis products of lignin include syringol (C₈H₁₀O₃, 154.062995 Da), guaiacol (C₇H₈O₂, 124.052430 Da)¹⁵⁸, vanillic acid (C₈H₈O₄, 168.042260 Da), syringaldehyde (C₉H₁₀O₄, 182.057910 Da), and syringic acid (C₉H₁₀O₅, 198.052824 Da)¹⁵⁵. In the 15-T FT-ICR MS, these were detected at relative intensities of 0.0 (not detected), 0.0, 9.0, 7.6. and 6.6%, while in the 21-T FT-ICR MS, we recorded only syringic acid with 9.7% relative intensity. Although syringol and guaiacol were not detected by either instrument, molecular formulae corresponding to their dimers (C₁₆H₁₈O₆; 306.110340 Da and C₁₄H₁₄O₄; 246.089210 Da), trimers (C₂₄H₂₆O₉; 458.157685 Da and C₂₁H₂₀O₆; 368.125990 Da), and tetramers (C₃₂H₃₄O₁₂; 610.205030 Da and C₂₈H₂₆O₈; 490.162770 Da) were detected by both instruments. Relative intensities recorded by the 21-T FT-ICR MS for the aforementioned syringol oligomers were quite high at 37.7, 27.7, and 6.1%. Relative intensities for guaiacol oligomers were similarly substantial at 21.4, 22.9, and 11.67%. A potential pentamer was detected only for guaiacol (C₃₅H₃₂O₁₀; 612.199550 Da) exclusively by the 21-T FT-ICR MS at 2.0% relative intensity. Owing to isomeric complexity, these formulae do not guarantee the presence of said oligomers. However, molecular species observed at *m/z* values >400 are considered to be an indication of oxidative aging and/or oligomerization¹⁵⁵, which lends reasonable credibility to our deductions .

Overall, in this sample, 85% of the molecular formulae were classified as extremely low volatility or low volatility organic compounds (ELVOC or LVOC) based on estimated C⁰⁹⁵. Amongst the CHO species, 86.6% were E/LVOC and 10.67% were semi volatile organic compounds (SVOC). The distributions of CHO

species with respect to m/z and $T_{g,RH}$ is depicted in Figure 4.2. Estimations of $T_{g,RH}$ and PSR^{97} showed 99.2% of CHO formulae (99.2%) to be present as solids or semi-solids under the meteorological conditions on the day of sample collection (RH and T_{amb} on Sep 05, 2017 were 30.5% and 300.15 K) with the rest being liquids. The liquid compounds were smaller species ($m/z < 200$) with C_{10-17} and O_{1-3} (Figure 4.2). Both C^0 and PSR estimations are corroborated by previous investigations on this BBOA mixture, where ~85-85% of particles sampled from it were morphologically identified to be TB with a dominance of non-volatile compounds⁴¹. Presence of molecular species that are likely low-volatility, solid building blocks of TB reiterates that this BBOA mixture was atmospherically processed during transport with the consequential production of SOA from particle-phase oligomerization of POA.

Table 4.1 Summary of bulk chemical metrics based on molecular formulae assigned to the BBOA mixture analyzed here. Considering that mass distributions and signal-to-noise ratio (SNR) recorded by the two instruments differed drastically, number averages and standard deviations are reported.

		Number of raw peaks ^a	Monoisotopic peaks ^b	Polyisotopic peaks ^b	Unassigned peaks ^c	Assigned MF ^d	O/C _{avg}	H/C _{avg}	DBE _{avg}	m/z _{avg}	AE _{avg}	AI _{mod,avg}
							Oxygen-to-carbon ratio	Hydrogen-to-carbon ratio	Double bond equivalence	Mass/charge	Absolute error (ppm)	Modified aromaticity index
15-T FT-ICR MS												
All MF		23701	8536	1403	2447	6053	0.56 (0.2)	1.19 (0.34)	9.38 (4.4)	430.85 (117.8)	0.066 (0.05)	0.29 (0.5)
						2618	0.48 (0.2)	1.12 (0.3)	10.3 (4.3)	433.98 (126.6)	0.063 (0.05)	0.33 (0.2)
CHNO						3080	0.60 (0.2)	1.22 (0.3)	9.20 (4.1)	441.47 (107.1)	0.066 (0.05)	0.27 (0.6)
CHOS						355	0.75 (0.3)	1.49 (0.3)	4.00 (2.6)	315.56 (68.4)	0.077 (0.05)	0.12 (0.6)
CHOP		Not detected										
CHNOP		Not detected										
21-T FT-ICR MS												
All MF	34557	11845	3467	1294	10533	0.46 (0.2)	1.23 (0.3)	10.38 (4.8)		474.06 (124.2)	0.032 (0.03)	0.28 (0.3)
	CHO				4629	0.39 (0.2)	1.22 (0.3)	10.80 (5.0)		483.06 (136.2)	0.027 (0.02)	0.30 (0.2)

CHNO				5022	0.45 (0.2)	1.17 (0.2)	11.18 (4.0)	474.43 (116.6)	0.031 (0.03)	0.29 (0.2)
CHOS				190	0.70 (0.3)	1.64 (0.2)	3.22 (1.8)	324.00 (64.7)	0.026 (0.02)	0.06 (0.5)
CHOP				488	0.83 (0.2)	1.75 (0.3)	3.66 (2.2)	455.93 (65.6)	0.078 (0.04)	0.002 (0.0)
CHNOP				204	1.02 (0.2)	1.78 (0.2)	3.74 (1.9)	444.05 (57.19)	0.077 (0.04)	0.36 (2.0)

^aTotal number of peaks in the mass list above the signal-to-noise cut-off; ^bNumber of peaks that were identified as monoisotopic and polyisotopic by IsoFiltR() in MFAssignR; ^cNumber of monoisotopic and polyisotopic peaks that could not be assigned a molecular formula within the specified constraints; ^dNumber of formulae assigned to monoisotopic peaks with ¹²C, ¹H, ¹⁶O, ¹⁴N, ³²S, and ³¹P only.

Figure 4.2 Volatility and viscosity classification of molecular formulae containing only C, H, and O using 21-T FT-ICR MS: (A) Volatility classes and (B) Phase states. In (A), identified species are distributed across a continuum of $\log_{10}$ of saturation mass concentration (C_0) and relative-humidity dependent glass transition temperature ($T_{g,RH}$). Based on the former parameter, a majority of species are classified as extremely low volatility organic compounds (ELVOC), followed by low volatility, semi volatile, and intermediate volatility organic compounds (LVOC, SVOC, and IVOC). Fully volatile organic compounds (VOC) were not detected in this mixture. The green dotted line denotes ambient temperature ($T_{amb} = 300.15$ K on September 05, 2017). Species with $T_{g,RH} > T_{amb}$ likely exist as solids and those with $0.8 \times T_{amb} < T_{g,RH} < T_{amb}$ are expected to be semi-solids. Those with $T_{g,RH} < 0.8 \times T_{amb}$ (blue dotted line) will exist as liquids. Volatility of organic compounds tends to decrease with the m/z . In (B), a majority of identified species are classified as solids. As we move toward species with higher O numbers, the fraction of solid components increases. This is not necessarily because of enhanced oxygenation, but mostly because of the tendency of larger and more viscous compounds to hold more O atoms.

4.3.3 CHNO species

We assigned 5022 N-containing formulae in total, i.e., the largest fraction of all formulae assigned, by 21-T FT-ICR MS analysis. These formulae belonged to 41 heteroatom classes laying across NO_2 - NO_{18} , N_2O_3 - N_2O_{16} , and N_3O_4 - N_3O_{13} with number of formulae assigned to each class decreasing in the order listed (Figure 4.3). The majority of formulae assigned were concentrated between DBE values of 7 to 16. Average O/C was 0.45 ± 0.2 and H/C_{avg} was 1.17 ± 0.2 for the CHNO molecular group. Previously, BBOA has been reported to be rich in CHNO species^{156, 160}. The distribution of CHNO species in the van Krevelen space was the same as CHO species, except for some highly saturated and oxygenated

species observed exclusively in 15-T FT-ICR analysis (Figure 4.1). These aliphatic CHNO species were present from m/z 319-519 (Figure S4.6). In the 21-T FT-ICR MS, 96.32% of all CHNO species identified were classified as E/LVOC with the rest being SVOC (Figure 4.3). The fraction of less volatile species increased as the number of N atoms per molecule increased. Physical parameters, such as $T_{g,RH}$ and PSR cannot be estimated for heteroatom-containing molecules using Li et al.'s model⁹⁵ although $T_{g,dry}$ for CHNO and CHOS can be estimated using a newer model (data not presented here)¹⁵².

Figure 4.3 Distribution of N-containing species detected by the 21-T FT-ICR MS across N_nO_o molecular classes and corresponding structural and physical nature. (A) DBE trends, (B) Modified aromaticity index trends, and (C) Volatility trends.

The CHNO species identified in this BBOA mixture were predominantly low-O olefinic compounds (Figure 4.3). The species with N_{1-2} had a considerable fraction of aromatic species as well, which are likely nitroaromatic compounds previously reported to be present in TB²⁴. Aliphatic and condensed aromatic species were present, but only sparsely. Average O/N values of 7.91 ± 4.1 and a minimum of at least two O atoms per CHNO molecule, along with a high average number of O atoms per molecule (9.76 ± 3.3), indicates these compounds to comprise organic species with up to three nitro (-NO₂) or nitrate (-ONO₂) groups along with other O functional groups. It must be noted that the elemental ratios alone are not sufficient to draw conclusions about the presence of these functional groups. However, reasonable inferences can be drawn based on findings of fragmentation studies on atmospheric organic aerosol that have reported loss of HNO₃^{59, 161}, HNO₂, and NH₃ from BBOA⁵⁹. There were 43 CHNO species with three N atoms whose O/N values were <2, i.e., they had an insufficient number of O to have -NO₂ or -ONO₂ functionalities. Considering that the N₃O_x class was detected exclusively by the 21-T FT-ICR MS, these compounds may likely be compounds of low ionizability that remained undetected in the 15-T FT-ICR and FT-OTE instruments. Their DBE values ranged from 7-12 and they were mostly classified as olefinic species with a small proportion of aromatic species (Figure 4.3). They could be unsaturated species containing reduced N, for instance, amine (-NH₃) or azo (-N₂) functional groups, with at least one other polar functional group as speculated previously for water-soluble organic aerosol¹⁵⁴. They could also be aromatic species, such as phenols, naphthalenes, or benzoic acids, substituted with nitroso (-NO) groups.

4.3.4 CHOS species

The 21-T FT-ICR MS detected 190 CHOS species with much lower intensities than CHNO and CHO species. The number of O atoms in these compounds ranged from 5-12 with O/S_{avg} of 7.61 ± 1.5 . Figures S4.5 shows the distribution of formulae belonging to these molecular groups across O_oS_s classes. These species had very low overall Al_{mod,avg} of 0.06 ± 0.5 and DBE_{avg} of 3.2 ± 1.8 suggesting CHOS species to be very H-saturated and highly oxygenated. This was supported by the O/C_{avg} and H/C_{avg} of this molecular group being 0.70 ± 0.3 and 1.64 ± 0.2 , both of which are much higher than average elemental ratios for CHO and CHNO groups (Table 4.1). Willoughby et al.¹⁵⁴ attributed aliphatic, oxygenated CHOS species in aerosol to the co-emission and reactions of sulphate (SO_x) aerosol with carbonaceous, hydrogen-saturated aerosol. This could also be the reason behind CHOS molecular group having a much higher percent fraction of SVOC and LVOC compounds, in comparison to CHNO and

CHO groups, where ELVOC compounds were prevalent (Figure S4.5). It must be noted here that 15-T FT-ICR MS detected almost twice the number of CHOS species ($n = 355$) with the same overall elemental ratios, but relatively higher unsaturation ($\text{DBE}_{\text{avg}} = 4.0 \pm 2.6$ and $\text{Al}_{\text{mod,avg}} = 0.12 \pm 0.6$). These CHOS species prevailed from m/z 200-450 (Figure S4.7). In both FT-ICR instruments, low levels of unsaturation, O/S of ≥ 4 , and a minimum of five O atoms per CHOS molecule implies S to be present as organic sulfonates ($-\text{SO}_3$) or sulphates ($-\text{OSO}_3$) with other O-containing functional groups. The neutral loss of these functional groups as SO_3 , SO_4 , and H_2SO_4 has been reported in the literature^{59, 161}.

4.3.5 CHOP and CHNOP species

The most prominent improvement in the information obtained from the 21-T FT-ICR MS – at much lower relative intensities – was P-containing assignments that fell into two molecular groups: CHOP ($n = 488$) and CHNOP ($n = 244$) that comprised 4.6 and 2.0% of all formulae detected by this instrument. The presence of P in aerosol has been explored in some studies^{154, 162, 163}, but their molecular characterization is essentially missing. Along with N, P is an important nutrient that may enter aquatic ecosystems via wet and dry atmospheric deposition. Despite the diffused flux of atmospherically deposited P, it can alter the biogeochemical balance of the receiving aquatic ecosystem with cascading effects on downstream water quality¹⁶³⁻¹⁶⁵. Excess deposition of P in aquatic systems, particularly in the euphotic zone, causes undesirable effects, such as eutrophication, that may lead to increase in primary productivity and O depletion¹⁶³. Biomass burning is among the major sources of P into the atmosphere^{162, 163, 166}. Previous studies have reported an increase in P-content in aerosol in the vicinity of anthropogenic coal burning or large-scale forest and peat fires^{154, 163}, especially on haze days caused in the aftermath of these events. Considering the environmental relevance, it is crucial to detect and identify P-containing species in BBOA as their molecular nature can inform studies on their deposition potential and influence on the recipient ecosystem.

Elemental analysis using energy-dispersive X-ray spectroscopy (EDX; data not shown) indicated the BBOA mixture analyzed here to contain a non-negligible fraction by atomic percentage of P. However, P-containing species were detected neither with the FT-OTE nor the 15-T FT-ICR MS. According to 21-T FT-ICR MS, two P-containing molecular groups, CHOP and CHNOP, were present and exhibited high oxygenation ($\text{O}/\text{C}_{\text{avg}} = 0.84 \pm 0.2$ and 1.02 ± 0.2) and a very aliphatic nature ($\text{DBE}_{\text{avg}} = 3.66 \pm 2.2$ and 3.74 ± 1.9). The formulae belonging to

these molecular groups were distributed across O_6P_p and $N_nO_6P_p$ classes ($O_7P-O_{18}P$ and $NO_9P-NO_{17}P$). The O/P values were ≥ 4 and O/P_{avg} was 12.80 ± 2.0 in CHOP suggesting these species to contain phosphate ($-OPO_3$) with other O-based functional groups. The values of $O/(P+N)$ were ≥ 3 with $[O/(P+N)]_{avg}$ of 6.64 ± 0.83 , implying the presence of organic species with both $-OPO_3$, $-NO_2$, and/or $-ONO_2$, with other O-based functionalities. Considering that these species were located in the van Krevelen space associated with bioavailable carbohydrate, lipid, or protein-like compounds (Figure 4.6), it is plausible that these could be long-chain phosphoalkane or phosphoalkene compounds. Estimations for physical parameters discussed for other molecular groups, including C^0 , T_g , and PSR, have not yet been developed for CHOP(N) compounds, and thus, are not included in this report.

4.3.6 Comparison of molecular composition delineated from 15-T and 21-T FT-ICR instruments

Altogether, 11,474 distinct molecular components were identified by the 15-T and 21-T FT-ICR instruments (Datasets S4.1 and S4.2). Of these, 5112 were commonly detected. The 21-T FT-ICR MS took the lead by exclusively detecting 5421 species as compared to only 941 for the 15-T FT-ICR MS. The spectra acquired from the instruments exhibited clear differences in mass and intensity distributions (Figures 4.1 and S4.6). Notably, more peaks can be found at higher mass defects in the spectrum from 21-T FT-ICR MS throughout the scan range and especially at higher m/z values (Figures 4.4 and 4.5). These peaks may correspond to either H-saturated or O-deficient species. Similar H/C_{avg} of species exclusively detected by either instrument with much lower O/C_{avg} of 0.41 ± 0.2 for the 21-T FT-ICR MS as compared to 0.82 ± 0.3 for the 15-T FT-ICR MS suggests the latter case to be prevalent in the selectivity of the 21-T FT-ICR MS. This is supported by the visualization of species exclusively detected by each instrument in the van Krevelen space, as well as DBE versus C plots, where they occupy distinct and complementary spaces (Figure 4.6). The species detected only by the 21-T FT-ICR MS had H/C_{avg} and DBE_{avg} of 1.31 ± 0.3 and 10.64 ± 5.4 , respectively, as compared to 1.40 ± 0.4 and 5.45 ± 3.3 for those detected by the 15-T FT-ICR MS. The latter had an m/z_{avg} of 376.345 ± 110.98 and comprised predominantly of H-saturated, oxygenated, aliphatic N-containing species ($n = 633$, $H/C_{avg} = 1.51 \pm 0.4$, $O/C_{avg} = 0.87 \pm 0.3$, $DBE_{avg} = 5.17 \pm 3.1$) and S-containing species ($n = 191$, $H/C_{avg} = 1.34 \pm 0.4$, $O/C_{avg} = 0.76 \pm 0.3$, $DBE_{avg} = 4.95 \pm 2.9$). Looking at Figure S4.6, we cannot comment on the

exclusivity of the few peaks between m/z 100-150 detected only in the 15-T FT-ICR MS as the 21-T FT-ICR MS scan range was set to be $m/z \geq 150$.

