

ABSTRACT

Title of Document: A DEUTERIUM LABELING METHOD FOR THE CHARACTERIZATION OF (CHROMOPHORIC) DISSOLVED ORGANIC MATTER USING ULTRAHIGH RESOLUTION ELECTROSPRAY IONIZATION MASS SPECTROMETRY

Daniel Robert Baluha, Ph.D., 2015

Directed By: Professor Neil Blough
Department of Chemistry and Biochemistry

Dissolved organic matter (DOM) is a complex ensemble of naturally occurring organic compounds found in virtually all aquatic environments. The overwhelming diversity of DOM makes it extremely difficult to understand the relationship between its bulk physicochemical properties and its molecular structure and composition.

This dissertation describes the development of a novel method to identify ketone/aldehyde-containing species within DOM, which are known to contribute substantially to the ultraviolet/visible (UV-vis) absorption and emission of chromophoric DOM. In this method, an aqueous sample is treated with sodium borodeuteride (NaBD_4) and is analyzed via ultrahigh resolution electrospray ionization (ESI) mass spectrometry. Ketone/aldehyde-containing species (at mass m) in the untreated sample are identified by searching the mass spectrum of the reduced sample for peaks corresponding to deuterated

derivatives (at mass $m+3.021927n$). Initial experiments demonstrated that this method reliably discriminates among mass spectral peaks in an untreated DOM sample that comprise species with zero, one, and/or two reducible moieties.

The reactivity and optical properties of reducible species within Suwannee River fulvic acid (SRFA) were studied by treating an aqueous sample with several amounts of NaBD₄. This study demonstrated that most species with at least one ketone/aldehyde moiety were reduced a single time under low [NaBD₄], while higher [NaBD₄] resulted primarily in additional reductions on multi-ketone/aldehyde species. Furthermore, the changes in UV-vis absorption and emission of the reduced aliquots relative to that of the untreated were correlated with the number of ketone/aldehyde-containing species reduced and identified by this method.

The fully developed protocol was used to compare DOM extracted from several aquatic environments. Two pools of ketone/aldehyde-containing species were tentatively identified: A terrestrially-produced group of lignin/tannin-derivatives and a microbially-produced group of carboxyl-rich alicyclic molecules. While the first pool has previously been shown to contribute substantially to the absorption/emission of chromophoric DOM, the second pool most likely would not. The mass labeling method developed here revealed compositional features that are not observable by common ESI mass spectrometric analyses and may serve as a useful way to link the physicochemical properties of DOM to its structure and composition.

A DEUTERIUM LABELING METHOD FOR THE CHARACTERIZATION OF
(CHROMOPHORIC) DISSOLVED ORGANIC MATTER USING ULTRAHIGH
RESOLUTION ELECTROSPRAY IONIZATION MASS SPECTROMETRY

By

Daniel Robert Baluha

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Neil Blough, Chair
Professor Daniel Falvey
Professor Catherine Fenselau
Professor Michael Gonsior
Professor Russell Dickerson (Dean's Representative)

© Copyright by
Daniel Robert Baluha
2015

Acknowledgements

I thank my advisor, Dr. Neil Blough for his support and guidance that he has given me throughout the last four and a half years. I thank all of my (past and present) group members: Dr. Rossana Del Vecchio, Dr. Andrea Andrew, Dr. Yi Zhang, Dr. Kelli Golanoski, Dr. Lynne Heighton, Tara Schendorf, Joe Devlin, Kevin Koech, Carmen Cartisano, Marla Bianca, and Abby Katilas. I thank my dissertation and candidacy committee members for their time and for their helpful feedback: Dr. Catherine Fenselau, Dr. Daniel Falvey, Dr. Yu Huang Wang, Dr. Russell Dickerson, and Dr. Michael Gonsior. I owe Dr. Gonsior an additional debt of gratitude for the collaboration and assistance on many aspects of this research. I also thank Dr. Melissa Kido-Soule and Dr. Krista Longnecker at the Woods Hole Oceanographic institution and Dr. Mourad Harir at the Helmholtz Center for Environmental Sciences for their assistance with FT-MS data acquisition and processing. I thank Dr. Earle Stone for playing a crucial role in my development as a science educator, and I thank my former chemistry professors (and future CSCU colleagues!) at Central Connecticut State University for all the mentorship and support that prepared me for the challenges of a Ph.D. program. Finally, I thank my wife, Christine, for all the love, support, and patience that made getting through all of this possible!