Much of the performance enhancement of the 21-T FT-ICR MS did not appear to be because of its superior resolving power; it was more attributable to its better sensitivity toward heterogenous species that were present in relatively small quantities in the BBOA mixture analyzed here. To some extent, better detection of such species may likely be due to much higher number of transients co-added from the 21-T FT-ICR MS (1000 as compared to 300 from the 15-T FT-ICR MS) in an attempt to optimize data acquisition. However, this was not the major decisive factor as there was a substantial number of species that were exclusively detected with strong peak intensities only in the 15-T FT-ICR MS (Figures 4.6 and S4.6). These notably included N and S-containing species, along with a large number of unidentified species throughout the scan range (refer to Sections 4.3.7). As commented in the following section, the differences, and specifically improvement, in compositional complexity of our BBOA mixture, when seen using the 15-T and 21-T FT-ICR instruments is due to the difference in instrumental designs downstream of the ionization source and data processing routines (absorption mode processing presents better resolution), as well as the selectivity of ESI sources used on each MS for this experimentation toward certain fractions of the BBOA mixture.

Figure 4.4 Reconstructed mass spectra depicting isobaric species at m/z 571: (top) 15-T FT-ICR MS and (bottom) 21-T FT-ICR MS. Delta mass of a few adjacent peaks detected exclusively by the latter are depicted to emphasize that these species were resolvable by it but remained undetected. Gray peaks in the background denote peaks that were either contaminants or noise. Black peaks represent unidentified species.

4.3.7 Instrumental considerations

The 21-T FT-ICR MS presents resolving powers of 1,200,000 at m/z 400 with absorption mode data processing, while the 15-T FT-ICR MS and FT-OTE MS have resolving powers of 1,000,000 and 240,000 at m/z 400. Resolving power and rate of acquiring spectra of an FT-ICR MS is linearly proportional to its magnetic field strength, while mass accuracy and dynamic range are quadratically related to it⁸⁷. The apparent improvements or insufficiencies in performances of the three mass spectrometers manifested as different molecular components being detected by the instruments and varying intensities of components detected by all. These differences could be due to multiple factors. Particularly, the enhanced sensitivity and dynamic range of 21-T FT-ICR MS that extended its analytical capabilities to identification of compounds present in smaller quantities in the BBOA mixture, notably P-containing compounds, was not solely because of greater resolving power. For instance, resolution of $C_{23}H_{41}O_{14}P$ from isobaric species, $C_{30}H_{36}O_{11}$ and $C_{25}H_{36}N_2O_{13}$, at m/z 571 by the 21-T FT-ICR MS required resolving powers of $\sim 239,000$ and $\sim 342,000$, respectively (Figure 4.4). The resolution of $C_{26}H_{36}O_{14}$ from $C_{28}^{13}CH_{33}NO_{11}$ at the same nominal mass required resolving power of 308,000. The resolution of $C_{14}H_{27}O_{15}P$ from $C_{23}^{13}CH_{19}NO_9$ required resolving power of 810,000 at m/z 465 (spectrum not shown). All these ion pairs were resolvable by the 15-T FT-ICR MS, but they remained undetected by it (Figure 4.4). Furthermore, stark differences in ion intensities were recorded. The seemingly better performance of the 21-T FT-ICR MS was owed not only to a much larger number of scans co-added to acquire its mass list, but also to the fact that stronger magnetic strength allows ion clouds to thrive in the ICR cell for longer time periods before coalescing¹³³.

A better understanding of differences in mass and intensity distributions between the 15-T and 21-T FT-ICR instruments can be obtained by visualizing molecular formulae in terms of the sum of normalized intensity, which serves as a proxy of total ion current here. The summed normalized intensity that is accounted for by formulae containing each integer value of C, O, and DBE is shown in Figure S4.4. Ion current was distributed across species with C_{4-36} and C_{4-41} in the 15-T and 21-T FT-ICR instruments, with C_{avg} of 19.62 ± 6.6 and 23.09 ± 7.3 . The 21-T FT-ICR MS exhibited a prominent step-up in ion current from species with ≤ 20 C atoms, where species with C_{20} accounted for the most ion current. Some of the most intense C_{20} species identified by 21-T FT-ICR MS analysis are presented in Table S4.1 with their corresponding normalized intensities in 15-T FT-ICR and FT-OTE MS⁴¹ analyses. The species detected by both instruments had very dissimilar intensities in the mass spectra, while some

C₂₀ species were entirely missed in detection by the 15-T FT-ICR MS (Figure S4.8). For instance, C₂₀H₂₆O₃ formed the tallest peak in 21-T FT-ICR but it occurred as a very diminutive peak in 15-T FT-ICR and FT-OTE instruments (Table S4.1). The C₂₀ species comprised both heteroatom-containing and CHO-only molecular groups. The same was observed for a molecular class containing O₇ within the CHO group only. Species belonging to this class contributed to the largest fraction of ion current in both FT-ICR instruments, but with marked differences in species detected and their mass and intensity distributions (Figures 4.5 and S4.9). As an example, three nominal masses from Figure S4.10 are reconstructed in Figure 4.5. At *m/z* 305, 401, and 505, the 21-T FT-ICR MS detected 2, 3, and 3 species with seven O atoms; however, the 15-T FT-ICR MS detected only 1, 3, and 0 with much smaller relative intensities.

Figure 4.5 Reconstructed mass spectra depicting isobaric species at (A) m/z 305, (B) 401, and (C) 505. In each panel, the top spectrum was acquired from 15-T FT-ICR MS and the bottom spectrum from 21-T FT-ICR MS. Species belonging to O_7 class are pinpointed with their neutral formulae. Formulae outlined in green were detected by both instruments, while those in red were detected only by the 21-T FT-ICR MS. Gray peaks in the background denote peaks that were either contaminants or noise. Black peaks represent unidentified species.

On the one hand, components of complex mixtures that were seemingly not detected in the 15-T FT-ICR MS analysis of this BBOA mixture were detected by the 21-T FT-ICR MS. On the other hand, we observed species that were very highly enriched in the 15-T FT-ICR MS, but suppressed in the 21-T FT-ICR MS. A synthetic linear alkyl benzene sulfonate⁷⁴ contaminant species, $C_{17}H_{28}O_3S$, presents an excellent example. This sulfonate was the most intense peak in the mass spectrum of 15-T FT-ICR MS that was approximately twice the height of the tallest analyte (non-contaminant) peak with a credible formula assignment (Figure S4.2). Additionally, approximately 2400 unassigned peaks (10.32% of the raw mass list) that accounted for 15% of the cumulative ion intensity were also enriched in the 15-T FT-ICR mass spectrum. The ions forming these peaks were either not detected by the 21-T FT-ICR MS (m/z 571 in Figure 4.4) or had very low intensity (m/z 401 in Figure 4.5). Most such peaks that were enriched in 15-T FT-ICR MS analysis could be organized in CH_2 -homologous series using Kendrick mass defect but could not be assigned a molecular formula with the routine used here. We could also not satisfactorily prove the presence of non-traditional elements/adducts (F, Cl, Na, etc.), odd-electrons, multiple charge, etc. Considering that these species were observed only in BBOA samples, such as the one reported here and others to be reported elsewhere in the future, but not in other natural complex mixtures, including Suwannee River fulvic acid and Maple leaf organic matter extracts, during the same experimentation period with the 15-T FT-ICR MS, we speculate that they are likely species unique to our BBOA mixtures. As most instrumental settings and even data acquisition protocols (such as the use of aFT processing) should have caused an overall better sensitivity of the 21-T FT-ICR MS, we speculate that these outlier peaks with high intensity in the 15-T FT-ICR MS may be due to the tendency of selective ionization by the ESI source on the 15-T FT-ICR MS.

Figure 4.6 Molecular species identified exclusively or commonly by 15-T and 21-T FT-ICR mass spectrometers. The upper panel shows the distribution of such species in the van Krevelen space, while the bottom panel shows them in the DBE versus C plot. Gray data points in the background represent all species detected by the mass spectrometer depicted in each diagram (third column has both). Exclusive or common species are overlaid in colour, where the colours represent molecular groups. Data points are scaled to abundance in the first two panels.

4.4 Acknowledgments

We would like to thank Dr Matthew Brege and Dr Simeon Schum for constructive discussions and advice in data processing with R and MFAssignR. This work was supported in part through a Fulbright Foreign Student Award to AI and a user project with Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA awarded to LRM.

4.5 Data availability

Datasets S4.1 and 4.2 with all molecular formulae assigned to the BBOA mixture analyzed here using 15-T and 21-T FT-ICR MS analysis.

4.6 Supplementary material

4.6.1 Bibliometric analysis

A bibliometric mini-analysis was performed using four search terms on Web of Science: “21-T FT-ICR”, “21-T FT-ICR”, “21 T FTICR”, and “21-T FTICR” that accommodated the various possible hyphenation options. The same were used as search terms with the acronyms spelled out. We looked for these terms in the title, abstract, author keywords, and keywords plus published papers. The time span was set to 1965-2020 and Web of Science Core Collection was selected as the database. The search resulted in a total of 24 distinct publications starting from 2006.

4.6.2 Calculation of estimable physical parameters

Modified aromaticity index (AI_{mod}) was calculated as described by Koch and Dittmar, 2006 (See also the erratum published in 2016)^{99, 100}: $AI_{mod} = (1 + C - (0.5 \times O) - S - (0.5 \times H) - (0.5 \times N) - (0.5 \times P)) / (C - (0.5 \times O) - S - N - P)$, where C, O, S, H, and N are the numbers of atoms of these elements in each molecular formula. Based on AI_{mod} , the formulae were then distributed in aromaticity categories as follows: $AI_{mod} \leq 0$ were categorized as aliphatic; $0 < AI_{mod} \leq 0.5$ and $O/C < 0.5$ were classified as unsaturated with low O; $0 < AI_{mod} \leq 0.5$ and $O/C > 0.5$ were classified as unsaturated with high O; $0.5 < AI_{mod} < 0.67$ were classified as aromatic, and $AI_{mod} \geq 0.67$ were classified as condensed aromatic.

Average oxidation state of C (OS_c) atoms was calculated as described by Kroll et al., 2011 but with a modified equation⁹⁴: $OS_c = 2 \times (O/C) - (H/C) - 5 \times (N/C) -$

$6 \times (S/C) - 5(P/C)$. Here, C, O, S, H, and N are the numbers of atoms of these elements in the molecular formulae.

Estimated saturation mass concentrations (C^0) were calculated using the equation described by Li et al., 2016⁹⁵: $\log_{10}C^0 = (nOC - C) \times bC - O \times bO - H \times bH - 2 \times ((C \times O)/(C + O)) \times bCO - N \times bN - S \times bS$. In this equation, C, O, S, H, and N are the numbers of atoms of these elements in each molecular formula.

Furthermore, nOC is the reference C number; bC , bO , bN , and bS denote the contribution of each atom to the estimated C^0 . bCO is the C-O non-ideality. The values of these constants are different based on molecular groups to which a given formula belongs (CHO, CHNO, CHOS, CH, CHNOS, or CHN) and were accessed from the report published by Li et al., 2016. Formulae with $\log_{10}C^0 \leq -3.52$ were categorized as extremely low volatility (ELVOC), $-3.52 < \log_{10}C^0 \leq -0.52$ as low volatility (LVOC), $-0.52 < \log_{10}C^0 \leq 2.47$ as semi-volatile (SVOC), $2.47 < \log_{10}C^0 \leq 6.47$ as intermediate volatility (IVOC), and those with $6.47 < \log_{10}C^0$ as volatile organic compounds (VOC).

Dry glass transition temperatures ($T_{g,dry}$) were calculated from molecular formulae using the equation described by DiRieux et al., 2018: $T_{g,dry} = nOC + \log(l) \times bC + \log(H) \times bH + \log(l) \times \log(H) \times bCH + \log(O) \times bO + \log(l) \times \log(O) \times bCO$. Here, nOC is the reference C number, bC , bH , and bO represent the numeric contribution of these elements to $T_{g,dry}$. Also, bCH and bCO are coefficients that reflect contributions from C-H and C-O bonds, respectively. This equation is applicable to CHO species only. The values of these constants were accessed from the report published by DiRieux et al., 2018.

Relative humidity-dependent glass transition temperatures ($T_{g,RH}$) were calculated as described by Schum et al., 2018⁷³ and Brege et al., 2024⁴¹, although this was originally explained in Shiraiwa et al., 2017⁹⁷. This was calculated using $T_{g,RH} = (1 - Worg) \times T_{g,w} + (1/Kgt) \times Worg \times T_{g,dry} / ((1 - Worg) + (1/Kgt) \times Worg)$. Here, $T_{g,w}$ is the glass transition temperature of water, i.e., 136 K, and Kgt is the Gordon-Taylor constant that is equal to 2.5. Furthermore, $Worg$ was calculated using the following equation: $Worg = 1.4 - ((1.4 \times RH)/100) / (1.4 - ((1.28 \times RH)/100))$. Here, RH is the measured relative humidity on the day of sampling in Richland, WA. Phase states ratios (PSR) were then calculated using the following: $PSR = T_{g,RH}/T_{amb}$, where T_{amb} is the average ambient temperature on the day of sampling in Richland, WA. According to meteorological data, RH and T_{amb} on Sep 05, 2017 were 30.5% and 300.15 K, respectively. Molecular formulae with $PSR \geq 1$ were classified as solids, those with $1 > PSR > 0.8$ were classified as semi-solids, and those with $PSR \leq 0.8$ were classified as liquids.

4.6.3 Supplementary tables

Table S4.1 Formulae assigned to top-10 tallest peaks detected by the 21-T FT-ICR MS and their normalized intensities recorded by both FT-ICR and the FT-OTE instruments. Intensities were normalized to the intensity of the tallest peak for the former instruments, while they were normalized to the sum of all intensities by Brege et al.⁴¹ for the latter.

Molecular formula	Normalized abundance in 21-T FT-ICR MS	Normalized abundance in 15-T FT-ICR MS	Normalized abundance in FT-OTE MS ⁴¹
C ₂₀ H ₂₆ O ₃	1.000000000	0.007121367	0.010236817
C ₂₂ H ₄₄ O ₂	0.957040573	0.010810526	ND
C ₂₀ H ₂₈ O ₃	0.877486078	0.007266178	0.006221464
C ₂₀ H ₂₆ O ₄	0.858392999	0.013913374	0.008012584
C ₂₄ H ₄₈ O ₂	0.790373906	0.005693337	ND
C ₂₀ H ₂₆ O ₅	0.719570406	0.025636379	0.005584254
C ₁₈ H ₁₆ O ₇	0.707637232	0.122125957	0.012497067
C ₂₀ H ₃₀ O ₅	0.678202069	0.017860613	0.004568523
C ₁₇ H ₁₆ O ₆	0.673826571	0.102051839	0.020376228
C ₂₀ H ₂₈ O ₅	0.663484487	0.015154205	0.004626966

ND, not detected.

4.6.4 Supplementary figures

Figure S4.1 Trends in research studies that employed 21-T FT-ICR MS: As of 9th Nov 2020, Web of Science indexes 24 publications that report the use of a 21-T FT-ICR MS, beginning in 2006-2008 with reports on conceptual design of a 21-T

FT-ICR system and supporting equipment¹⁶⁷⁻¹⁷⁰. Initial performance evaluations appeared in 2015¹³⁴ and 2016⁸⁸ from the National High Magnetic Field Laboratory, Tallahassee, FL and the Pacific Northwest National Laboratory, Richland, WA, respectively. Bubble graph prepared with VOSviewer¹⁷¹, where the bubbles show the major keywords and the lines represent the relationship of different publications via these keywords.

Figure S4.2 Reconstructed mass spectra from 15-T and 21-T FT-ICR mass spectrometers depicting peaks that were assigned a molecular formula in green and unassigned peaks in red. The tallest peak in spectrum from the 15-T FT-ICR MS was a sulfonate contaminant species. As the entire dataset was normalized to its intensity, most analyte peaks resided at normalized intensity of <0.3. Gray peaks going up to higher m/z are noise peaks. Insets show the cumulative ion current (in terms of sum of normalized abundance) that was accounted for by assigned and unassigned peaks.

Figure S4.3 Error distribution of molecular formulae assigned to species detected by FT-OTE, 15-T FT-ICR, and 21-T FT-ICR mass spectrometers. Zoomed-in error distributions of both FT-ICR instruments are presented in the inset.

Figure S4.4 Distribution of total ion current, in terms of sum of normalized abundance, that was accounted for by molecular formulae with each integer value of: (A) Number of C atoms, (B) Number of O atoms, and (C) Double bond equivalence. Bars outlined with black represent 21-T FT-ICR MS, while the ones

outlined in gray represent 15-T FT-ICR MS. Bars are filled to represent fraction of formulae in each molecular group.

Figure S4.5 Distribution of S-containing species detected by the 21-T FT-ICR MS across O₀S_s molecular classes and their: (A) Aromaticity trends and (B) Volatility trends.

Figure S4.6 Reconstructed mass spectra from 15-T (top) and 21-T (bottom) FT-ICR mass spectrometers. Species exclusively identified by each instrument above the signal-to-noise ratio are shown as coloured peaks overlaid on commonly identified species (gray peaks).

Figure S4.7 Distribution of P-containing species detected by the 21-T FT-ICR MS across O_nP_p and $N_nO_oP_p$ molecular classes and corresponding aromaticity trends: (A) CHOP species, and (B) CHNOP species.

Figure S4.8 Reconstructed mass spectra from 15-T (top) and 21-T (bottom) FT-ICR mass spectrometers. Species identified to have C₂₀ are shown in either black (commonly identified) or by colours of molecular groups (exclusively identified by each instrument). The rest of the data is drawn in gray. Insets on the right show C₂₀ species overlaid on the rest of the formula assignments (gray) for either instrument (right, 15-T; left, 21-T). Data points are sized by normalized intensity in the insets.