Table of Contents

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| Acknowledgements | ii |
| Table of Contents | iii |
| List of Tables | v |
| List of Figures | vii |
| List of Abbreviations | xii |
| Chapter 1: Aquatic natural organic matter (NOM): Definitions, sub-fractions, and environmental significance | 1 |
| 1.1. A broad definition of natural organic matter (NOM) | 1 |
| 1.2. Description of chromophoric dissolved organic matter (CDOM) | 4 |
| 1.3. Physicochemical justification of the optical properties of CDOM..... | 9 |
| Chapter 2: Advanced characterization of DOM using ultrahigh resolution mass spectrometry (an overview) | 13 |
| 2.1. Analyzing the structure and composition of DOM..... | 13 |
| 2.2. Instrumentation for mass spectrometric analysis of DOM | 15 |
| 2.3. Analysis, interpretation, and representation of mass spectrometric data..... | 24 |
| Chapter 3: Development of a mass labeling method for identifying ketone/aldehyde-containing species in DOM | 34 |
| 3.1. Introduction..... | 34 |
| 3.2. Materials and Methods | 36 |
| 3.3. Results and Discussion | 39 |
| 3.4. Conclusions..... | 54 |
| Chapter 4: Reactivity of ketone/aldehyde-containing species in Suwannee River fulvic acid and their contribution to bulk optical properties | 56 |
| 4.1. Introduction..... | 56 |
| 4.2. Materials and Methods | 57 |
| 4.3. Results and Discussion | 65 |
| 4.4. Conclusions..... | 82 |
| Chapter 5: Comparison of the presence and composition of ketone/aldehyde-containing species in DOM from various aquatic environments | 83 |
| 5.1. Introduction..... | 83 |
| 5.2. Materials and Methods | 85 |
| 5.3. Results and Discussion | 89 |
| 5.4. Conclusions..... | 103 |
| Chapter 6: Conclusions and Future Work | 105 |
| Appendix 1: Internal calibrants and extraction blanks | 108 |
| Appendix 2: Full scan negative ion mass spectra | 113 |

| | |
|------------------------------------------------------------------------------------------------------------------|------------|
| Appendix 3: Peak and molecular formula lists of UNT SRFA at selected nominal masses (for Chapter 3) | 116 |
| Appendix 4: Reduction and mass spectrometric analysis of acetovanillone | 119 |
| Appendix 5: Molecular formula assignment details for Chapter 4 | 121 |
| Appendix 6: Supporting optical data for Chapter 4 | 127 |
| Appendix 7: Sample collection details for C₁₈ extracts | 132 |
| Appendix 8: SRFA and PLFA at 311 <i>m/z</i> (for Chapter 5) | 134 |
| Appendix 9: Custom MATLAB functions and description of calculations | 135 |
| A9.1. Chapter 3 analysis code | 135 |
| A9.2. Blank subtraction and excision of ions with multiple charges | 140 |
| A9.3. Molecular formula assignment (“brute-force”) | 143 |
| A9.4. Molecular formula assignment (“low-mass moiety”)..... | 148 |
| A9.5. Ketone/aldehyde identification | 153 |
| A9.6. Peak/formula list subset selection..... | 157 |
| A9.7. Comparison of peaks and/or molecular formulae..... | 158 |
| A9.8. UV-visible absorption and emission analysis..... | 163 |
| A9.9. Calculations of average molecular weight, H/C and O/C molar ratios, and mass percentages. | 172 |
| Bibliography | 174 |