Figure S4.9 Reconstructed mass spectra from 15-T (top) and 21-T (bottom) FT-ICR mass spectrometers. Species identified to have O7 are shown in colour overlaid on the rest of the molecular formulae. Insets on the right show O7 species overlaid on the rest of the data (gray) for both instruments (right, 15-T; left, 21-T). Data points are sized by normalized intensities in the insets.

5 Molecular Insights into the Anatomy of Ambient Tar Balls in Wildfire Smoke from the Pacific Northwest

Abstract

Tar balls (TB) are a major carbonaceous product of wildfires and other large-scale biomass burning events that often exceeds soot in number and mass, especially in smouldering phase. They are an active scientific interest owing to their highly uncertain light-absorbing/scattering characteristics that allow them to influence air quality, radiation budget, and climate. Being a relatively recalcitrant fraction of organic carbon in wildfire smoke, TB are capable of long-range atmospheric transport, and thus, exert influence not only in the vicinity of wildfires but also in remote regions. A molecular signature specific to TB-rich, atmospherically aged smoke plume collected in the Pacific Northwest was delineated here in comparison to relatively fresh, non-TB smoke using 15-Tesla Fourier transform – ion cyclotron resonance mass spectrometry (FT-ICR MS). Molecular composition from complementary ionisation methods, laser desorption and electrospray ionisations, demonstrated spheroid TB to comprise a less-oxygenated amorphous core and a more oxygenated, solidified exterior. Large number of species exclusive to TB-rich aerosol relative to non-TB aerosol are proof that molecular composition plays a key role in dictating TB formation during maturation of smoke plumes. Tar ball-specific molecular signatures are a critical resource that must be considered in atmospheric and climate models to account for this persistent form of organic carbon not only in the vicinity of wildfires but in far-off communities and ecosystems that are ill-prepared to manage environmental and health challenges associated with wildfire emissions.

5.1 Introduction

Wildfires have immense destructive power that not only threatens animal and human lives, but also substantially deteriorates air quality¹¹. In the last 10 years (2011-2020), ~62,693 wildfires occurred annually in the United States, each of which burned nearly 7.5 million acres¹³. Although wildfires are more frequent in Eastern and Central United States, only a small acreage is impacted; Western United States suffers from less recurrent, but larger and longer wildfire episodes

^{13, 14}, making them a major determinant of atmospheric organic aerosol (OA) concentrations, especially during summertime ¹⁵. Since the 1980s, western wildfires have become 4-fold frequent and 64% longer in duration from discovery to control, burning 6.5 times more acreage annually ¹⁴. An increased occurrence of wildfires is predicted for the future, owing to a warmer, drier climate, and changes in fuel availability ^{4, 14}.

Extreme wildfire events across the globe occasionally generate pyrocumulonimbus (pyroCb) clouds that inject smoke into the stratosphere ¹⁷²⁻¹⁷⁵. Smoke from Pacific Northwest wildfires in August 2017 led to an unusually strong pyroCb occurrence, where the ascent of smoke plume was associated with black carbon (BC; 2% by carbonaceous aerosol mass; 98% was OC) that absorbed heat energy and allowed self-lofting of the plume. Smoke from this event persisted in the stratosphere for 8 months, which was only 40% shorter lifetime than that simulated under conditions of no OC in the plume ¹⁷⁶. These findings countered the notion of short lifetimes and photochemical lability associated with smoke OC, highlighting the need to understand persistent OC fractions likely to reach far-off communities and ecosystems that are ill-prepared to manage environmental and health challenges associated with wildfire emissions.

The goal of this study was to examine molecular composition in wildfire smoke and to understand if it plays a key role in enabling smoke to persist even in the warmer, more humid conditions of the boundary layer, specifically by forming tar balls (TB) that are a relatively recalcitrant product of atmospheric aging of biomass-burning smoke plumes. Tar balls are carbonaceous, highly viscous, spherical entities with diameters of <500 nm; they are devoid of any crystalline microstructure and are resistant to high-energy radiation ^{16, 17, 34}. They form within minutes to hours of emission from wildfires ³⁹ and may constitute a large portion (even >95%) of particulate mass in the vicinity of large-scale biomass burning events ^{24, 40, 41}. They have been reported in extremely remote areas, such as the Himalayas, after long-range transport ¹⁷⁷. While suspended as aerosol, TB directly exert influence over air quality, radiation budget, and climate due to their highly variable optical characteristics ^{16, 37}, while their hygroscopic nature has been associated with indirect effects on climate, such as via their potential to function as cloud condensation nuclei (CCN)³⁸.

Smoke plumes from Pacific Northwest fires in August 2018 were collected at Benton County, WA, United States on the days of intense burn episodes and after they had lingered on-site for a few days to target smoke of varying ages that showed in the distinct evolution of liquid-like particles to TB from the relatively fresh to aged plumes (Figure 5.1). An unequivocal molecular signature specific to TB was isolated here from ultrahigh-resolution mass spectrometric analysis that provides valuable insight into their structure, phase state, and optical properties. More importantly, this molecular signature is a critical resource that must be integrated in atmospheric and climate models to improve the representation of OC as a persistent component of wildfire smoke, and thus, better predict transport, transformation, and fate of particulate emissions, as well as their climatic influence regionally and via long-range transport.

5.2 Methods

5.2.1 Sample collection and preparation

Sampling was conducted at the Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, in Richland, WA, USA. Three fresh aerosol samples that had arrived on site in <12 hours post-emission were collected on Aug 09 (day and night) and 11 (Figures 1 and S1). Aged aerosol (>12 hours post-emission) was collected on Aug 13, when smoke had lingered on site for a few days (Table S5.1). Aerosol has intercepted larger fires in Boreal forests of British Columbia, Canada and smaller fires in Washington, US. Average ambient temperature and relative humidity varied across sampling days from 23-32°C and 22-32% (Table S5.1).

5.2.2 High-resolution mass spectrometry

Sample were analysed with 15-Tesla Fourier transform – ion cyclotron resonance mass spectrometry using two complementary techniques in negative-ion mode: laser desorption ionisation on aerosol deposited on substrate and extracted aerosol and electrospray ionisation (Supplementary Information). Molecular

formulae were assigned to measured masses using MFAssignR⁹³ within the following constraints: $C_cH_hO_oN_{0-3}S_{0-1}^{13}C_{0-2}^{34}S_{0-1}$; $-13 \leq DBE-O \leq 20$; $0 \leq O/C \leq 2.0$; $0.3 \leq H/C \leq 2.5$, and maximum permissible error of 0.5 ppm. Structural information was inferred from assigned formulae in terms of double bond equivalence (DBE) and modified aromaticity index (AI_{mod})^{99, 100}. Physicochemical parameters, including saturation mass concentration of pure compounds (C_0) and volatility classification^{30, 95}, and dry⁹⁶ and relative humidity-dependent⁹⁷ glass transition temperatures ($T_{g,dry}$ and $T_{g,RH}$) were estimated as described in their respective reports (Table S5.2).

5.3 Results and Discussion

5.3.1 Tar balls are more abundant in aged wildfire plumes

Aerosol samples in this study were collected downwind of numerous wildfires in the Pacific Northwest in 2018. The plumes had likely undergone varying extent of oxidative aging *en route*. Backward trajectories show smoke plumes sampled on Aug 09 (daytime and nighttime) and Aug 10 to have travelled <12 hours from heavy wildfires before reaching the sampling site (Figures 5.1 and S5.1). August 13 sample was influenced by smoke that had been lingering at site for a few days, as well as smoke plume that reached here after travelling >12 hours. Samples codes used henceforth are given in Table S5.1. Age of a plume is directly proportional to the number of TB^{24, 39}. Here, TB formed 9 and 11% of particles in BBAug09(DT) and BBAug1011, as compared to 56% in BBAug1314. Furthermore, shape of particles on the substrate was substantially different between the two sample types, where BBAug09(DT) and BBAug1011 turned to flattened domes on impact (Figures 5.1A-C). The shape of particulate matter upon impact with the substrate is dictated by their viscosity^{178, 179}, where freshly emitted OA is prone to spreading due to low viscosity. Individual particles in BBAug1314 showed a stark difference in shape from non-TB samples by holding distinct spherical forms following impact with the substrate, which is a characteristic of TB (Figure 5.1C). This property is attributable to increase in viscosity and surface tension as OA ages, confirming the matured stage of this smoke plume with respect to the others.

Transmission electron microscopy (TEM) showed TB in Aug 13 sample to have average diameters of $\sim 0.4 \mu\text{m}$ (400 nm) (Figure 5.1D), which is consistent with earlier studies that have reported a range of diameters from 30-1500 nm, with an average size of $<500 \text{ nm}$ ^{16, 24, 35, 36}. An overwhelming majority of particles were identified to be carbonaceous from computer-controlled scanning electron microscopy (CCSEM) with sizes from $0.05 \mu\text{m}$ to $0.65 \mu\text{m}$ (Figure S5.2), indicating a heavy influence of biomass burning on all three samples. Dust, sea salt/sulphates, and Na-rich particles are prevalent in BBAug1011 that is likely because of its path from across the Pacific Ocean before intercepting wildfires in Washington State, US (Figures 5.1 and S5.1). Tar balls in BBAug1314 had no inorganic inclusions (IC), which is the typical mixing state observed by others as well ^{24, 36}. Minor elemental carbon inclusions were observed in some TB; 72% of single particles analysed by scanning transmission electron microscopy (STXM) and near-edge X-ray absorption fine structure spectroscopy (STXM/NEXAFS) in BBAug1314 were pure OC (Figure S5.3A). Based on aspect ratios and optical depth, majority of single particles were classified as solids or semi-solids with average total carbon absorption (TCA) of 0.22 as compared to 0.18 and 0.11 for single particles largely classified as semi-solids or liquids in BBAug09(DT) and BBAug1011 (Figure S5.3A).

5.3.2 An oxygenated exterior with a less oxygenated amorphous core

Early studies speculated TB to arise from gas-to-particle condensation, OH $\bullet$ -induced polymerisation of low-volatility species with simultaneous loss of water by evaporation in biomass burning OA ^{16, 35}, and heat shock-induced solidification of tar droplets that are directly emitted from pores of burning biomass ¹⁸⁰. Sedlacek III et al. ³⁹, as well as other researchers²⁴, classified TB as ‘processed primary aerosol’ rather than SOA. They reasoned that a requisite for classification of aged aerosol as SOA is gas-to-particle condensation that should manifest as an increase in size of TB with atmospheric age. This was unequivocally not the case in their (and others ^{16, 17}) study: TB increased in number as distance increased from the site of emission, and their average diameter was independent of the photochemical age. Furthermore, a layer of different composition, specifically higher oxygenation and less volatility, has been proposed to exist on the surface of TB ^{36, 40}. Thickness of 30 nm ³⁸ and 41 nm ³⁶ has been deduced for the oxygenated layer by electron energy loss spectrometry (EELS) and STXM/NEXAFS analyses, respectively, where increased O content was associated with enhanced concentration of carboxylic C. The

exterior is shielded and does not undergo extensive chemical transformation, but the outer layer of a TB is exposed, which manifests as enhanced oxidation and associated physical transformation, i.e., lowered viscosity and volatility ^{24, 39}).

Figure 5.2. A comparison of TB-rich (BBAug1314) and TB-poor (BBAug1011) aerosol and ionisation methods: Van Krevelen diagrams in the top panel depict the ionization selectivity of ESI and LDI on the two aerosol mixtures. Species ionized by both methods are drawn in grey in the background and are overlaid by peaks that were exclusively ionized by either method (blue, ESI only; maroon, LDI only). The two bottom panels depict the species exclusively present in one aerosol mixture and not the other

Electron energy loss spectrometry on BBAug1314 showed carbon to be uniformly distributed in TB (Figure 5.1). Oxygen, on the other hand, was localised toward the outer edge of TB, which is consistent with the current known facts about TB morphology. The presence of an oxygenated layer and its potential molecular composition was demonstrated here by subjecting aerosol-loaded filters to HRMS in conjunction with LDI imaging and ESI, both of which are soft-ionisation techniques capable of ionising intact molecules⁶⁷. Laser shots of 50-100 μm diameter fired randomly on the substrate were not expected to exclusively ionise TB, but the probability of hitting and ionising a TB was stronger on substrates largely covered with TB as in BBAug1314. Electrospray ionisation on ACN extracts requires extractive procedures (sonication and

dissolution in our study) that would disintegrate TB and combine species from the core and outer layer. It must be noted that LDI has a greater propensity to ionise species of a more condensed, aromatic character, with less oxygenation⁶⁰, while ESI is selective for more oxygenated species in complex mixtures. We hypothesised that in a TB-rich sample, LDI should largely have access to oxygenated outer layers of TB, so the overlap between the two ionisation methods should increase with the increase in percent fraction of TB. Alignment of species identified by LDI in TB with non-TB samples, should reveal the former to have a more oxygenated character, while ESI should depict TB-rich aerosol to be enriched in fresher, less oxygenated species of the amorphous core.

Figure 5.2 shows the percent fraction of species commonly identified by ESI and LDI to increase from 20.6% for BBAug0910(DT) to 45.7% for BBAug1314 with most species above $O/C > 1.0$ being commonly ionised by LDI and ESI in the latter. Furthermore, molecules exclusively present in BBAug1314 with respect to non-TB samples occupy distinct zones in van Krevelen diagrams drawn from the two ionisation methods. In confirmation of our hypothesis, TB-specific species occupy a low- O/C region in ESI and a very high O/C region in LDI (Figure 5.2; Table 1). In LDI, the species exclusive to BBAug1314 had $O/C_{wa} = 0.72 \pm 0.35$, in comparison to $O/C_{wa} = 0.36 \pm 0.17$ for BBAug1011. In ESI, O/C_{wa} for BBAug1314 was computed to be 0.54 ± 0.27 as compared to $O/C_{wa} = 0.86 \pm 0.43$ for BBAug1011. The large overlap between ESI and LDI imaging provides a great confirmation of the latter as a promising technique for analysis of aerosol-loaded filters that has not been extensively employed for the purpose; however, for further confirmation, we determined the similarity of LDI imaging on a non-TB and TB sample each with molecular composition delineated from LDI on ACN extracts. Overall, >80% of assigned formulae from LDI imaging were shared with one or both ionisation methods (Figure S5.4).

Figure 5.3 A comparison of TB-rich (BBAug1314) and non-TB (BBAug1011) aerosol: The panels depict the trends of DBE-(O/2) and RH-dependent glass transition temperatures ($T_{g,RH}$) against the number of C atoms in the molecular fingerprint of each method, i.e., species exclusively present in one aerosol mixture and not the other. These are further grouped and colour-coded by the method(s) which ionized them as in the van Krevelen.

Tar ball formation is a consequence of the loss of volatile species and increase in viscosity^{24, 39}. Volatilities were estimated using elemental compositions from HRMS analysis in terms of saturation mass concentrations (C^0) using the parameterisations of Li et al.⁹⁵, which showed extremely low volatility organic compounds (ELVOC and LVOCs) to be dominant in all samples. More than 75% of identified species in BBAug1314 were E/LVOCs with a minor contribution of semi-volatile and intermediate volatility compounds as compared to TB-poor aerosol (Figure S5.5). No volatile organic compounds were detected in either TB-rich or TB-poor aerosol. In line with TB being solidified entities^{24, 39}, very high average values of $T_{g,dry}$ and $T_{g,RH}$ (>320 K) were noted, which was indicative of a vast majority of species to be semi-solids or solids (data not shown); average

values of $T_{g,RH,wa}$ computed for all aerosols using ESI analysis were comparable, but a substantial difference was noted from LDI analysis (Table 5.1). BBAug1314 had an $T_{g,RH,wa}$ of 348.89 ± 189.93 K as compared to 263.75 ± 244.60 K for BBAug1011. Figure 3 shows the distribution trends of $T_{g,RH}$ for species exclusively detected by either LDI or ESI in only BBAug1011 or BBAug1314. With LDI, the latter shows consistently higher $T_{g,RH}$ compared to BBAug1011 throughout the entire C range providing further evidence of enhanced solidification.

5.3.3 Tar balls are extremely complex at the molecular level

Immense chemical richness was noted for all aerosol mixtures analysed here, but the number of species assigned an unambiguous formula was much higher in TB-rich samples. Unprecedented molecular complexity has been demonstrated for TB-rich aerosol downwind of August 2017 Pacific Northwest wildfires⁴¹; how much of this is attributable to TB specifically is an open question. Even when present as the dominant particle type in aerosol, TB are externally mixed with non-TB organic and inorganic material, where physical separation of TB from this matrix is impossible¹⁸¹. Considering the high similarity generally reported between organic aerosol originating from different biomass burning events^{73, 105, 155}, we assumed here that most – if not all – non-TB molecular species can be accounted for by comparison with multiple non-TB aerosol, which will allow molecular species unique to TB-rich sample to be identified as a TB signature. Formulae assigned to BBAug1314 were aligned with two non-TB samples, BBAug0910(DT) and BBAug1011. Tar balls were expected to be rich in organonitrates and nitroaromatic compounds²⁴, therefore, a nighttime sample, BBAug09(NT), was used to account for non-TB N-containing compounds.

Table S5.3 summarises the number of monoisotopologues identified and the chemical parameters deduced from all assigned formulae. With ESI analysis, 3663, 3686, 3922, and 4841 species were identified in BBAug09(DT), BBAug0910(NT), BBAug1011, and BBAug1314. A good signal could be obtained for BBAug09(DT) with LDI imaging; 1516, 2383, and 4299 species were identified in the other three samples, respectively. Corresponding mass spectra are presented in Figures S5.6 and S5.7 with further details on formula assignment quality in Supplementary Information. Frequency of all possible differences in measured masses of ions (Δm) with ESI analysis were calculated for both full-day aerosol samples, BBAug1011 and BBAug1314. In both

datasets, $\geq 70\%$ of ions were present at the following Δm from another ion(s): 14.01565 (CH_2), 12.00000 (C), 27.99491 (CO), 30.01056 (CH_2O), 42.01056 ($\text{C}_2\text{H}_2\text{O}$), 15.99491 (O), 2.01565 (H_2), 26.01565 (C_2H_2), 44.02621 ($\text{C}_2\text{H}_4\text{O}$), 18.01056 (H_2O).

These mass spectral observations were corroborated by STXM/NEXAFS analysis on single TB and non-TB particles. Absorption peaks were recorded at 285.2, 286.7, 288.5, and 290.3 eV showing the presence of sp^2 , ketonic, carboxylic, and carbonate carbon by (Figure 5.4). These functional groups have commonly been associated with TB^{24, 36, 181} as well as hydroxyl units¹⁸¹. Notably, as shown in pie-charts from STXM/NEXAFS analysis, carboxylate was more prevalent (46%) in TB as compared to non-TB (<25%), while aliphatic character was much weaker in TB. Increase in oxidation state of carbon by formation of carboxylate is a common path toward atmospheric ageing¹⁸². In our samples, the ratio of carboxylate to carbonyl consistently increased from BBAug09(DT) to BBAug1011 (Figure S5.8). Furthermore, frequently occurring Δm of >78 Da in HRMS datasets could be indicative of aromatic rings, but are often much sparse as compared to the aforementioned highly prevalent Δm . Therefore, we focused on them by Kendrick mass defect (KMD) analysis with $\text{C}_6\text{H}_5\text{O}$, $\text{C}_7\text{H}_6\text{O}$, and $\text{C}_6\text{H}_5\text{NO}_3$ as base units. In this analysis, individual peak intensities were normalized in each dataset to the sum of abundance of species that were identified to be a part of a KMD series with C (12.00000) as base unit, which indicated CH_2 , CO_2 , OCH_2 , and H_2O to be evenly present throughout all samples. Notably, as drawn in Figure S9, total abundance of ions that potentially had ≥ 1 aromatic moieties, specifically phenol ($\text{C}_6\text{H}_5\text{O}$) and nitroaromatic ($\text{C}_6\text{H}_5\text{NO}_3$) functionalities, was substantially higher in BBAug1314 as compared to BBAug1011.

Figure 5.4 Absorption spectra from scanning transmission X-ray microscopy analysis of two non-TB aerosol samples, BBAug 09(DT) and BBAug1011, relative to the TB-rich aerosol, BBAug1314. Each spectrum shows the average absorption for particles analysed for the respective sample, and the ribbon around covers the standard deviation. Major absorption peaks correspond to sp^2 , ketonic, carboxylic, and carbonate functionalities

Molecules containing only C, H, and O (CHO species) or those containing N_{0-2} (CHNO species) formed the largest fractions in ESI analysis, with S-containing molecules (CHOS species) constituting a small portion of the aerosol sample set here (Table S5.3). The dominance of N-containing compounds in BBOA, and particularly in TB-rich aerosol, is consistent with previous observations of organic N being comparable to C and O in TB²⁴. Formation of TB is accompanied by the addition of O (possibly as -COOH) and N (possibly as organic N) to low-viscosity organic particles²⁴. In ESI analysis, CHNO species in these aerosol mixtures ranged from NO_3 to NO_{19} and N_2O_5 to N_2O_{16} with only a few additional classes in BBAug1314 and BBAug0910(NT). These CHNO species may have - NO_2 , - NO_3 , or nitrophenolic functionalities, in combination with other O-based functional groups. This inference is merely a speculation based on elemental ratios alone, but is reasonable in view of our KMD analysis (Figure S5.8), as well as fragmentation studies that have reported loss of HNO_3 ^{59, 161}, HNO_2 , and NH_3 ⁵⁹

from BBOA. On the contrary, a smaller number of CHNO species were identified by LDI as compared to CHO species, which may be due to the selectivity of LDI for less polar, aromatic species and/or difference in sample preparation/extraction for the two methods. Additional classes were seen with this ionization method: NO-NO₂ and N₂O₃-N₂O₄, which were not detected in ESI and could be indicative of reduced N or -NO groups. The presence of S and its apparent enhancement in BBAug1314 is attributed to the atmospheric maturity of this aerosol sample. Sulphur has previously been detected in TB in association with processed/aged smoke plumes that had travelled in the atmosphere for 2-3 days after having intercepted wildfires³⁸. Most S species identified in this study was highly aliphatic and oxygenated, which is expected as addition of SO₄²⁻ to organic moieties is a major consequence of atmospheric processing of BBOA^{183, 184}.

Table 5.1. Summary of average chemical characteristics of aerosol mixtures inferred from molecular species exclusively detected by negative electrospray (ESI) and laser desorption/ionization (LDI) with high resolution mass spectrometric analysis. All averages are weighted to normalized abundance. For a summary derived from entire datasets, and not exclusive species only, please refer to Table S3.

Aerosol mixture	Molecular group	MF	O/C _{wa}	H/C _{wa}	DBE _{wa}	Al _{mod}	T _{g,RH} ⁺ (K)	MF	O/C _{wa}	H/C _{wa}	DBE _{wa}	Al _{mod}	T _{g,RH} ⁺ (K)	
		ESI						LDI**						
BBAug09(DT)	All	98	0.73 (0.19)	1.00 (0.34)	11.07 (3.22)	0.31 (0.26)	--	Data not acquired						
	CHO	57	0.72 (0.17)	1.02 (0.33)	11.38 (2.81)	0.29 (0.24)	351.13 (296.70)							
	CHNO	36	0.70 (0.17)	0.90 (0.32)	11.41 (3.25)	0.40 (0.27)	--							
	CHOS	5	0.96 (0.35)	1.43 (0.12)	5.16 (1.44)	00.00 (0.00)	--							
BBAug0910(NT)	All	105	0.84 (0.33)	1.65 (0.24)	4.16 (1.80)	0.17 (0.78)	--	53	0.16 (0.15)	0.88 (0.43)	8.83 (3.82)	0.61 (0.29)	--	
	CHO	3	0.47 (0.02)	1.78 (0.06)	2.95 (0.80)	0.00 (0.00)	317.87 (278.38)	20	0.21 (0.16)	1.20 (0.53)	7.72 (4.45)	0.45 (0.34)	234.70 (232.02)	
	CHNO	96	0.86 (0.34)	1.65 (0.24)	4.21 (1.78)	0.18 (0.81)	--	22	0.21 (0.11)	0.89 (0.30)	8.12 (2.40)	0.62 (0.21)	--	
	CHOS	6	0.81 (0.07)	1.64 (0.22)	3.86 (2.01)	0.00 (0.00)	--	Not detected.						
BBAug1011	All	137	0.86 (0.43)	1.64 (0.31)	4.65 (2.70)	0.62 (2.3)	--	66	0.36 (0.17)	1.13 (0.46)	7.11 (3.69)	0.42 (0.31)	--	
	CHO	25	0.33 (0.14)	1.48 (0.29)	7.43 (3.64)	0.17 (0.18)	321.29 (280.17)	29	0.27 (0.12)	1.29 (0.51)	6.47 (4.61)	0.33 (0.31)	263.75 (244.60)	
	CHNO	65	1.18 (0.38)	1.77 (0.34)	3.82 (2.49)	1.23 (3.22)	--	37	0.43 (0.18)	1.00 (0.34)	7.64 (2.57)	0.49 (0.30)	--	
	CHOS	47	0.67 (0.15)	1.54 (0.20)	4.58 (1.45)	0.00 (0.04)	--	Not detected.						
BBAug1314	All	851	0.54 (0.27)	1.21 (0.34)	10.07 (4.70)	0.26 (0.21)	--	2006	0.72 (0.35)	1.11 (0.32)	9.81 (4.88)	0.34 (0.81)	--	

	CHO	349	0.46 (0.19)	1.08 (0.29)	12.03 (4.24)	0.37 (0.18)	326.08 (282.65)	879	0.57 (0.24)	1.00 (0.27)	12.32 (3.86)	0.36 (0.23)	348.89 (189.93)
	CHNO	444	0.59 (0.31)	1.29 (0.31)	9.19 (4.18)	0.20 (0.20)	--	698	0.54 (0.19)	0.98 (0.25)	12.02 (3.37)	0.38 (0.23)	--
	CHOS	58	0.66 (0.29)	1.56 (0.42)	3.30 (2.29)	0.04 (0.08)	--	423	1.09 (0.30)	1.40 (0.26)	4.38 (2.05)	0.28 (1.42)	--

*Average ambient temperatures recorded at the Hanford Meteorology Station (25 miles northwest of Richland, WA) during sample collection were 304.26 (Aug 09, 2018), 305.10 (Aug 09-10, 2018), 302.32 (Aug 10-11, 2018), and 296.76 K (Aug 13-14, 2018), respectively, with average relative humidity of 21.9, 24.4, 29.8, and 32.0%.

**Two additional molecular groups: CH and CHN were detected here with LDI. They made up only <1% of all assignments, and hence, have been omitted from this summary.

5.4 Research implications

Tar balls efficiently scatter and absorb light in the ultraviolet⁴⁴ and near-infrared region³⁶. Previously, they have been deemed similar to high-molecular weight humic-like substances formed by polymerization and gas-to-particle transition of breakdown products of lignin, such as syringol, guaiacol, etc. and their derivatives, which is considered to confer their unique optical properties^{36, 185}. The high-resolution molecular analysis carried out in our study provides evidence for a majority of TB constituents to be a complex mixture of unsaturated moieties of low O and high N content. Our findings further support the presence of an oxygenated exterior, which may be an important determinative factor in the hygroscopicity of aged tar balls. Under suitable conditions of supersaturation and temperature, TB may potentially function as CCN, and thus, indirectly effect Earth's radiative balance and climate³⁸. Previously, low hygroscopicity and insolubility in water has been reported for fresh TB³⁵. Increase in hydrophilicity and water-solubility has been observed with atmospheric aging that was attributed to mixing with SO_4^{2-} or incorporation/formation of hydrophilic particles with time¹⁶. Similarly, Hand et al.³⁸ observed weak hygroscopicity of TB at 83.4% relative humidity (RH) with no hygroscopic growth at <83% RH. Complete TB degradation occurred at >92% RH. Knowing the exact molecular nature of TB will provide invaluable insight into the very basis of its polarity and oxidation state, which are major determinative factors for hygroscopicity. Furthermore, mass spectrometric studies as ours estimate physical parameters, such as phase state under different RH and ambient temperature conditions, with the aid of suitable models, all of which is crucial in extrapolating the atmospheric fate, function, and eventual large-scale and long-term influence of TB released into the environment.

5.5 Data availability

Supplementary Dataset S1 contains a full list of masses and formulae assigned by ESI for all samples. Supplementary Dataset S2 contains a full list of masses and formulae assigned by LDI for all samples. Supplementary Dataset S3 present 851 species that were uniquely identified in BBAug1314 with respect to all three Non-TB samples: BBAug1011, BBAug09(DT), BBAug0910(NT).

5.6 Supplementary material

Sampling campaign. Aerosol samples were collected on 90-mm PTFE filters by pulling 80 L of air per minute for varied lengths of times on four days in August 2018 at the Pacific Northwest National Laboratory, Richland, WA, United States. Intense smoky conditions were recorded on all days owing to wildfires burning northwest of the site of sample collection (Figure 5.1). Backward trajectories extending 72 hours from the time that sample collection was concluded were modelled using HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) at heights of 350, 1000, and 1500 meters above ground level (Figure S5.1). These models indicated all aerosol samples to be representative of plumes that had come directly in contact with or travelled in close vicinity of wildfires ~XX km upwind. Two samples were collected over full days (25-28 hours), while one each (<15 hours) was collected in daytime and nighttime (Table S5.1). One-fourth of each filter was immersed in analytical-grade acetonitrile and shaken at 60 rpm for 90 minutes. The filter was then discarded, and liquid I-aerosol mixture was passed through a 0.2- μ m PTFE syringe to remove suspended solid material.

5.6.1 Ultrahigh-resolution mass spectrometry

5.6.1.1 Electrospray ionisation

Immediately prior to analysis, aerosol samples extracted in 100% acetonitrile (I) were diluted 10-fold in 50:50 H₂O:I. Spectra were acquired on a 15-Tesla Bruker Solarix XR FT-ICR MS equipped with a ParaCell and actively shielded superconducting magnet. Samples were infused at 5 μ L/min into an Apollo II ESI source operating in negative-ion mode with a capillary voltage of 4 kV, dry gas flow of 4 L/min, temperature of 180°C, and nebulizer gas pressure of 1 bar. The Q1 mass was set to m/z 100. Ions were accumulated for between 50 and 80 ms prior to transfer to the analyser cell with a time-of-flight of 0.75 ms. All spectra were acquired between m/z 92.1 and 1200.0 into an 8 Mword time domain with transient length of 1.67 s. Three hundred transients were co-added prior to zero filling and Fourier transform. The measured resolving power was about 600,000 at m/z 400. Spectra were peak picked in DataAnalysis 5.0 (Bruker Daltonics, Billerica MA) with a S/N threshold of 4.0 and absolute intensity threshold of 1,000,000.

5.6.1.2 Laser desorption ionisation

Substrates with aerosol deposited on them were affixed to a polished steel matrix-assisted LDI (MALDI) target with double-sided conductive copper tape. Spectra were acquired on a 15-Tesla Bruker Solarix XR FTICR MS equipped with a ParaCell and actively shielded superconducting magnet. Laser desorption ionisation in negative-ion mode was performed with a solid-state smartbeam-II (355 nm) laser. Given the heterogeneity of the sample, and resulting scan-to-scan variation, spectra were acquired in imaging mode. Laser focus was set to small with power at 60%; 100 laser shots were fired at 200 Hz per pixel. Approximately 600 pixels were acquired per region at 100 μm square pixel width, or up to 1000 pixels with 50 μm pixel width. The Q1 mass was set to m/z 100. Each spectrum was recorded between m/z 92.1 and 1200 in an 8 Mword time domain with transient length of 1.67s. Time of flight (ToF) was 0.75 ms. Post-processing of the spectra was performed using custom Python routines. Briefly, the total ion current (TIC) values for each region of interest and pixel were collated, and a narrow subset ($n = 48$ transients) of similar TICs – per ROI – were selected for transient co-addition. These were then zero filled and Fourier transformed prior to frequency-to-mass calibration and peak picking. This ionisation technique was also applied to aerosol samples extracted in 100% I aerosol, dried, and reconstituted in 50:50 H₂O:I. Two microliters of samples were deposited on a polished steel MALDI target plate. Spectra were acquired by co-adding 64 transients recorded from m/z 92.14-1400. Two hundred laser shots were fired at random with 400 Hz per pixel and laser power of 28%. Settings for time domain, transient length, Q1 mass, and ToF were same as for LDI in imaging mode.

5.6.1.3 Data processing and formula assignment

Mass lists acquired from FT-ICR MS analyses above were subjected to noise estimation using the KMDNoise() function in MFAssignR1 to remove residual noise, if any. Monoisotopic and polyisotopic peaks were identified using IsoFiltR(). CH₂-homologous series were chosen from within the preliminary C, H, and O-containing formulae assigned to each aerosol using Recal(). Molecular formulae were assigned to recalibrated m/z values using MFAssign() within the following constraints: C₆H₈O₆N₀-3S₀-113C₀-234S₀-1; $-13 \leq \text{DBE-O} \leq 20$; $0 \leq \text{O/C} \leq 2.0$; and $0.3 \leq \text{H/C} \leq 2.5$. Maximum error 0.5 ppm was allowed between

measured mass of ion and theoretical mass of [M-H]⁻ formula assigned to it. Assignment of odd-electron ions was considered for mass lists generated by LDI. Unless otherwise indicated, assignments containing ¹³C1-2 and/or ³⁴S atoms (or those assigned as odd-electron ions in LDI) are not included in our discussion, considering that they are duplicates of corresponding monoisotopic (and/or deprotonated) formulae. Where required for the sake of comparison between samples, peak intensities in a given dataset were normalised to the sum of abundance of C,H,O-containing species that it shared with $\geq(n-1)$ samples, where n is the total number of samples under consideration. Structural information was inferred by calculating double bond equivalence (DBE) and modified aromaticity index (Almod)^{2, 3}. Physicochemical parameters, including saturation mass concentration of pure compounds (C₀) and volatility classification^{4, 5}, and dry⁶ and relative humidity-dependent⁷ glass transition temperatures (T_{g,dry} and T_{g,RH}) were modelled using the equations listed in Table S2.

General observation from mass spectral data. On average, 85.30 ± 0.35 , 68.45 ± 0.97 , 70.55 ± 3.26 , $72.83 \pm 2.15\%$ peaks were assigned an unambiguous formula in duplicate spectral acquisitions of BBAug1314, BBAug1011, BBAug09(DT), and BBAug0910(NT) in ESI and 88.54 ± 2.67 , 88.52 ± 2.09 , and $82.96 \pm 1.07\%$ peaks were assigned a formula in LDI analysis. In ESI analysis, a bulk of unassigned peaks was quite intense relative to the assigned portion of the spectra. These peaks constituted 8.0-30% of the total ion current above the noise threshold in the different replicated spectral acquisitions for the four samples. We speculate that their ions contained unknown elements or highly condensed species, whose composition lay outside of the formula assignment constraints applied in this study. To gauge whether the lack of assignments for these peaks affected the comparative results we aimed to present here, we aligned their masses within 0.5 ppm of one another to find that these peaks were largely shared between all samples, and thus, did not contribute to the variability between them.