List of Tables

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 1.1. Common fluorophores identified in CDOM. | 9 |
| Table 2.1. Mass spacing patterns observed in the ESI mass spectra of DOM. | 29 |
| Table 3.1. Percentage of peaks in the mass spectra of untreated and borodeuteride-reduced SRFA and DEUB identified as $m+3.021927n$ ($n = 1, 2,$ and/or 3) masses. | 48 |
| Table 3.2. Percentage of peaks in the mass spectrum of untreated SRFA and DEUB identified as comprising ketone and/or aldehyde-containing species (based on a search for M+3.0219 and M+6.0438 masses in the corresponding borodeuteride-reduced mass spectrum. | 50 |
| Table 4.1. Descriptions, pH, and absorption data for untreated (UNT) and all borodeuteride-reduced (BDR) samples. | 59 |
| Table 4.2. Distribution of peaks in the full (50 – 2000) and analyzed (200 – 600) m/z ranges. | 67 |
| Table 4.3. Numbers of non, singly, and/or doubly reduced species in UNT (between 200 and 600 m/z) following a 2.3-fold (BDR-1) and 20-fold (BDR-3) addition of NaBD₄. Regions above and below the (shaded) diagonal correspond to species which decreased or increased in reduction, respectively, from BDR-1 to BDR-3.” | 73 |
| Table 4.4 Number of formulae in UNT at each m/z range identified as non-, singly-, and/or doubly-reduced. | 75 |
| Table 5.1. Mass spectral peak distribution of all samples in the full (50 – 2000) and analyzed (250 – 550) m/z ranges. | 91 |
| Table 5.2. Summary of molecular formulae identified in all untreated samples in the 250 to 550 m/z range. | 93 |
| Table A1.1. m/z and corresponding molecular formula (of [M-H]⁻ ions) used for post-acquisition calibration of the 7 T ESI FT-ICR mass spectra (calibration performed by Dr. Krista Longnecker, Woods Hole Oceanographic Institution). | 108 |
| Table A1.2. m/z and corresponding molecular formula (of [M-H]⁻ ions) used for post-acquisition calibration of the 12 T ESI FT-ICR mass spectra (calibration performed by Dr. Michael Gonsior, University of Maryland Center for Environmental Sciences). | 109 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table A1.3. <i>m/z</i> and tentative assignments of peaks present in the extraction blank shown in Figure A1.1 which were removed from the mass spectra of SRFA, EACR, EAUW, EADO, and EASO (untreated and reduced). | 110 |
| Table A1.4. <i>m/z</i> and tentative assignments of peaks present in the extraction blank shown in Figure A1.2 which were removed from the mass spectra of PLFA, DELB, and DERV (untreated and reduced). | 112 |
| Table A2.1. Number of peaks and total ion count (TIC) of the full scanned range (200 – 1000 <i>m/z</i>) and the analyzed range (200 – 600 <i>m/z</i> shaded rows) for all 7 T ESI FT-ICR mass spectra. | 114 |
| Table A3.1. List of peaks, relative intensities, and molecular formulae at 467 <i>m/z</i> in the 7 T ESI FT-ICR mass spectra of SRFA (highlighted: Peaks A – E in Figure 3.4). | 116 |
| Table A3.2. List of peaks, relative intensities, and molecular formulae at 469 <i>m/z</i> in the 7 T ESI FT-ICR mass spectra of SRFA (highlighted: Peaks A’ – E’ in Figure 3.4). | 117 |
| Table A3.3. List of peaks, relative intensities, and molecular formulae at 470 <i>m/z</i> in the 7 T ESI FT-ICR mass spectra of SRFA (highlighted: Peaks A’’ – E’’ in Figure 3.4). | 118 |
| Table A5.1. Molecular formula assignment details for UNT in the 200 to 400 <i>m/z</i> and 400 to 500 <i>m/z</i> ranges using 0.2 and 0.3 ppm error limits, respectively and various maximum allowances of N and S. | 122 |
| Table A5.2. Molecular formula for peaks at 311 and 314 <i>m/z</i> in SRFA UNT. Intensities are normalized to the most intense peak the 295 to 314 <i>m/z</i> range. | 124 |
| Table A5.3. Molecular formula for peaks at 311 and 314 <i>m/z</i> in SRFA BDR-1. Intensities are normalized to the most intense peak the 295 to 314 <i>m/z</i> range. | 125 |
| Table A5.4. Molecular formula for peaks at 311 and 314 <i>m/z</i> in SRFA BDR-4. Intensities are normalized to the most intense peak the 295 to 314 <i>m/z</i> range. | 126 |
| Table A7.1. Sample collection date and location for all C₁₈ extracts.^a | 133 |

List of Figures

- Figure 1.1.** Two major categories of aquatic NOM and their primary sources and sinks. .2
- Figure 1.2.** Common separation processes used for isolating various sub-fractions of NOM from a water sample collected from an aquatic environment.3
- Figure 1.3.** Absorption, wavelength of maximum emission, and luminescence quantum yield versus excitation wavelength for Suwannee River fulvic acid (SRFA). Adapted from Del Vecchio and Blough, 2004.²²6
- Figure 1.4.** Fluorescence EEM of two CDOM samples, depicted using 3-dimensional wireframe (left) and contour plot (right) renderings. Adapted from Coble et al., 1996.²⁹8
- Figure 1.5.** Schematic of an absorption spectrum of CDOM before (left) and after (right) treatment by NaBH₄ showing the disruption of charge transfer interactions (by reduction of acceptor moieties) and the corresponding loss of longer wavelength (>300 nm) absorption bands. 11
- Figure 2.1.** Early hypothetical structures for (A) fulvic and (B) humic acids from seawater (structures from Zafiriou et al., 1984⁴¹) and (C) representations of various classes of chemical compounds (lipids, peptides, cellulose, condensed hydrocarbons, lignin, condensed tannins) that may exist in a DOM sample. 14
- Figure 2.2.** Schematic of an ESI interface operated in negative ion mode..... 15
- Figure 2.3.** ESI mass spectra of various DOM samples collected using a range of operating resolving powers: (A) DOM extracted from the North Pacific Ocean using 12 T FT-ICR with a resolving power of > 500,000 (adapted from Sleighter et al., 2012⁵⁸) and humic acid from Mount Rainier using (B) 7 T FT-ICR and (C) quadrupole time-of-flight mass analyzers (adapted from Kujawinski et al., 2002⁵⁹). 19
- Figure 2.4.** Illustrations of (A) excitation of ions with small initial incoherent cyclotron orbits to larger and more detectable coherent motions and (B) subsequent detection of the currents in the ICR cell resulting from the coherent motions. (C) A measured FID signal and the corresponding mass spectrum obtained by Fourier transformation of the FID (adapted from Amster, 2002⁶⁴). 22
- Figure 2.5.** ESI FT-ICR mass spectra of DOM extracted from (A) the Dismal Swamp and (B) the Atlantic Ocean off the coast of northern Virginia (adapted from Sleighter and Hatcher, 2008⁶⁷). 24
- Figure 2.6.** Number (and percentage) of peaks with at least possible molecular formulae, the respective sum of all possible assignments, and average number of formulae per

peak (total number of possible formula divided by number of peaks assigned) for a negative ion ESI FT-ICR mass spectrum of Suwannee River fulvic acid (8429 total peaks with S/N > 3; 5029 at odd nominal masses, 3400 at even nominal masses) for three different sets of the number of elements used for formula determination: (A) $C_{0-\infty}H_{0-\infty}O_{0-\infty}$, (B) $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-30}$, and (C) $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-30}S_{0-2}P_{0-2}$. Mass accuracy < 1 ppm (adapted from Koch et al., 2007⁷⁷).28

Figure 2.7. Van Krevelen plots of the molecular formulae (obtained using negative ion ESI FT-ICR mass spectrometry) within DOM collected from the Toivola Swamp and Brule River (adapted from Minor et al., 2012⁸²).31

Figure 2.8. Possible improvements on the representation of Van Krevelen plots for the clarification of compositional differences: Relative intensities displayed using a color plot (left; adapted from Perminova et al., 2014⁸⁵), and different sizes of data points (right; adapted from Yekta et al., 2012⁷⁹).33

Figure 3.1. UV-vis absorption spectra (left) and fraction of absorbance remaining after reduction (right) of UNT (black), BHR (blue), and BDR (green) SRFA. Spectra were taken prior to purification.40

Figure 3.2. Full scan negative ion ESI mass spectra of (A) UNT, (B) BHR, and (C) BDR SRFA. The apparent truncation of peaks at about 620 *m/z* is due to most peaks at higher *m/z* falling below the S/N cutoff. The peak cluster at 814 *m/z* was of unknown origin and were present in all spectra in this study acquired using the 7 T ESI FT-ICR MS (at the Woods Hole Oceanographic Institution).....42

Figure 3.3. Expanded region from 446 – 470 *m/z* of the negative ion ESI 7 T FT-ICR mass spectra of (A) UNT, (B) BHR, and (C) BDR SRFA. Peaks in red are those which were not present in UNT SRFA.43

Figure 3.4. Mass spectra of UNT, BHR, and BDR SRFA at 467 (left; M), 469 (middle), and 470 (right) *m/z*. Red peaks are those that were not present in UNT.45

Figure 3.5. Venn diagrams comparing the number of peaks detected in the mass spectrum of SRFA UNT to those in an untreated replicate (UNT-2), BHR, and BDR mass spectra. Percentages of peaks belonging to each subset to which molecular formulae were assigned are given in parenthesis. Below each Venn diagram is the percentage of all species in the two spectra that were in *both* spectra.46

Figure 3.6. Venn diagrams comparing the number of peaks detected in the mass spectrum of DEUB UNT to those in the corresponding BHR, and BDR mass spectra. Percentages of peaks belonging to each subset to which molecular formulae were assigned are given in parenthesis. Below each Venn diagram is the percentage of all species in the two spectra that were in *both* spectra.46

Figure 3.7. Plots of DBE versus *m/z* of all molecular formulae in untreated SRFA (top) and DEUB (bottom) showing peaks comprising species with no reducible groups

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| (grey), both one and two reducible groups (red, A & C), one group only (green), and two reducible groups only (purple, B & D)..... | 51 |
| Figure 3.8. Plots of H/C versus O/C molar ratios of all molecular formulae in untreated SRFA (top) and DEUB (bottom) showing peaks comprising species with no reducible groups (grey), both one and two reducible groups (red, A & C), one group only (green), and two reducible groups only (purple, B & D). | 52 |
| Figure 3.9. Two hypothetical molecular structures for the molecular formulae $C_{24}H_{28}O_{11}$ (O/C = 0.46, H/C = 1.17) containing one borohydride-reducible group: (A) carboxyl-rich alicyclic molecule (CRAM) and (B) lignin-derived structure. | 53 |
| Figure 4.1. Analysis of mass spectral peak lists (white boxes labeled P(UNT) and P(BDR)) by custom-written MATLAB functions (black arrows). Colored boxes represent matrices of molecular formulae calculated by the indicated functions F1 , F1* , or F2 | 60 |
| Figure 4.2. Number of D-containing molecular formulae identified in BDR-3 (A and C) and UNT (B and D) in four m/z ranges by algorithms F1* and/or F2 with various maximum numbers of N and S using maximum formula error (FE) tolerances of 0.2 (A and B) and 0.3 (C and D) ppm. | 70 |
| Figure 4.3. Percentage of peaks in UNT with an assigned molecular formula identified as comprising singly- and/or doubly- reducible species for m/z ranges (A) 200 – 600, (B) 200 – 300, (C) 300 – 400, (D) 400 – 500, and (E) 500 – 600..... | 71 |
| Figure 4.4. Van Krevelen plots of formula of peaks in UNT identified as non, singly, and/or doubly-reduced based on a search for $m+3.021927$ and $m+6.043854$ mass markers in (top) BDR-1 and (bottom) BDR-3 in the four m/z ranges analyzed (see Table 4.4 for numbers of formulae)..... | 75 |
| Figure 4.5. Absorption and emission of all SRFA samples. A: Blank-subtracted absorption spectra, B: Percent loss of absorption (relative to UNT) following reduction, and C: quantum yield. | 77 |
| Figure 4.6. Corrected emission spectra of untreated and reduced SRFA (25 mg/L), in quinine units (QSU), where 1 QSU = emission intensity of a 1 ppb quinine sulfate solution at (EX, EM) = (350, 450). Mass equivalents of $NaBD_4$ are given in parentheses (see Appendix 6 for difference spectra)..... | 79 |
| Figure 4.7. Number of reduced species identified via ESI FT-ICR MS versus percent change in $A(\lambda)$, $F(EX,EM)$, and $QY(360)$ relative to UNT. “All singly reduced” = singly reduced only + singly & doubly reduced; “all doubly reduced” = doubly reduced only + singly & doubly reduced; and “total reductions” = singly reduced only + 2(doubly reduced only) + 3(singly & doubly reduced). Total reductions (purple triangles) are given on the right-hand vertical axes. BDR-5 data are excluded for percent change in $F(275,306)$, $F(290,480)$, and $QY(360)$. Percent changes are absolute values..... | 81 |