5.6.2 Supplementary tables

Table S5.1 Sample ID of aerosol mixtures analyzed here with duration and timings of sample collection

Sample ID	Start	End	Duration of collection	Status
BBAug09(DT)2018	Aug 09 (09:00)	Aug 09 (17:00)	8 hours	Non-TB

BBAug0910(NT)2018	Aug 09 (18:00)	Aug 10 (07:00)	13 hours	Non-TB
BBAug10112018	Aug 10 (11:00)	Aug 11 (14:00)	28 hours	Non-TB
BBAug13142018	Aug 13 (10:00)	Aug 14 (11:00)	25 hours	TB-rich

Table S5.2 Equations/models used to compute physicochemical parameters from formulae assigned to high-resolution mass spectrometric data

Parameter	Equation	Reference
Modified aromaticity index	$AI_{mod} = (1 + C - (0.5 \times O) - S - (0.5 \times H) - (0.5 \times N)) / (C - (0.5 \times O) - S - N)$	99, 100
C, O, S, H, and N are the numbers of atoms of these elements in each molecular formula. $AI_{mod} \leq 0$ were categorized as aliphatic; $0 < AI_{mod} \leq 0.5$ and $O/C < 0.5$ were classified as unsaturated with low O; $0 < AI_{mod} \leq 0.5$ and $O/C > 0.5$ were classified as unsaturated with high O; $0.5 < AI_{mod} < 0.67$ were classified as aromatic, and $AI_{mod} \geq 0.67$ were classified as condensed aromatic.		
Average oxidation state of C	$OS_c = 2 \times (O/C) - (H/C) - 5 \times (N/C) - 6 \times (S/C) - 5 \times (P/C)$	94
C, O, S, H, and N are the numbers of atoms of these elements in the molecular formulae.		
Estimated saturation mass concentrations	$\log_{10}C_0 = (nC - C) \times bC - O \times bO - H \times bH - 2 \times ((C \times O) / (C + O)) \times bCO - N \times bN - S \times bS$	95
C, O, S, H, and N are the numbers of atoms of these elements in each molecular formula. nC is the reference C number; bC, bO, bN, and bS denote the contribution of each atom to the estimated C_0 . bCO is the C-O non-ideality. The values of these constants are different based on molecular groups to which a given formula belongs (CHO, CHNO, CHOS, CH, CHNOS, or CHN) and were accessed from the report published by Li et al., 2016 ⁹⁵ . Formulae with $\log_{10}C_0 \leq -3.52$ were categorized as extremely low volatility (ELVOC), $-3.52 < \log_{10}C_0 \leq -0.52$ as low volatility (LVOC), $-0.52 < \log_{10}C_0 \leq 2.47$ as semi-volatile (SVOC), $2.47 < \log_{10}C_0 \leq 6.47$ as intermediate volatility (IVOC), and those with $6.47 < \log_{10}C_0$ as volatile organic compounds (VOC).		
Dry glass transition temperature	$T_{g,dry} = nC + \log I \times bC + \log(H) \times bH + \log I \times \log(H) \times bCH + \log(O) \times bO + \log I \times \log(O) \times bCO$	96

nOC is the reference C number, bC, bH, and bO represent the numeric contribution of these elements to T _{g,dry} . Also, bCH and bCO are coefficients that reflect contributions from C-H and C-O bonds, respectively. This equation is applicable to CHO species only. The values of these constants were accessed from the report published by DiRieux et al., 2018 ⁹⁶ .		
Relative humidity-dependent glass transition temperatures	T _{g,RH} = (1 – Worg)×T _{g,w} + (1/Kgt)×Worg×T _{g,dry})/((1 – Worg) + (1/Kgt)×Worg)	97
	Worg = 1.4 – ((1.4×RH)/100))/(1.4 – ((1.28×RH)/100))	
T _{g,w} is the glass transition temperature of water, i.e., 136 K, and Kgt is the Gordon-Taylor constant that is equal to 2.5 ± 1.5. Furthermore, Worg was calculated using the following equation: Worg = 1.4 – ((1.4×RH)/100))/(1.4 – ((1.28×RH)/100)). Here, RH is the measured relative humidity on the day of sampling in Richland, WA. Average ambient temperatures recorded at the Hanford Meteorology Station (25 miles northwest of Richland, WA) during sample collection were 304.26 (Aug 09, 2018), 305.10 (Aug 09-10, 2018), 302.32 (Aug 10-11, 2018), and 296.76 K (Aug 13-14, 2018), respectively, with average RH of 21.9, 24.4, 29.8, and 32.0%.		

Table S5.3 Summary of the chemical nature of aerosol mixtures inferred from molecular species detected by negative electrospray (ESI) and laser desorption/ionization (LDI) with high resolution mass spectrometric analysis. All averages are weighted to normalized abundance.

Aerosol mixture	Molecular group	MF	O/C _{wa}	H/C _{wa}	DBE _{wa}	Al _{mod}	T _{g,RH} [*] (K)	MF	O/C _{wa}	H/C _{wa}	DBE _{wa}	Al _{mod}	T _{g,RH} [*] (K)
		ESI							LDI**				
BBAug09(DT)	All	3663	0.68 (0.18)	1.16 (0.28)	7.44 (3.06)	0.24 (0.26)	--						
	CHO	1847	0.68 (0.17)	1.14 (0.27)	7.39 (2.93)	0.26 (0.20)	313.92 (272.35)						
	CHNO	1608	0.68 (0.16)	1.16 (0.28)	8.29 (2.96)	0.22 (0.20)	--						
	CHOS	208	1.00 (0.27)	1.62 (0.20)	2.89 (1.20)	0.12 (0.87)	--						
BBAug0910(NT)	All	3686	0.69 (0.17)	1.25 (0.29)	6.60 (2.79)	0.20 (0.27)	--	1516	0.48 (0.21)	1.06 (0.33)	7.12 (2.95)	0.40 (0.25)	--
	CHO	1554	0.67 (0.16)	1.19 (0.27)	6.69 (2.70)	0.23 (0.20)	308.13 (269.98)	967	0.50 (0.21)	1.10 (0.34)	6.76 (2.95)	0.37 (0.25)	280.04 (253.69)
	CHNO	1876	0.70 (0.19)	1.30 (0.30)	6.97 (2.76)	0.15 (0.22)	--	523	0.43 (0.17)	0.94 (0.25)	8.11 (2.35)	0.50 (0.22)	--
	CHOS	256	0.96 (0.26)	1.63 (0.19)	2.90 (1.22)	0.10 (0.80)	--	Not detected					
BBAug1011	All	3922	0.68 (0.20)	1.25 (0.29)	6.63 (2.92)	0.21 (0.41)	--	2383	0.53 (0.21)	1.01 (0.31)	7.88 (3.11)	0.41 (0.24)	--
	CHO	1734	0.65 (0.16)	1.20 (0.26)	6.76 (2.76)	0.23 (0.19)	304.65 (265.93)	1447	0.54 (0.22)	1.04 (0.32)	7.54 (3.18)	0.38 (0.25)	286.25 (254.83)
	CHNO	1865	0.68 (0.20)	1.27 (0.30)	7.36 (2.89)	0.20 (0.62)	--	928	0.48 (0.17)	0.92 (0.24)	8.89 (2.58)	0.49 (0.22)	--
	CHOS	323	0.94 (0.26)	1.62 (0.21)	2.95 (1.27)	0.09 (0.74)	--	Not detected					
BBAug1314	All	4841	0.67 (0.21)	1.17 (0.29)	7.46 (3.21)	0.25 (0.32)	--	4299	0.59 (0.26)	1.01 (0.30)	8.87 (3.59)	0.40 (0.44)	--
	CHO	2221	0.65 (0.19)	1.12 (0.27)	7.50 (3.14)	0.28 (0.20)	303.85 (266.34)	2281	0.56 (0.22)	0.99 (0.29)	8.89 (3.51)	0.40 (0.23)	297.65 (262.78)
	CHNO	2366	0.67 (0.21)	1.22 (0.28)	7.97 (3.05)	0.20 (0.31)	--	1581	0.51 (0.18)	0.94 (0.24)	10.01 (3.07)	0.44 (0.22)	--

	CHOS	254	1.01 (0.31)	1.63 (0.26)	2.64 (1.35)	0.19 (1.04)	--	423	1.09 (0.30)	1.40 (0.26)	4.38 (2.05)	0.28 (1.42)	--

5.6.3 Supplementary figures

Figure S5.1 Backward trajectories obtained from HYSPLIT modelling. The plots show 72-hour backward trajectory (6-hour intervals) starting from 14:00 hrs on Aug 11 (going back to Aug 09) and 11:00 hrs on Aug 13 (going back to Aug 11).

Figure S5.2 Percent fraction of major particle types based on composition in particles of size ranges in non-TB (BBAug09(DT) and BBAug1011) and TB (BBAug1314) aerosol.

Figure S5.3 Optical thickness of TB in BBAug1314 determined by scanning transmission electron microscopy and near-edge X-ray absorption fine structure spectroscopy, (A) Distribution of organic and elemental carbon in single TB, (B) total carbon absorption as a function of particle size

Figure S5.4 van Krevelen diagrams depicting the similarity of LDI imaging with two other ionisation methods: LDI and ESI on acetonitrile extracts

Figure S5.5 Van Krevelen diagrams highlighting the volatility of species exclusively identified in each aerosol mixture by ESI (panel on the left) and LDI (panel on the right) analysis. Gray species in the background depict entire datasets and are overlaid by exclusively detected species coloured by their estimated saturation mass concentration (C_0), i.e., volatility.

Figure S5.6 Reconstructed mass spectra from ESI/15-T FT-ICR MS analysis of aerosol mixtures. Unassigned peaks, contaminants, and/or noise are not shown. Pie charts show the fraction of assigned dataset belonging to CHO, CHNO, and CHOS molecular groups.

Figure S5.7 Reconstructed mass spectra from LDI/15-T FT-ICR MS analysis of aerosol mixtures. Unassigned peaks, contaminants, and/or noise are not shown. Pie charts represent fractions of the assigned dataset belonging to CHO, CHNO, CHOS, CH, and CHN molecular groups.

Figure S5.8 Ratios between major functional groups determined using scanning transmission X-ray microscopy analysis of two non-TB aerosol samples, BBAug 09(DT) and BsAug1011, relative to the TB-rich aerosol, BBAug1314

Figure S5.9 A comparison of the contribution of a variety of aromatic and non-aromatic units to the total ion current of TB-rich (BBSep05 and BBAug1314) and TB-poor samples (BBAug1011).

6 Conclusions

Natural organic matter consists of thousands or even millions of individual chemical species that are present as undefinable solutes in a single soup. In aquatic ecosystems, OM participates in crucial processes, such as transport of nutrients; degradation, photolysis, and volatilization of organic pollutants¹⁰⁹; chelation of metals to alter their availability and toxicity in the environment^{110, 111}, photochemical reactions¹¹², and nutrient cycling¹¹³, etc. When present in the atmosphere, OM is extremely important for its ability to directly and indirectly affect the radiative balance of the Earth, and thus, can profoundly affect the climate^{182, 186}. These capabilities of environmental OM are dependent upon its physicochemical properties that are, in turn, determined by their molecular composition. Aerosol mass spectrometers have been employed to obtain real-time measurements on aerosol composition, however, these instruments are not high-resolution instruments and provide information on bulk chemistry only. Offline measurements following sampling are required for a detailed look into the molecular composition^{187, 188}. In this dissertation, the major objective was to elucidate the molecular composition of complex environmental mixtures from aquatic and atmospheric sources. This work was presented in four chapters. Specific conclusions derived from each are presented below.

6.1 Conclusions from Chapter 2

Expanding the Molecular Characterization of Natural Organic Matter with Complementary Ionization Methods

Conventionally, either positive- or negative-mode ESI is used as a method of choice for mass spectrometric analysis of NOM. Our findings on the molecular composition of three commercial IHSS mixtures using six popular ionization methods, ESI, APCI, and APPI, in positive and negative polarities corroborates the concept of ionization selectivity and the need for employing more than one method – preferably ionization methods that can access and ionize complementary molecular fractions – for a more comprehensive acquisition of information on the composition of complex mixtures. The apparent differences in composition of the same sample with six methods shows that even if ≥ 2 samples are used, there will always be a part of natural complex mixtures that will remain elusive. It is crucial that mass spectrometric analysis be performed as

part of integrated studies that can support the information derived from mass spectrometry.

6.2 Conclusions from Chapter 3

Qualitative Molecular Characterization of Fulvic Acid Mixtures from the International Humic Substances Society with High-Resolution Orbitrap Mass Spectrometry

Reference and standard mixtures made available by the IHSS have long been an integral component of analytical studies involving complex organic mixtures to ensure objectivity. Owing to their origin being natural environments, they are sufficiently complex to be of environmental relevance, but the uniformity with which they are prepared makes them of special interest for sensitive analytical studies. Therefore, IHSS references and standards have been applied and studied extensively for diverse purposes ranging from method development to gauging instrumental performances. Our findings on the molecular composition of six DOM mixtures from the IHSS using electrospray ionization and an FT-OTE MS, in combination with data processing and formula assignment tool, MFAssignR, is in agreement with previously published data on the molecular composition of these mixtures. We have compiled our findings to provide mass lists comprising distinct molecular formulae that can be applied as a reference to gauge the accuracy of mass spectrometric studies employing any one of these IHSS mixtures as a control. Furthermore, based on the molecular compositions of the DOM mixtures described here, researchers must be able to choose the most chemically and practically pertinent IHSS reference or standard in their study, not only for the sake of analytical validation but to provide a basis for comparison with organic matter from similar ecosystems as well.

6.3 Conclusion from Chapter 4

High-Resolution Molecular Characterization of Tar Ball-Rich Biomass Burning Organic Aerosol with a 21-T FT-ICR Mass Spectrometer

Molecular characterization of a BBOA mixture rich in tar balls was performed to support well-established notions about the composition of such aerosol, as well

as to provide novel insights into their relatively less abundant chemical components. A high-resolution mass spectrometer, the 21-T FT-ICR MS, detected 10,533 distinct formulae that belonged to five molecular groups: CHO, CHNO, CHOS, CHOP, and CHNOP. A dominance of CHNO species, in combination with oligomerization of CHO species, suggested this aerosol mixture to have undergone extensive atmospheric processing, while low O/C_{avg} (0.46 ± 0.2) and O_{avg} (9.80 ± 3.6) are indicative of conglomerated particles, such as tar balls, that prevent it from oxygenating despite extensive processing during transport. Furthermore, although the CHOP and CHNOP groups formed only 6.57% of all assigned formulae, they are extremely environmentally relevant due to their tendency for wet and dry deposition into aquatic ecosystem from aerosol, leading to changes in nutritional balance, particularly eutrophication. Despite evidence of the occurrence of P-containing species in BBOA, detailed molecular analysis of this fraction is sparse. We recorded similar, although not identical, findings from a 15-T FT-ICR MS instrument as well. Characterization studies as those presented here reveal the complex nature of aerosol mixtures, and thus, build foundation for future such studies on aerosol mixtures from other sources as well. Such detailed molecular-level information is crucial to understanding chemical signatures of aerosol source, atmospheric aging of aerosol, as well as the interaction and influence of organic aerosol mixtures with ecosystems and their biotic and abiotic components. Comparative analysis of observations from the two FT-ICR instruments suggested that superior performance of the 21-T FT-ICR instrument was not solely because of the strength of the magnet and the benefits incurred by it in terms of resolving power, but better sensitivity and dynamic range play a substantial role. The ESI sources and instrumental designs are important determinative factors in the fine-scale analytical information provided by these instruments as well.

6.4 Conclusions from Chapter 5

Insights into the Molecular Composition of Ambient Tar Balls: An Analysis Using 15-T FT-ICR Mass Spectrometry with Electrospray and Laser Desorption Ionization

We have presented detailed qualitative high-resolution mass spectrometric analysis of TB-rich aerosol. Our findings from 15-T FT-ICR MS analysis, in conjunction with ESI and LDI, support the designation of TB being processed primary particles. Their characteristic atmospheric processing is clearly exhibited in the presence of an outer oxygenated interface that encloses a complex mixture of less oxygenated organic matter. Furthermore, by aligning molecular

species that were exclusive to TB-rich aerosol from the year 2018 with another TB-rich aerosol collected a year before showed a much higher overlap between these two as compared to any other aerosol mixture collected in the same year/week. Identification of molecular species commonly occurring in TB-rich aerosol collected farther apart in time will be immensely valuable in credibly identifying a signature set of compounds, whose presence unequivocally encourages the formation of TB in one plume, while their absence may lead to no TB formation. The role of ambient conditions, such as temperature, relative humidity, height of plume, etc. in dictating these processes is another crucial area of study.

6.5 Future recommendations

In this research, a total of eight complex mixtures were analyzed using high-resolution mass spectrometry: three from aquatic sources and five from atmospheric sources. Regardless of source, one aspect remained certain – the molecular complexity of these natural mixtures was astounding. Some mixtures, such as the BBOA mixture presented in Chapter 4 was analyzed using three types of mass spectrometers differing in their resolving powers. There seemed to be no end to the complexity; with every enhancement in resolving power, we could see more constituents. This observation asserts the elusive nature of such complex mixtures and also provides motivation for continuous development of better analytical instruments and approaches. A major approach toward elucidating the chemistry of complex mixtures with high-resolution mass spectrometry analysis is to look at them as the average of the characteristics of accessible molecules. The more molecules we are able to access, the more representative the conclusions would be of the whole. Reduction of mixture complexity is one way to better access species that would otherwise be lost in the original complexity due to ion suppression¹⁸⁹. For instance, in case of organic aerosol, separation of molecular components by their size is an interesting aspect that was not explored in this dissertation. Ambient organic aerosol can be size-resolved into multiple stages by techniques such as the micro-orifice uniform deposition cascade impactor (MOUDI)¹⁹⁰. Such a method would distribute a single complex organic aerosol mixture into several stages, and thus, relatively simpler mixtures that would not only be easier to analyse but would provide insight into formation of SOA under ambient conditions as well. High-resolution analysis of molecular changes in SOA formed by plant-based VOCs, such as pinene and isoprene, has been extensively studied at lab-scale¹⁹¹⁻¹⁹⁶, but studies on ambient SOA are still sparse.

One way that high-resolution mass spectrometric data is put to good use is to make educated inferences about the physicochemical nature of complex mixtures. For instance, Section 1.2.4.2 discussed several parameters, such as glass transition temperatures, volatility, aromaticity, etc. that can be predicted using the elemental compositions experimentally determined from mass spectrometry. However, such models are sparse and are applicable to C, H, and O-containing species only. Although elucidation of the individual molecular constituents of one aerosol mixture after the next may seem like a redundant task, the importance for such a reservoir of information for these models cannot be denied. The major way that atmospheric aerosol asserts its influence is via climatic changes. The distribution of aerosol, their hygroscopicity, and optical properties are important factors that dictate whether they can have a direct influence or indirect influence on the climate^{4, 9}. All of these properties can be estimated using elemental compositions, which highlights the need to acquire and analyse aerosol from not only biomass burning sources, but other sources producing OM as well, such as trash burning, suspension of soil or dust in the atmosphere, etc.

High-resolution mass spectrometric datasets can also be used to isolate molecular signatures^{118, 127} and to indicate molecular markers of different sources¹⁵⁸. In this dissertation, we were able to find high similarity between two aerosol samples collected a year apart, which is an indication of a TB molecular signature. However, such a signature for TB cannot be validated with the utmost confidence with only two TB-rich samples. Continuous sampling – preferably farther apart in time – can effectively help in substantiating such a signature for TB. Once particular molecular entities are associated with TB, it would provide a strong basis to predict molecular structures, generate potential such species in lab, and to understand their optical and physicochemical properties.

7 Works Cited

1. Nebbioso, A.; Piccolo, A., Molecular characterization of dissolved organic matter (DOM): a critical review. *Analytical and bioanalytical chemistry* **2013**, *405* (1), 109-124.
2. Dittmar, T.; Stubbins, A., Dissolved organic matter in aquatic systems. *Treatise on Geochemistry, 2nd edn. Elsevier: Oxford* **2014**, 125-156.
3. Hoffman, E. d.; Stroobant, V., Mass spectrometry: principles and applications. *West Sussex: John Wiley & Sons, Bruxellas, Bélgica* **2007**, 1 (2), 85.
4. Boucher, O.; Randall, D.; Artaxo, P.; Bretherton, C.; Feingold, G.; Forster, P.; Kerminen, V.-M.; Kondo, Y.; Liao, H.; Lohmann, U.; Rasch, P.; Satheesh, S. K.; Sherwood, S.; Stevens, B.; Zhang, X. Y., Clouds and Aerosols. In *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*, Stocker, T.; Qin, D.; Plattner, G. K.; Tignor, M.; Allen, S. K.; Boschung, J.; Nauels, A.; Xia, Y.; Bex, V.; Midgley, P. M., Eds. Cambridge University Press: Cambridge, United Kingdom and New York, USA, 2014.
5. Mouli, P. C.; Mohan, S. V.; Reddy, S. J., A study on major inorganic ion composition of atmospheric aerosols at Tirupati. *Journal of hazardous materials* **2003**, *96* (2-3), 217-228.
6. Seinfeld, J., Atmospheric science: Black carbon and brown clouds. *Nature geoscience* **2008**, *1* (1), 15.
7. Lohmann, U.; Feichter, J., Global indirect aerosol effects: a review. *Atmospheric Chemistry and Physics* **2005**, *5* (3), 715-737.
8. Sun, J.; Ariya, P. A., Atmospheric organic and bio-aerosols as cloud condensation nuclei (CCN): A review. *Atmospheric Environment* **2006**, *40* (5), 795-820.
9. Seinfeld, J. H.; Pandis, S. N., Atmospheric Organic Aerosols. In *Atmospheric chemistry and physics: from air pollution to climate change*, 3rd ed.; John Wiley & Sons: 2016.
10. Friedlander, S. K., *Smoke, dust, and haze*. Oxford university press: 2000.
11. Odman, M. T.; Hu, Y.; Garcia-Menendez, F.; Davis, A. Y.; Chang, M. E.; Russell, A. G., Fires and Air Quality Forecasts: Past, Present, and Future *Air & Waste Management Association* 2013.
12. Wielgoński, G., Pollutant formation in combustion processes. *Advances in Chemical Engineering, Zeeshan Nawaz & Shahid Naveed (Ed.), InTech Rijeka* **2012**, 295-324.
13. Hoover, K.; Hanson, L. A. *Wildfire Statistics*; Congressional Research Service (CRS): Washington DC, USA, 2021.

14. Westerling, A. L.; Hidalgo, H. G.; Cayan, D. R.; Swetnam, T. W., Warming and earlier spring increase western US forest wildfire activity. *science* **2006**, *313* (5789), 940-943.
15. Spracklen, D. V.; Logan, J. A.; Mickley, L. J.; Park, R. J.; Yevich, R.; Westerling, A. L.; Jaffe, D. A., Wildfires drive interannual variability of organic carbon aerosol in the western US in summer. *Geophysical Research Letters* **2007**, *34* (16).
16. Pósfai, M.; Gelencsér, A.; Simonics, R.; Arató, K.; Li, J.; Hobbs, P. V.; Buseck, P. R., Atmospheric tar balls: Particles from biomass and biofuel burning. *Journal of Geophysical Research: Atmospheres* **2004**, *109* (D6).
17. Pósfai, M.; Simonics, R.; Li, J.; Hobbs, P. V.; Buseck, P. R., Individual aerosol particles from biomass burning in southern Africa: 1. Compositions and size distributions of carbonaceous particles. *Journal of Geophysical Research: Atmospheres* **2003**, *108* (D13).
18. Andreae, M. O.; Merlet, P., Emission of trace gases and aerosols from biomass burning. *Global biogeochemical cycles* **2001**, *15* (4), 955-966.
19. Bond, T. C.; Doherty, S. J.; Fahey, D. W.; Forster, P. M.; Bernsten, T.; DeAngelo, B. J.; Flanner, M. G.; Ghan, S.; Kärcher, B.; Koch, D., Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of geophysical research: Atmospheres* **2013**, *118* (11), 5380-5552.
20. Robinson, A. L.; Donahue, N. M.; Shrivastava, M. K.; Weitkamp, E. A.; Sage, A. M.; Grieshop, A. P.; Lane, T. E.; Pierce, J. R.; Pandis, S. N., Rethinking organic aerosols: Semivolatile emissions and photochemical aging. *Science* **2007**, *315* (5816), 1259-1262.
21. Sun, H.; Biedermann, L.; Bond, T. C., Color of brown carbon: A model for ultraviolet and visible light absorption by organic carbon aerosol. *Geophysical Research Letters* **2007**, *34* (17).
22. Andreae, M. O.; Gelencsér, A., Black carbon or brown carbon? The nature of light-absorbing carbonaceous aerosols. *Atmospheric Chemistry and Physics* **2006**, *6* (10), 3131-3148.
23. Desyaterik, Y.; Sun, Y.; Shen, X.; Lee, T.; Wang, X.; Wang, T.; Collett, J. L., Speciation of "brown" carbon in cloud water impacted by agricultural biomass burning in eastern China. *Journal of Geophysical Research: Atmospheres* **2013**, *118* (13), 7389-7399.
24. Adachi, K.; Sedlacek, A. J.; Kleinman, L.; Springston, S. R.; Wang, J.; Chand, D.; Hubbe, J. M.; Shilling, J. E.; Onasch, T. B.; Kinase, T., Spherical tarball particles form through rapid chemical and physical changes of organic matter in biomass-burning smoke. *Proceedings of the National Academy of Sciences* **2019**, *116* (39), 19336-19341.
25. Lee, A. K.; Zhao, R.; Li, R.; Liggió, J.; Li, S.-M.; Abbatt, J. P., Formation of light absorbing organo-nitrogen species from evaporation of droplets containing glyoxal and ammonium sulfate. *Environmental science & technology* **2013**, *47* (22), 12819-12826.

26. Hawkins, L. N.; Welsh, H. G.; Alexander, M. V., Evidence for pyrazine-based chromophores in cloud water mimics containing methylglyoxal and ammonium sulfate. *Atmospheric Chemistry and Physics* **2018**, *18* (16), 12413-12431.
27. Saleh, R.; Cheng, Z.; Atwi, K., The brown-black continuum of light-absorbing combustion aerosols. *Environmental Science & Technology Letters* **2018**, *5* (8), 508-513.
28. Yu, L.; Smith, J.; Laskin, A.; Anastasio, C.; Laskin, J.; Zhang, Q., Chemical characterization of SOA formed from aqueous-phase reactions of phenols with the triplet excited state of carbonyl and hydroxyl radical. *Atmospheric Chemistry and Physics* **2014**, *14* (24), 13801-13816.
29. Yu, L.; Smith, J.; Laskin, A.; George, K. M.; Anastasio, C.; Laskin, J.; Dillner, A. M.; Zhang, Q., Molecular transformations of phenolic SOA during photochemical aging in the aqueous phase: competition among oligomerization, functionalization, and fragmentation. *Atmospheric Chemistry and Physics* **2016**, *16* (7), 4511-4527.
30. Donahue, N.; Epstein, S.; Pandis, S. N.; Robinson, A., A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics. **2011**.
31. Donahue, N. M.; Kroll, J.; Pandis, S. N.; Robinson, A. L., A two-dimensional volatility basis set-Part 2: Diagnostics of organic-aerosol evolution. *Atmospheric Chemistry and Physics* **2012**, *12* (2), 615-634.
32. Jimenez, J. L.; Canagaratna, M.; Donahue, N.; Prevot, A.; Zhang, Q.; Kroll, J. H.; DeCarlo, P. F.; Allan, J. D.; Coe, H.; Ng, N., Evolution of organic aerosols in the atmosphere. *Science* **2009**, *326* (5959), 1525-1529.
33. Donahue, N.; Chuang, W.; Epstein, S. A.; Kroll, J.; Worsnop, D.; Robinson, A.; Adams, P.; Pandis, S., Why do organic aerosols exist? Understanding aerosol lifetimes using the two-dimensional volatility basis set. *Environmental Chemistry* **2013**, *10* (3), 151-157.
34. Kleinman, L.; Adachi, K.; Chand, D.; Sedlacek, A. J. I.; Buseck, P. R., Transmission Electron Microscopy Analysis of Tarball Formation and Volatility from Biomass-burning Aerosol Particles during the 2013 BBOP Campaign (Abstract A52A-03). In *AGU Fall Meeting Talk 2014*, 2014 Fall Meeting, AGU, San Francisco, Calif., 15-19 December: 2014.
35. Adachi, K.; Buseck, P. R., Atmospheric tar balls from biomass burning in Mexico. *Journal of Geophysical Research: Atmospheres* **2011**, *116* (D5).
36. Tivanski, A. V.; Hopkins, R. J.; Tyliszczak, T.; Gilles, M. K., Oxygenated interface on biomass burn tar balls determined by single particle scanning transmission X-ray microscopy. *The Journal of Physical Chemistry A* **2007**, *111* (25), 5448-5458.
37. Liu, C.; Chung, C. E.; Zhang, F.; Yin, Y., The colors of biomass burning aerosols in the atmosphere. *Scientific reports* **2016**, *6* (1), 1-9.
38. Hand, J. L.; Malm, W.; Laskin, A.; Day, D.; Lee, T.-b.; Wang, C.; Carrico, C.; Carrillo, J.; Cowin, J. P.; Collett Jr, J., Optical, physical, and chemical

- properties of tar balls observed during the Yosemite Aerosol Characterization Study. *Journal of Geophysical Research: Atmospheres* **2005**, *110* (D21).
39. Sedlacek III, A. J.; Buseck, P. R.; Adachi, K.; Onasch, T. B.; Springston, S. R.; Kleinman, L., Formation and evolution of tar balls from northwestern US wildfires. *Atmospheric Chemistry and Physics* **2018**, *18* (15), 11289-11301.
  40. China, S.; Mazzoleni, C.; Gorkowski, K.; Aiken, A. C.; Dubey, M. K., Morphology and mixing state of individual freshly emitted wildfire carbonaceous particles. *Nature communications* **2013**, *4* (1), 1-7.
  41. Brege, M.; China, S.; Zelenyuk-Imre, A.; Mazzoleni, L., Extreme molecular complexity resulting in a continuum of carbonaceous species in biomass burning tar balls. *In process* **2021**.
  42. Alexander, D. T.; Crozier, P. A.; Anderson, J. R., Brown carbon spheres in East Asian outflow and their optical properties. *Science* **2008**, *321* (5890), 833-836.
  43. Bhandari, J., Morphology and Mixing State of Soot and Tar Balls: Implications for Optical Properties and Climate. **2018**.
  44. Chakrabarty, R.; Moosmüller, H.; Chen, L.-W.; Lewis, K.; Arnott, W.; Mazzoleni, C.; Dubey, M.; Wold, C.; Hao, W.; Kreidenweis, S., Brown carbon in tar balls from smoldering biomass combustion. *Atmospheric Chemistry and Physics* **2010**, *10* (13), 6363-6370.
  45. Thurman, E. M., *Organic geochemistry of natural waters*. Springer Science & Business Media: 2012; Vol. 2.
  46. Leenheer, J., Chemistry of dissolved organic matter in rivers, lakes, and reservoirs. **1994**.
  47. Leenheer, J.; Huffman, E. W., Classification of organic solutes in water by using macroreticular resin. *JR Stat. Soc. (United States)* **1976**, *4* (6).
  48. McCarthy, P.; Malcolm, R., The need to establish a reference collection of humic substances. In *Trace Organic Analysis: A New Frontier in Analytical Chemistry*, US National Bureau of Standards: 1979; pp 519-793.
  49. McCarthy, P., A proposal to establish a reference collection of humic materials for interlaboratory comparisons. *Geoderma* **1976**, *16* (2), 179-181.
  50. Abbt-Braun, G. In *30 years of IHSS*, 9th International Conference Humic Substances in Ecosystems (HSE9), 26 - 29. May 2011; 2011.
  51. Averett, R. C.; Leenheer, J.; McKnight, D. M.; Thorn, K. *Humic substances in the Suwannee River, Georgia; interactions, properties, and proposed structures*, USGPO; US Geological Survey, Map Distribution: 1994.
  52. Thorn, K. A.; Folan, D. W.; McCarthy, P., Characterization of the International Humic Substances Society standard and reference fulvic and humic acids by solution state carbon-13 (¹³C) and hydrogen-1 (¹H) nuclear magnetic resonance spectrometry. *Water-Resources Investigations Report* **1989**, *89* (4196), 1-93.
  53. Select Products. <https://ihss.humicsubstances.org/orders.html>.

54. 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.
55. Hawkes, J. A.; D'Andrilli, J.; Agar, J. N.; Barrow, M. P.; Berg, S. M.; Catalán, N.; Chen, H.; Chu, R. K.; Cole, R. B.; Dittmar, T., An international laboratory comparison of dissolved organic matter composition by high resolution mass spectrometry: Are we getting the same answer? *Limnology and Oceanography: Methods* **2020**.
56. 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. *J Anal Chem* **2016**, *88* (15), 7698-7704.
57. Bianca, M. R.; Baluha, D. R.; Gonsior, M.; Schmitt-Kopplin, P.; Del Vecchio, R.; Blough, N. V., Contribution of ketone/aldehyde-containing compounds to the composition and optical properties of Suwannee River fulvic acid revealed by ultrahigh resolution mass spectrometry and deuterium labeling. *Analytical and Bioanalytical Chemistry* **2020**, 1-11.
58. Patriarca, C.; Bergquist, J.; Sjöberg, P. J.; 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.
59. Schum, S. K. Molecular characterization of free tropospheric organic aerosol and the development of computational tools for molecular formula assignment. Dissertation, Michigan Technological University, Houghton, MI, USA, 2019.
60. Blackburn, J. W.; Kew, W.; Graham, M. C.; Uhrin, D. a., Laser desorption/ionization coupled to FTICR mass spectrometry for studies of natural organic matter. *Analytical chemistry* **2017**, *89* (8), 4382-4386.
61. Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B.; Schmitt-Kopplin, P.; Perdue, E. J. A. c., Natural organic matter and the event horizon of mass spectrometry. *Anal Chem* **2008**, *80* (23), 8908-8919.
62. Kim, D.; Kim, S.; Son, S.; Jung, M.-J.; Kim, S., Application of Online Liquid Chromatography 7 T FT-ICR Mass Spectrometer Equipped with Quadrupolar Detection for Analysis of Natural Organic Matter. *Analytical chemistry* **2019**, *91* (12), 7690-7697.
63. Baluha, D. R.; Blough, N. V.; Del Vecchio, R., Selective mass labeling for linking the optical properties of chromophoric dissolved organic matter to structure and composition via ultrahigh resolution electrospray ionization mass spectrometry. *Environmental science & technology* **2013**, *47* (17), 9891-9897.
64. Hawkes, J. A.; Sjöberg, P. J.; Bergquist, J.; Tranvik, L., Complexity of dissolved organic matter in the molecular size dimension: insights from coupled size exclusion chromatography electrospray ionisation mass spectrometry. *Faraday discussions* **2019**, *218*, 52-71.