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5.1. Full scan negative ion 12 T FT-ICR mass spectra for all samples included in this study. Purple triangles indicate the base peak in spectra with rescaled relative intensity axes (249.03222 m/z in EAUW and EADO; 209.00919 m/z in EASO-UNT). Blue circles indicate secondary peaks that were also off-scale. All spectra exclude peaks which were doubly or triply charged, were present in the corresponding blank spectrum at high intensity, and/or had S/N < 10. No peaks with $m/z > 800$ or $m/z < 100$ were observed in any sample. | 90 |
| Figure 5.2. Percentage of peaks with assigned molecular formulae in all 7 samples in the three m/z ranges identified as comprising at least one reducible species (A : control search, B : experimental search). | 96 |
| Figure 5.3. Van Krevelen plots of all formulae identified in all mass spectra within three m/z ranges (grey: non-reduced; red: singly-reduced only; green: doubly- and singly-reduced; blue/purple: doubly-reduced only). | 99 |
| Figure 5.4. Percentage of formulae identified in each sample which were also identified in SRFA and PLFA. A : all formulae, B : formulae identified as peaks comprising reducible species. In parenthesis are the total numbers of formulae in each sample within the studied 250 – 550 m/z range. | 101 |
| Figure A1.1. Negative ion ESI FT-ICR mass spectrum of the extraction blank associated with SRFA, EACR, EAUW, EADO, and EASO. Asterisks (*) denote m/z values which were excluded from sample peak lists. | 110 |
| Figure A1.2. Negative ion ESI FT-ICR mass spectrum of the extraction blank associated with PLFA, DELB, and DERV. | 111 |
| Figure A2.1. Negative ion 7 T ESI FT-ICR mass spectra of (A) SRFA UNT replicate 2, DEUB (B) UNT, (C) BHR, and (D) BDR. | 113 |
| Figure A2.2. 12 T ESI FT-ICR mass spectra of SRFA UNT and BDR-1 though BDR-5. All spectra depicted exclude peaks which were doubly or triply charged, were present in the blank spectrum at high intensity, and/or had S/N < 10. No peaks with $m/z > 800$ were observed in any sample. | 115 |
| Figure A4.1. Negative ion ESI-orbitrap mass spectra of 42 μ M acetovanillone in 50:50 methanol/water: (a) untreated, (b) borohydride-reduced, and (c) borodeuteride-reduced aliquots. TIC = total ion count. Masses of molecular formula assignments are within 4 ppm of the measured masses. | 120 |
| Figure A5.1. Molecular formula error vs. m/z for (A) UNT and (B) BDR-3 calculated using a 0.3 ppm error limit using function F2 (mass shift-based assignment of D-containing compounds). | 121 |
| Figure A5.2. Expanded regions of the negative ion 12T ESI FT-ICR mass spectra of (A) UNT, (B) BDR-1, and (C) BDR-4 SRFA at 311 (left) and 314 (right) m/z . Black markers on peaks indicate those which have been assigned a molecular formula. . | 123 |

Figure A6.1. Shifts in the wavelength of maximum emission (EM_{MAX}) for various excitation wavelengths (EX) in BDR-1 through BDR-5 relative to those in UNT.
 $\Delta EM_{MAX} = EM_{MAX} (UNT) - EM_{MAX} (BDR)$ 127

Figure A6.2. Percent loss of absorption (relative to UNT) of all samples at selected wavelengths versus mass equivalents of $NaBD_4$ 128

Figure A6.3. Absorption spectra (left) of whole and PPL extracted SRFA (untreated and reduced) and difference (middle) and fractional (right) spectra comparing the ratios of the four samples. 130

Figure A6.4. Difference emission spectra of BDR SRFA samples relative to UNT SRFA. 131

List of Abbreviations

| | |
|--------------|----------------------------------------------------------|
| AI | aromaticity index |
| $A(\lambda)$ | absorbance at wavelength λ |
| AMW_N | number averaged molecular weight |
| AMW_W | weighted averaged molecular weight |
| APCI | atmospheric pressure chemical ionization |
| APPI | atmospheric pressure photoionization |
| BDR | borodeuteride-reduced sample/spectrum |
| BHR | borohydride-reduced sample/spectrum |
| CDOM | chromophoric dissolved organic matter |
| CRAM | carboxyl-rich alicyclic molecule(s) |
| DBE | double bond equivalent |
| DELB | Lower Delaware Bay C_{18} extract |
| DERV | Delaware River C_{18} extract |
| DEUB | Upper Delaware Bay C_{18} extract |
| DOM | dissolved organic matter |
| EACR | Congo River plume C_{18} extract |
| EADO | Northern Atlantic Ocean C_{18} extract (deep ocean) |
| EASO | Northern Atlantic Ocean C_{18} extract (surface ocean) |
| EAUW | Equatorial upwelling region C_{18} extract |
| EEMS | excitation/emission matrix spectroscopy |
| EM | emission wavelength |
| EM_{MAX} | wavelength of maximum emission |

| | |
|------------|--------------------------------------------------------------------|
| ESI | electrospray ionization |
| EX | excitation wavelength |
| F1 | formula assignment algorithm (low-mass moiety, deuterium excluded) |
| F1* | formula assignment algorithm (low-mass moiety, deuterium included) |
| F2 | formula assignment and ketone/aldehyde identification algorithm |
| FDOM | fluorescent dissolved organic matter |
| FE | formula error |
| FID | free induction decay |
| FT-ICR | Fourier transform ion cyclotron resonance |
| FTIR | Fourier transform infrared |
| FWHM | full width at half maximum |
| h | hour(s) |
| H/C | hydrogen-to-carbon molar ratio |
| HCl | hydrochloric acid |
| HS | humic substances |
| I(EX, EM) | emission intensity at excitation/emission wavelengths EX and EM |
| λ | wavelength (general) |
| LTQ | (Quadrupolar) linear ion trap |
| <i>m</i> | mass |
| M | molar/molarity |
| ME | mass equivalents |
| MeOH | methanol |
| mg | milligram(s) |

| | |
|-------------------|----------------------------------------------------------------------|
| mL | milliliter(s) |
| m/z | mass to charge ratio |
| NaBD ₄ | sodium borodeuteride |
| NaBH ₄ | sodium borohydride |
| NaOH | sodium hydroxide |
| NMR | nuclear magnetic resonance |
| NOM | natural organic matter |
| O/C | oxygen-to-carbon molar ratio |
| ω_c | ion cyclotron frequency |
| PD | polydispersity |
| Pg | petagrams |
| PLFA | Pony Lake fulvic acid |
| POM | particulate organic matter |
| $\Phi(\text{EX})$ | fluorescence quantum yield at excitation wavelength EX |
| PPL | functionalized polystyrene divinyl benzene (proprietary solid phase) |
| ppm | parts per million |
| rf | radiofrequency |
| RP | resolving power |
| QSU | quinine sulfate units |
| S _A | absorption spectral slope |
| SEC | size exclusion chromatography |
| S/N | signal-to-noise ratio |
| SPE | solid phase extraction |

| | |
|--------|-----------------------------------------|
| SRFA | Suwannee River fulvic acid |
| TIC | total ion count |
| UNT | untreated (non-reduced) sample/spectrum |
| UV-vis | ultraviolet-visible |

Chapter 1: Aquatic natural organic matter (NOM): Definitions, sub-fractions, and environmental significance

1.1. A broad definition of natural organic matter (NOM)

Virtually all aquatic environments contain complex, heterogeneous ensembles of organic compounds. These ensembles comprise what is collectively referred to as natural organic matter (NOM), which can vary greatly in complexity, diversity, and concentration from one geographic locale to another. The worldwide size of the total pool of NOM within aquatic environments has been estimated to be over 1,000 Pg (petagrams), which surpasses the 720 Pg of atmospheric carbon dioxide.^{1,2} This nebulous description of NOM attests to the difficulty that researchers still face today in an attempt to fully understand its composition and reactivity, even after over a century of active research.