65. 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* (1-4), 23-37.
66. Remucal, C. K.; Cory, R. M.; Sander, M.; McNeill, K., Low molecular weight components in an aquatic humic substance as characterized by membrane dialysis and orbitrap mass spectrometry. *Environmental science & technology* **2012**, *46* (17), 9350-9359.
67. Gross, J. H., *Mass spectrometry: A textbook*. Springer Science & Business Media: 2006.
68. Cook, R. D.; Lin, Y.-H.; Peng, Z.; Boone, E.; Chu, R. K.; Dukett, J. E.; Gansch, M. J.; Zhang, W.; Tolic, N.; Laskin, A., Biogenic, urban, and wildfire influences on the molecular composition of dissolved organic compounds in cloud water. *Atmospheric Chemistry and Physics* **2017**, *17* (24), 15167-15180.
69. Altieri, K.; Turpin, B.; Seitzinger, S., Oligomers, organosulfates, and nitrooxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry. *Atmospheric Chemistry and Physics* **2009**, *9* (7), 2533-2542.
70. Lin, P.; Fleming, L. T.; Nizkorodov, S. A.; Laskin, J.; Laskin, A., Comprehensive molecular characterization of atmospheric brown carbon by high resolution mass spectrometry with electrospray and atmospheric pressure photoionization. *Analytical chemistry* **2018**, *90* (21), 12493-12502.
71. Lin, P.; Rincon, A. G.; Kalberer, M.; Yu, J. Z., Elemental composition of HULIS in the Pearl River Delta Region, China: Results inferred from positive and negative electrospray high resolution mass spectrometric data. *Environmental science & technology* **2012**, *46* (14), 7454-7462.
72. Zhao, Y.; Hallar, A.; Mazzoleni, L., Atmospheric organic matter in clouds: exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry. *Atmospheric Chemistry and Physics* **2013**, *13* (24), 12343-12362.
73. Schum, S. K.; Zhang, B.; Džepina, K.; Fialho, P.; Mazzoleni, C.; Mazzoleni, L. R., Molecular and physical characteristics of aerosol at a remote free troposphere site: implications for atmospheric aging. *J Atmos Chem Phys* **2018**, *18* (19), 14017-14036.
74. Mazzoleni, L. R.; Ehrmann, B. M.; Shen, X.; Marshall, A. G.; Collett Jr, J. L., Water-soluble atmospheric organic matter in fog: exact masses and chemical formula identification by ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry. *Environmental science & technology* **2010**, *44* (10), 3690-3697.
75. 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.

76. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass spectrometry reviews* **2001**, *20* (6), 362-387.

77. Zuth, C.; Vogel, A. L.; Ockenfeld, S.; Huesmann, R.; Hoffmann, T., Ultrahigh-resolution mass spectrometry in real time: atmospheric pressure chemical ionization Orbitrap mass spectrometry of atmospheric organic aerosol. *Analytical chemistry* **2018**, *90* (15), 8816-8823.

78. Reinnig, M. C.; Warnke, J.; Hoffmann, T., Identification of organic hydroperoxides and hydroperoxy acids in secondary organic aerosol formed during the ozonolysis of different monoterpenes and sesquiterpenes by on-line analysis using atmospheric pressure chemical ionization ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2009**, *23* (11), 1735-1741.

79. Kückelmann, U.; Warscheid, B.; Hoffmann, T., On-line characterization of organic aerosols formed from biogenic precursors using atmospheric pressure chemical ionization mass spectrometry. *Analytical chemistry* **2000**, *72* (8), 1905-1912.

80. Giorio, C.; Bortolini, C.; Kourtchev, I.; Tapparo, A.; Bogialli, S.; Kalberer, M., Direct target and non-target analysis of urban aerosol sample extracts using atmospheric pressure photoionisation high-resolution mass spectrometry. *Chemosphere* **2019**, *224*, 786-795.

81. Jiang, B.; Liang, Y.; Xu, C.; Zhang, J.; Hu, M.; Shi, Q., Polycyclic aromatic hydrocarbons (PAHs) in ambient aerosols from Beijing: characterization of low volatile PAHs by positive-ion atmospheric pressure photoionization (APPI) coupled with Fourier transform ion cyclotron resonance. *Environmental science & technology* **2014**, *48* (9), 4716-4723.

82. Mirabelli, M. F.; Zenobi, R., Rapid screening and quantitation of PAHs in water and complex sample matrices by solid-phase microextraction coupled to capillary atmospheric pressure photoionization-mass spectrometry. *Journal of Mass Spectrometry* **2021**, *56* (1), e4656.

83. 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.

84. Scigelova, M.; Makarov, A., Orbitrap mass analyzer-overview and applications in proteomics. *Proteomics* **2006**, *6* (S2), 16-21.

85. Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Müller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M., Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS

- and versatile peptide fragmentation modes. *Molecular & Cellular Proteomics* **2012**, *11* (3), 0111. 013698.
86. Keministi, FTICR cell. Wikimedia, 2017.
  87. Marshall, A. G.; Guan, S., Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry* **1996**, *10* (14), 1819-1823.
  88. Shaw, J. B.; Lin, T.-Y.; Leach III, F. E.; Tolmachev, A. V.; Tolić, N.; Robinson, E. W.; Koppenaal, D. W.; Paša-Tolić, L., 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer greatly expands mass spectrometry toolbox. *Journal of The American Society for Mass Spectrometry* **2016**, *27* (12), 1929-1936.
  89. Reemtsma, T., Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry: status and needs. *Journal of chromatography A* **2009**, *1216* (18), 3687-3701.
  90. Tolić, N.; Liu, Y.; Liyu, A.; Shen, Y.; Tfaily, M. M.; Kujawinski, E. B.; Longnecker, K.; Kuo, L.-J.; Robinson, E. W.; Paša-Tolić, L., Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Analytical chemistry* **2017**, *89* (23), 12659-12665.
  91. Leefmann, T.; Frickenhaus, S.; Koch, B. P., UltraMassExplorer: a browser-based application for the evaluation of high-resolution mass spectrometric data. *Rapid Communications in Mass Spectrometry* **2019**, *33* (2), 193-202.
  92. Fu, Q.-L.; Fujii, M.; Riedel, T., Development and comparison of formula assignment algorithms for ultrahigh-resolution mass spectra of natural organic matter. *Analytica Chimica Acta* **2020**, *1125*, 247-257.
  93. Schum, S. K.; Brown, L. E.; Mazzoleni, L. R., MFAssignR: Molecular formula assignment software for ultrahigh resolution mass spectrometry analysis of environmental complex mixtures. *Environmental Research* **2020**, *191*, 110114.
  94. Kroll, J. H.; Donahue, N. M.; Jimenez, J. L.; Kessler, S. H.; Canagaratna, M. R.; Wilson, K. R.; Altieri, K. E.; Mazzoleni, L. R.; Wozniak, A. S.; Bluhm, H., Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. *Nature Chemistry* **2011**, *3* (2), 133-139.
  95. Li, Y.; Pöschl, U.; Shiraiwa, M., Molecular corridors and parameterizations of volatility in the chemical evolution of organic aerosols. *Atmos. Chem. Phys* **2016**, *16* (5), 3327-3344.
  96. DeRieux, W.-S. W.; Li, Y.; Lin, P.; Laskin, J.; Laskin, A.; Bertram, A. K.; Nizkorodov, S. A.; Shiraiwa, M., Predicting the glass transition temperature and viscosity of secondary organic material using molecular composition. *Atmospheric Chemistry and Physics* **2018**, *18* (9), 6331-6351.
  97. Shiraiwa, M.; Li, Y.; Tsimpidi, A. P.; Karydis, V. A.; Berkemeier, T.; Pandis, S. N.; Lelieveld, J.; Koop, T.; Pöschl, U., Global distribution of particle phase state in atmospheric secondary organic aerosols. *Nature communications* **2017**, *8* (1), 1-7.

98. Koop, T.; Bookhold, J.; Shiraiwa, M.; Pöschl, U., Glass transition and phase state of organic compounds: dependency on molecular properties and implications for secondary organic aerosols in the atmosphere. *Physical Chemistry Chemical Physics* **2011**, *13* (43), 19238-19255.
99. Koch, B.; Dittmar, T., From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. *Rapid Comm Mass Spectrometry* **2006**, *20* (5), 926-932.
100. Koch, B. P.; Dittmar, T., From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Comm Mass Spectrometry* **2016**, *30* (1), 250-250.
101. 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 Comm Mass Spectrometry* **2015**, *29* (24), 2385-2401.
102. D'Andrilli, J.; Foreman, C.; Marshall, A. G.; McKnight, D., Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy. *J Org Geochem* **2013**, *65*, 19-28.
103. 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.
104. Kew, W.; Mackay, C. L.; Goodall, I.; Clarke, D. J.; Uhrin, D. a., Complementary ionization techniques for the analysis of scotch whisky by high resolution mass spectrometry. *Analytical chemistry* **2018**, *90* (19), 11265-11272.
105. Brege, M.; Paglione, M.; Gilardoni, S.; Decesari, S.; Facchini, M. C.; Mazzoleni, L. R., Molecular insights on aging and aqueous-phase processing from ambient biomass burning emissions-influenced Po Valley fog and aerosol. *J Atmos Chem Phys* **2018**, *18* (17), 13197-13214.
106. Ghosh, K.; Schnitzer, M., Macromolecular structures of humic substances. *Soil Science* **1980**, *129* (5), 266-276.
107. Petersen, R. C., The contradictory biological behavior of humic substances in the aquatic environment. In *Humic substances in the aquatic and terrestrial environment*, Springer: 1991; pp 367-390.
108. Gerke, J., Concepts and misconceptions of humic substances as the stable part of soil organic matter: A review. *Agronomy* **2018**, *8* (5), 76.
109. Carter, C. W.; Suffet, I. H., Binding of DDT to dissolved humic materials. *Environ Sci Technol* **1982**, *16* (11), 735-740.
110. Mahdavi, M.; Jafari, J., Environmental risks due to application of sewage sludge in farmlands. *Ozean J Appl Sci* **2010**, *3* (2), 303-311.
111. Aiken, G. R.; Hsu-Kim, H.; Ryan, J. N., Influence of dissolved organic matter on the environmental fate of metals, nanoparticles, and colloids. ACS Publications: 2011.

112. Bertilsson, S.; Tranvik, L., Photochemical transformation of dissolved organic matter in lakes. *J Limnology Oceanography* **2000**, *45* (4), 753-762.
113. Fellman, J. B.; Petrone, K. C.; Grierson, P. F., Source, biogeochemical cycling, and fluorescence characteristics of dissolved organic matter in an agro-urban estuary. *J Limnol Oceanograph* **2011**, *56* (1), 243-256.
114. McKnight, D. M.; Boyer, E. W.; Westerhoff, P. K.; Doran, P. T.; Kulbe, T.; Andersen, D. T., Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *J Limnology Oceanography* **2001**, *46* (1), 38-48.
115. Tsuda, K.; Mori, H.; Asakawa, D.; Yanagi, Y.; Kodama, H.; Nagao, S.; Yonebayashi, K.; Fujitake, N., Characterization and grouping of aquatic fulvic acids isolated from clear-water rivers and lakes in Japan. *Water Res* **2010**, *44* (13), 3837-3846.
116. Mao, J.-D.; Schmidt-Rohr, K., Accurate quantification of aromaticity and nonprotonated aromatic carbon fraction in natural organic matter by ^{13}C solid-state nuclear magnetic resonance. *Environmental science & technology* **2004**, *38* (9), 2680-2684.
117. Ghabbour, E. A.; Davies, G., Spectrophotometric analysis of fulvic acid solutions — A second look. *Annals of Environmental Science* **2009**.
118. Zark, M.; Dittmar, T., Universal molecular structures in natural dissolved organic matter. *Nature Comm* **2018**, *9* (1), 3178.
119. Avneri-Katz, S.; Young, R. B.; McKenna, A. M.; Chen, H.; Corilo, Y. E.; Polubesova, T.; Borch, T.; Chefetz, B., Adsorptive fractionation of dissolved organic matter (DOM) by mineral soil: Macroscale approach and molecular insight. *J Org Geochem* **2017**, *103*, 113-124.
120. Herzsprung, P.; Osterloh, K.; von Tümpling, W.; Harir, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Meissner, R.; Bernsdorf, S.; Friese, K., Differences in DOM of rewetted and natural peatlands—Results from high-field FT-ICR-MS and bulk optical parameters. *Sci Total Environ* **2017**, *586*, 770-781.
121. An, Y.; Xu, J.; Feng, L.; Zhang, X.; Liu, Y.; Kang, S.; Jiang, B.; Liao, Y., Molecular characterization of organic aerosol in the Himalayas: insight from ultra-high-resolution mass spectrometry. *J Atmos Chem Phys* **2019**, *19* (2), 1115-1128.
122. Stubbins, A.; Hood, E.; Raymond, P. A.; Aiken, G. R.; Sleighter, R. L.; Hernes, P. J.; Butman, D.; Hatcher, P. G.; Striegl, R. G.; Schuster, P., Anthropogenic aerosols as a source of ancient dissolved organic matter in glaciers. *Nature Geoscience* **2012**, *5* (3), 198-201.
123. Spencer, R. G.; Mann, P. J.; Dittmar, T.; Eglinton, T. I.; McIntyre, C.; Holmes, R. M.; Zimov, N.; Stubbins, A., Detecting the signature of permafrost thaw in Arctic rivers. *Geophysical Research Letters* **2015**, *42* (8), 2830-2835.
124. 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.

125. Kujawinski, E. B.; Behn, M. D., Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Analytical Chemistry* **2006**, *78* (13), 4363-4373.
126. Brown, A.; McKnight, D. M.; Chin, Y.-P.; Roberts, E. C.; Uhle, M., Chemical characterization of dissolved organic material in Pony Lake, a saline coastal pond in Antarctica. *Marine Chem* **2004**, *89* (1-4), 327-337.
127. Niu, X.-Z.; Harir, M.; Schmitt-Kopplin, P.; Croué, J.-P., Characterisation of dissolved organic matter using Fourier-transform ion cyclotron resonance mass spectrometry: Type-specific unique signatures and implications for reactivity. *Sci Total Environ* **2018**, *644*, 68-76.
128. Mississippi river: A cultural treasure.
<https://www.americanrivers.org/river/mississippi-river/> (accessed June 01).
129. 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. *Analytical chemistry* **2003**, *75* (6), 1275-1284.
130. Hiraoka, K.; Murata, K.; Kudaka, I., Species-selectivity effects in the production of electrospray ions. *Rapid communications in mass spectrometry* **1993**, *7* (5), 363-373.
131. Herzsprung, P.; Hertkorn, N.; von Tümppling, 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. *J Anal Bioanal Chem* **2014**, *406* (30), 7977-7987.
132. Herzsprung, P.; Hertkorn, N.; von Tümppling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P., Molecular formula assignment for dissolved organic matter (DOM) using high-field FT-ICR-MS: chemical perspective and validation of sulphur-rich organic components (CHOS) in pit lake samples. *J Anal Bioanal Chem* **2016**, *408* (10), 2461-2469.
133. 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.
134. Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C., 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer: a national resource for ultrahigh resolution mass analysis. *Journal of the American Society for Mass Spectrometry* **2015**, *26* (9), 1626-1632.
135. He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Hendrickson, C. L.; Marshall, A. G., Analysis of monoclonal antibodies in human serum as a model for clinical monoclonal gammopathy by use of 21 tesla FT-ICR top-down and middle-down MS/MS. *Journal of The American Society for Mass Spectrometry* **2017**, *28* (5), 827-838.

136. He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Dasari, S.; Dispenzieri, A.; Hendrickson, C. L.; Marshall, A. G., Classification of plasma cell disorders by 21 Tesla Fourier transform ion cyclotron resonance top-down and middle-down MS/MS analysis of monoclonal immunoglobulin light chains in human serum. *Analytical chemistry* **2019**, *91* (5), 3263-3269.
137. Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E.; Early, L.; Mullen, C.; Dunyach, J.-J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J., Front-end electron transfer dissociation coupled to a 21 Tesla FT-ICR mass spectrometer for intact protein sequence analysis. *Journal of the American Society for Mass Spectrometry* **2017**, *28* (9), 1787-1795.
138. Shaw, J. B.; Gorshkov, M. V.; Wu, Q.; Paša-Tolić, L., High speed intact protein characterization using 4X frequency multiplication, ion trap harmonization, and 21 Tesla FTICR-MS. *Analytical chemistry* **2018**, *90* (9), 5557-5562.
139. He, L.; Weisbrod, C. R.; Marshall, A. G., Protein de novo sequencing by top-down and middle-down MS/MS: limitations imposed by mass measurement accuracy and gaps in sequence coverage. *International Journal of Mass Spectrometry* **2018**, *427*, 107-113.
140. He, L.; Rockwood, A. L.; Agarwal, A. M.; Anderson, L. C.; Weisbrod, C. R.; Hendrickson, C. L.; Marshall, A. G., Diagnosis of hemoglobinopathy and β -thalassemia by 21 tesla Fourier transform ion cyclotron resonance mass spectrometry and tandem mass spectrometry of hemoglobin from blood. *Clinical chemistry* **2019**, *65* (8), 986-994.
141. Weisbrod, C. R.; Anderson, L. C.; Greer, J. B.; DeHart, C. J.; Hendrickson, C. L., Increased Single-Spectrum Top-Down Protein Sequence Coverage in Trapping Mass Spectrometers with Chimeric Ion Loading. *Analytical Chemistry* **2020**, *92* (18), 12193-12200.
142. Walker, L. R.; Tfaily, M. M.; Shaw, J. B.; Hess, N. J.; Paša-Tolić, L.; Koppelaar, D. W., Unambiguous identification and discovery of bacterial siderophores by direct injection 21 Tesla Fourier transform ion cyclotron resonance mass spectrometry. *Metallomics* **2017**, *9* (1), 82-92.
143. Rowland, S.; Rodgers, R.; Smith, D.; Blakney, G.; Corilo, Y. E.; Hendrickson, C. In *Dual column on-line liquid chromatography coupled to ultrahigh resolution 21 T FT-ICR mass spectrometry for the determination of molecular-level changes in petroleum samples*, ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY, AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA: 2018.
144. Marshall, A.; Chacon, M.; Corilo, Y. E.; Hendrickson, C.; Krajewski, L.; Niles, S.; Putman, J.; Rowland, S.; Smith, D.; Ware, R. In *Widening the window for environmental analysis: On-line HPLC monitored by 21 tesla Fourier transform ion cyclotron resonance mass spectrometry*, ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY, AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA: 2019.