NOM can be divided roughly into two location-based subcategories: Riverine and marine NOM (Figure 1.1). The first of these two categories is generally accepted to comprise byproducts from the decomposition and runoff of terrestrial plant and animal matter, which is gradually transported to larger bodies of water such as lakes and estuaries. The origin and composition of the latter of these two categories is far more ambiguous and, as a result, has been the subject of much debate. While some studies show that marine NOM is greatly influenced by riverine and estuarine output,³⁻⁵ other studies have provided evidence that marine NOM arises largely from the *in situ* production from marine biomass such as algae and phytoplankton.^{6,7} Conversely, others

have demonstrated that marine NOM may act as a food source for such biota.⁸ In most cases, it is likely that marine NOM arises from a combination of both riverine input and *in situ* microbial production and consumption,⁹ and future studies may help elucidate the relative contribution of riverine input and *in situ* production of marine NOM in various geographic locales.

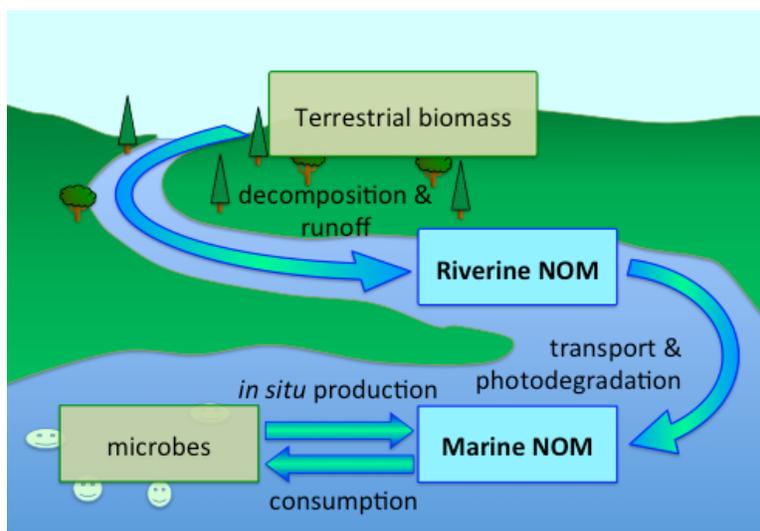


Figure 1.1. Two major categories of aquatic NOM and their primary sources and sinks.

In order to fully understand and predict the reactivity and dynamics of NOM within aquatic environments, an explicit description of the molecular composition of NOM must be realized. Early attempts at this goal relied on the assumption that if a sample of NOM was divided into different fractions using various physical and chemical techniques (e.g. chromatography, precipitation), each fraction would comprise a distinctive class of chemical compounds. Although this is not necessarily true, many of the separation techniques that were formerly used to characterize NOM are commonly practiced today as useful operational definitions of NOM sub-fractions (Figure 1.2). For

instance, the distinction between particulate organic matter (POM) and dissolved organic matter (DOM) is made using a 0.2 or 0.45 μm membrane filter, where the organic material that is retained is defined as POM, while the material that passes through the filter comprises DOM.¹⁰ The DOM in this permeate can then be separated from inorganic compounds via techniques such electro-dialysis/reverse osmosis or, more commonly, solid phase extraction (SPE).¹¹

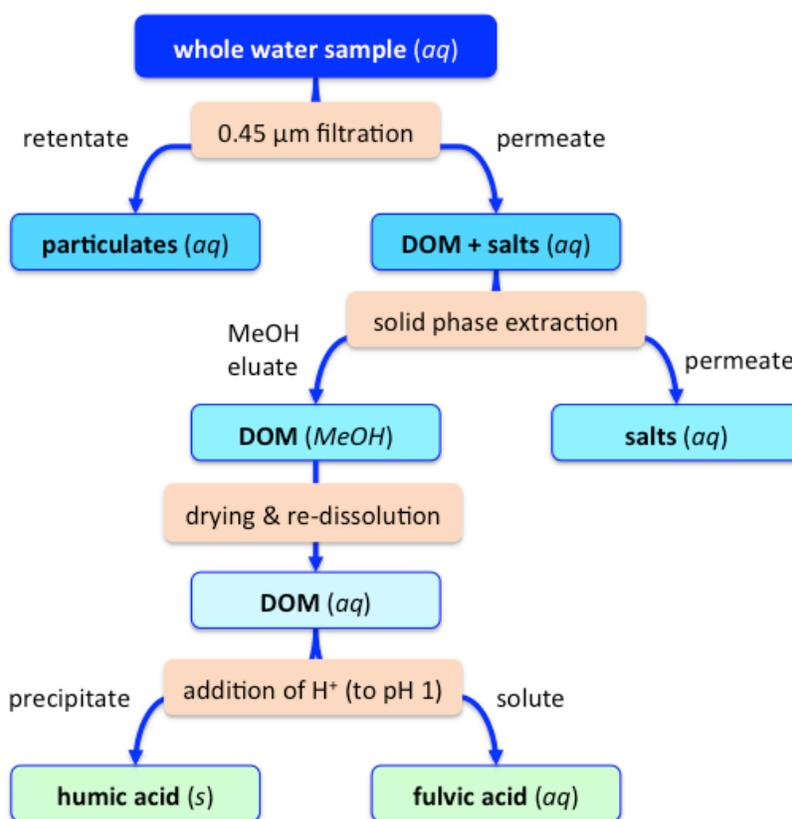


Figure 1.2. Common separation processes used for isolating various sub-fractions of NOM from a water sample collected from an aquatic environment.