145. Stopka, S. A.; Samarah, L. Z.; Shaw, J. B.; Liyu, A. V.; Veličković, D. a.; Agtuca, B. J.; Kukolj, C.; Koppenaal, D. W.; Stacey, G.; Paša-Tolić, L., Ambient metabolic profiling and imaging of biological samples with ultrahigh molecular resolution using laser ablation electrospray ionization 21 tesla FTICR mass spectrometry. *Analytical chemistry* **2019**, *91* (8), 5028-5035.
146. Bowman, A. P.; Blakney, G. T.; Hendrickson, C. L.; Ellis, S. R.; Heeren, R. M.; Smith, D. F., Ultra-high mass resolving power, mass accuracy, and dynamic range MALDI mass spectrometry imaging by 21-T FT-ICR MS. *Analytical chemistry* **2020**, *92* (4), 3133-3142.
147. Rockström, J.; Steffen, W.; Noone, K.; Persson, Å.; Chapin, F. S.; Lambin, E. F.; Lenton, T. M.; Scheffer, M.; Folke, C.; Schellnhuber, H. J., A safe operating space for humanity. *nature* **2009**, *461* (7263), 472-475.
148. Zhou, S.; Collier, S.; Jaffe, D. A.; Briggs, N. L.; Hee, J.; Sedlacek III, A. J.; Kleinman, L.; Onasch, T. B.; Zhang, Q., Regional influence of wildfires on aerosol chemistry in the western US and insights into atmospheric aging of biomass burning organic aerosol. *Atmos. Chem. Phys* **2017**, *17* (3), 2477-2493.
149. Garofalo, L. A.; Pothier, M. A.; Levin, E. J.; Campos, T.; Kreidenweis, S. M.; Farmer, D. K., Emission and evolution of submicron organic aerosol in smoke from wildfires in the western United States. *ACS Earth and Space Chemistry* **2019**, *3* (7), 1237-1247.
150. Gunsch, M. J.; May, N. W.; Wen, M.; Bottenus, C. L.; Gardner, D. J.; VanReken, T. M.; Bertman, S. B.; Hopke, P. K.; Ault, A. P.; Pratt, K. A., Ubiquitous influence of wildfire emissions and secondary organic aerosol on summertime atmospheric aerosol in the forested Great Lakes region. *Atmospheric Chemistry and Physics (Online)* **2018**, *18* (PNNL-SA-136284).
151. Palm, B. B.; Peng, Q.; Fredrickson, C. D.; Lee, B. H.; Garofalo, L. A.; Pothier, M. A.; Kreidenweis, S. M.; Farmer, D. K.; Pokhrel, R. P.; Shen, Y., Quantification of organic aerosol and brown carbon evolution in fresh wildfire plumes. *Proceedings of the National Academy of Sciences* **2020**, *117* (47), 29469-29477.
152. Li, Y.; Day, D. A.; Stark, H.; Jimenez, J. L.; Shiraiwa, M., Predictions of the glass transition temperature and viscosity of organic aerosols from volatility distributions. *Atmospheric Chemistry and Physics* **2020**, *20* (13), 8103-8122.
153. Zhurov, K. O.; Kozhinov, A. N.; Fornelli, L.; Tsybin, Y. O., Distinguishing analyte from noise components in mass spectra of complex samples: where to cut the noise? *Analytical chemistry* **2014**, *86* (7), 3308-3316.
154. Willoughby, A.; Wozniak, A.; Hatcher, P. G., A molecular-level approach for characterizing water-insoluble components of ambient organic aerosol particulates using ultrahigh-resolution mass spectrometry. *Atmospheric Chemistry and Physics* **2014**, *14* (18).
155. Dzepina, K.; Mazzoleni, C.; Fialho, P.; China, S.; Zhang, B.; Owen, R. C.; Helmig, D.; Hueber, J.; Kumar, S.; Perlinger, J. A., Molecular characterization of free tropospheric aerosol collected at the Pico Mountain Observatory: a case

- study with a long-range transported biomass burning plume. *Atmospheric Chemistry and Physics* **2015**, *15* (9), 5047-5068.
156. Laskin, A.; Smith, J. S.; Laskin, J., Molecular characterization of nitrogen-containing organic compounds in biomass burning aerosols using high-resolution mass spectrometry. *Environmental science & technology* **2009**, *43* (10), 3764-3771.
157. Simoneit, B. R.; Elias, V. O., Detecting organic tracers from biomass burning in the atmosphere. *Marine Pollution Bulletin* **2001**, *42* (10), 805-810.
158. Simoneit, B. R., Biomass burning—a review of organic tracers for smoke from incomplete combustion. *Applied Geochemistry* **2002**, *17* (3), 129-162.
159. Bai, X.; Johnston, P.; Sadula, S.; Brown, R. C., Role of levoglucosan physiochemistry in cellulose pyrolysis. *Journal of Analytical and Applied Pyrolysis* **2013**, *99*, 58-65.
160. Schmitt-Kopplin, P.; Gelencser, A.; Dabek-Zlotorzynska, E.; Kiss, G.; Hertkorn, N.; Harir, M.; Hong, Y.; Gebefügi, I., Analysis of the unresolved organic fraction in atmospheric aerosols with ultrahigh-resolution mass spectrometry and nuclear magnetic resonance spectroscopy: organosulfates as photochemical smog constituents. *Analytical Chemistry* **2010**, *82* (19), 8017-8026.
161. 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.
162. Yu, L.; Ma, X.; Gao, H.; Zong, H.; Yao, X.; Lin, Z.; Zhang, Z.; Zhang, C.; Yao, X.; Zhang, Z., Distribution and source identification of nitrogen and phosphorus in aerosols in the Qinhuangdao coast, north China. *Atmospheric Environment* **2020**, 117475.
163. Sundarambal, P.; Balasubramanian, R.; Tkalic, P.; He, J., Impact of biomass burning on ocean water quality in Southeast Asia through atmospheric deposition: field observations. *Atmospheric Chemistry and Physics* **2010**, *10* (23), 11323-11336.
164. Prospero, J.; Barrett, K.; Church, T.; Dentener, F.; Duce, R.; Galloway, J.; Levy, H.; Moody, J.; Quinn, P., Atmospheric deposition of nutrients to the North Atlantic Basin. In *Nitrogen Cycling in the North Atlantic Ocean and its Watersheds*, Springer: 1996; pp 27-73.
165. Tipping, E.; Benham, S.; Boyle, J.; Crow, P.; Davies, J.; Fischer, U.; Guyatt, H.; Helliwell, R.; Jackson-Blake, L.; Lawlor, A. J., Atmospheric deposition of phosphorus to land and freshwater. *Environmental Science: Processes & Impacts* **2014**, *16* (7), 1608-1617.
166. Mahowald, N.; Jickells, T. D.; Baker, A. R.; Artaxo, P.; Benitez-Nelson, C. R.; Bergametti, G.; Bond, T. C.; Chen, Y.; Cohen, D. D.; Herut, B., Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. *Global biogeochemical cycles* **2008**, *22* (4).
167. Painter, T. A.; Markiewicz, W. D.; Miller, J. R.; Bole, S. T.; Dixon, I. R.; Cantrell, K. R.; Kenney, S. J.; Trowell, A. J.; Kim, D. L.; Lee, B. S., Requirements

- and conceptual superconducting magnet design for a 21 T fourier transform ion cyclotron resonance mass spectrometer. *IEEE transactions on applied superconductivity* **2006**, *16* (2), 945-948.
168. Choi, Y. S.; Painter, T. A.; Kim, D. L.; Lee, B. S.; Yang, H. S.; Miller, J. R., Conceptual design of current leads for a 21 T FT-ICR magnet system. *IEEE transactions on applied superconductivity* **2007**, *17* (2), 2228-2231.
169. Choi, Y. S.; Kim, D. L.; Painter, T. A.; Markiewicz, W. D.; Lee, B. S.; Yang, H. S.; Yoo, J. S., Closed-loop cryogenic cooling for a 21 T FT-ICR magnet system. *IEEE transactions on applied superconductivity* **2008**, *18* (2), 1471-1474.
170. Choi, Y. S.; Painter, T. A.; Kim, D. L.; Bole, S. T.; Adkins, T., Semi-retractable current leads for a 21 T FT-ICR magnet system. *IEEE transactions on applied superconductivity* **2008**, *18* (2), 1431-1434.
171. Van Eck, N. J.; Waltman, L., Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics* **2010**, *84* (2), 523-538.
172. Ansmann, A.; Baars, H.; Chudnovsky, A.; Mattis, I.; Veselovskii, I.; Haarig, M.; Seifert, P.; Engelmann, R.; Wandinger, U., Extreme levels of Canadian wildfire smoke in the stratosphere over central Europe on 21–22 August 2017. *Atmospheric Chemistry and Physics* **2018**, *18* (16), 11831-11845.
173. Ndalila, M. N.; Williamson, G. J.; Fox-Hughes, P.; Sharples, J.; Bowman, D. M., Evolution of a pyrocumulonimbus event associated with an extreme wildfire in Tasmania, Australia. *Natural Hazards and Earth System Sciences* **2020**, *20* (5), 1497-1511.
174. Khaykin, S.; Godin-Beekmann, S.; Hauchecorne, A.; Pelon, J.; Ravetta, F.; Keckhut, P., Stratospheric smoke with unprecedentedly high backscatter observed by lidars above southern France. *Geophysical Research Letters* **2018**, *45* (3), 1639-1646.
175. Fromm, M.; Bevilacqua, R.; Servranckx, R.; Rosen, J.; Thayer, J. P.; Herman, J.; Larko, D., Pyro-cumulonimbus injection of smoke to the stratosphere: Observations and impact of a super blowup in northwestern Canada on 3–4 August 1998. *Journal of Geophysical Research: Atmospheres* **2005**, *110* (D8).
176. Yu, P.; Toon, O. B.; Bardeen, C. G.; Zhu, Y.; Rosenlof, K. H.; Portmann, R. W.; Thornberry, T. D.; Gao, R.-S.; Davis, S. M.; Wolf, E. T., Black carbon lofted wildfire smoke high into the stratosphere to form a persistent plume. *Science* **2019**, *365* (6453), 587-590.
177. Yuan, Q.; Xu, J.; Liu, L.; Zhang, A.; Liu, Y.; Zhang, J.; Wan, X.; Li, M.; Qin, K.; Cong, Z., Evidence for Large Amounts of Brown Carbonaceous Tarballs in the Himalayan Atmosphere. *Environmental Science & Technology Letters* **2020**, *8* (1), 16-23.
178. O'Brien, R. E.; Neu, A.; Epstein, S. A.; MacMillan, A. C.; Wang, B.; Kelly, S. T.; Nizkorodov, S. A.; Laskin, A.; Moffet, R. C.; Gilles, M. K., Physical

- properties of ambient and laboratory-generated secondary organic aerosol. *Geophysical Research Letters* **2014**, *41* (12), 4347-4353.
179. Järvinen, E.; Ignatius, K.; Nichman, L.; Kristensen, T. B.; Fuchs, C.; Hoyle, C. R.; Höppel, N.; Corbin, J. C.; Craven, J.; Duplissy, J., Observation of viscosity transition in α -pinene secondary organic aerosol. *Atmospheric Chemistry and Physics* **2016**, *16* (7), 4423-4438.
180. Tóth, A.; Hoffer, A.; Nyirő-Kósa, I.; Pósfai, M.; Gelencsér, A., Atmospheric tar balls: aged primary droplets from biomass burning? *Atmospheric Chemistry and Physics* **2014**, *14* (13), 6669-6675.
181. Tóth, Á.; Hoffer, A.; Pósfai, M.; Ajtai, T.; Kónya, Z.; Blazsó, M.; Czégény, Z.; Kiss, G.; Bozóki, Z.; Gelencsér, A., Chemical characterization of laboratory-generated tar ball particles. *Atmospheric Chemistry and Physics* **2018**, *18* (14), 10407-10418.
182. Shrivastava, M.; Cappa, C. D.; Fan, J.; Goldstein, A. H.; Guenther, A. B.; Jimenez, J. L.; Kuang, C.; Laskin, A.; Martin, S. T.; Ng, N. L., Recent advances in understanding secondary organic aerosol: Implications for global climate forcing. *Reviews of Geophysics* **2017**, *55* (2), 509-559.
183. McNeill, V. F., Aqueous organic chemistry in the atmosphere: Sources and chemical processing of organic aerosols. ACS Publications: 2015.
184. McNeill, V. F.; Woo, J. L.; Kim, D. D.; Schwier, A. N.; Wannell, N. J.; Sumner, A. J.; Barakat, J. M., Aqueous-phase secondary organic aerosol and organosulfate formation in atmospheric aerosols: a modeling study. *Environmental science & technology* **2012**, *46* (15), 8075-8081.
185. Hoffer, A.; Tóth, A.; Nyirő-Kósa, I.; Pósfai, M.; Gelencsér, A., Light absorption properties of laboratory-generated tar ball particles. *Atmospheric Chemistry and Physics* **2016**, *16* (1), 239-246.
186. Tsigaridis, K.; Kanakidou, M., Secondary organic aerosol importance in the future atmosphere. *Atmospheric Environment* **2007**, *41* (22), 4682-4692.
187. Aiken, A. C.; Decarlo, P. F.; Kroll, J. H.; Worsnop, D. R.; Huffman, J. A.; Docherty, K. S.; Ulbrich, I. M.; Mohr, C.; Kimmel, J. R.; Sueper, D., O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high-resolution time-of-flight aerosol mass spectrometry. *Environmental science & technology* **2008**, *42* (12), 4478-4485.
188. DeCarlo, P. F.; Kimmel, J. R.; Trimborn, A.; Northway, M. J.; Jayne, J. T.; Aiken, A. C.; Gonin, M.; Fuhrer, K.; Horvath, T.; Docherty, K. S., Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Analytical chemistry* **2006**, *78* (24), 8281-8289.
189. Reemtsma, T.; These, A.; Springer, A.; Linscheid, M., Differences in the molecular composition of fulvic acid size fractions detected by size-exclusion chromatography-on line Fourier transform ion cyclotron resonance (FTICR-) mass spectrometry. *Water research* **2008**, *42* (1-2), 63-72.
190. Nguyen, T. B.; Roach, P. J.; Laskin, J.; Laskin, A.; Nizkorodov, S. A., Effect of humidity on the composition of isoprene photooxidation secondary

- organic aerosol. *Atmospheric Chemistry and Physics* **2011**, *11* (14), 6931-6944.
191. Eddingsaas, N.; Loza, C.; Yee, L.; Chan, M.; Schilling, K.; Chhabra, P.; Seinfeld, J.; Wennberg, P., α -pinene photooxidation under controlled chemical conditions–Part 2: SOA yield and composition in low-and high-NO_x environments. *Atmospheric Chemistry and Physics* **2012**, *12* (16), 7413-7427.
192. McVay, R. C.; Zhang, X.; Aumont, B.; Valorso, R.; Camredon, M.; La, Y. S.; Wennberg, P. O.; Seinfeld, J. H., SOA formation from the photooxidation of α -pinene: systematic exploration of the simulation of chamber data. *Atmospheric Chemistry and Physics* **2016**, *16* (5), 2785-2802.
193. Henze, D. K.; Seinfeld, J. H., Global secondary organic aerosol from isoprene oxidation. *Geophysical Research Letters* **2006**, *33* (9).
194. Kroll, J. H.; Ng, N. L.; Murphy, S. M.; Flagan, R. C.; Seinfeld, J. H., Secondary organic aerosol formation from isoprene photooxidation. *Environmental science & technology* **2006**, *40* (6), 1869-1877.
195. Boyd, C.; Sanchez, J.; Xu, L.; Eugene, A. J.; Nah, T.; Tuet, W.; Guzman, M. I.; Ng, N., Secondary organic aerosol formation from the β -pinene+ NO₃ system: effect of humidity and peroxy radical fate. *Atmospheric Chemistry and Physics* **2015**, *15* (13), 7497-7522.
196. Winterhalter, R.; Dingenen, R. V.; Larsen, B.; Jensen, N.; Hjorth, J., LC-MS analysis of aerosol particles from the oxidation of α -pinene by ozone and OH-radicals. *Atmospheric Chemistry and Physics Discussions* **2003**, *3* (1), 1-39.