Other important operationally defined components of NOM are humin, humic acid, and fulvic acid, all of which are collectively referred to as humic substances (HS).¹² These sub-categories are defined based on their pH-dependent water solubility; humin is the portion that is insoluble under all pH conditions, and is therefore not considered a component of DOM. Of the two water-soluble HS, species which are insoluble under low pH (<2) are defined as humic acids, while those which are soluble under all pH conditions are defined as fulvic acids. While some researchers continue to rely heavily on separation techniques to study the composition of DOM,¹³ this approach must be used with caution due to the importance of intermolecular interactions (e.g. hydrogen bonding and complexation), since destroying these interactions via separation can lead to significant changes in many physicochemical properties of the bulk NOM ensemble.^{14,15}

1.2. Description of chromophoric dissolved organic matter (CDOM)

Presence and significance of chromophoric dissolved organic matter in the environment. Chromophoric dissolved organic matter (CDOM) is the name given to the fraction of DOM which absorbs light over a broad range of ultraviolet (UV) and visible wavelengths (~200 – 700 nm). Due to the characteristic brown to yellow color of these materials, colored natural organic matter has been referred to as *yellow substance* in the older literature.¹⁶ The currently accepted term “CDOM” refers to all species within an given DOM extract that participate in the observable photophysical properties of the bulk DOM samples, and unlike the operationally defined fractions described above, it may be impossible to physically isolate the “chromophoric” compounds from the “non chromophoric” ones within a DOM sample.

Owing to the absorption of light over a broad spectral range, CDOM can have a considerable impact on many biogeochemical processes within an aquatic environment. For instance, light absorption can substantially alter the solar radiation to which an aquatic ecosystem is exposed. This alteration can protect organisms (e.g. coral reefs) by filtering out harmful UV radiation.¹⁷ Conversely, CDOM's absorption at visible wavelengths (400 – 700 nm) can decrease the amount of solar radiation available to chlorophyll, thus potentially hindering photosynthesis. Furthermore, absorption of light by chromophoric species can initiate myriad photochemical reactions such as the photosensitization of other chemical species within a given aquatic environment (e.g. pollutants)^{18,19} as well as the generation of various inorganic species such as carbon monoxide, carbonyl sulfide,¹⁵ and reactive oxygen species such as hydrogen peroxide.¹⁶ Overall, CDOM's participation in this complex web of biogeochemical processes is still poorly understood. Thus, research aimed at determining the chemical structures that influence the optical properties of CDOM would help explain and predict its role within aquatic ecosystems, as well as the environmental factors that influence its composition.

Description of the optical properties of CDOM. Despite the virtually limitless variability in composition that is possible for such a complicated ensemble of molecular species, the absorption and emission of light by CDOM extracted from nearly all aquatic environments follow remarkably similar trends. The absorption of CDOM decreases exponentially with increasing wavelength and extends into the visible and even near-infrared region of the electromagnetic spectrum, with very few, if any, discernable peaks or shoulders (Figure 1.3). Similarly, the wavelength of maximum emission and quantum yield also display monotonic dependencies on the excitation wavelength used.²²

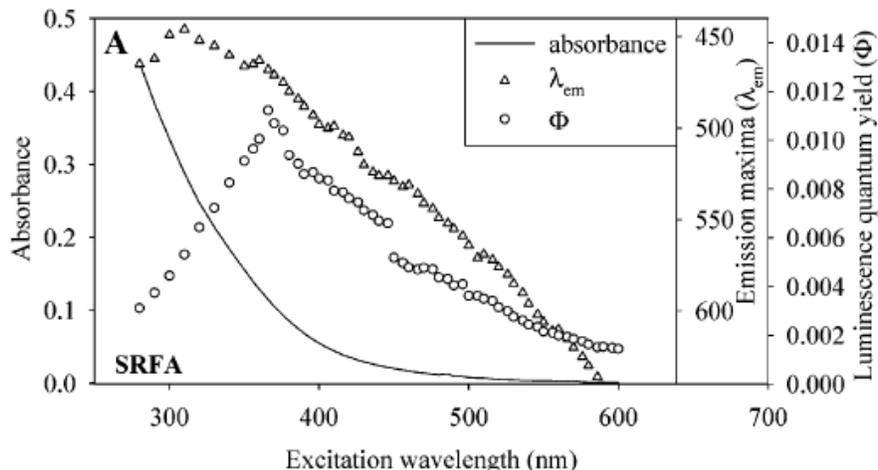


Figure 1.3. Absorption, wavelength of maximum emission, and luminescence quantum yield versus excitation wavelength for Suwannee River fulvic acid (SRFA). Adapted from Del Vecchio and Blough, 2004.²²

Owing to the steady, exponential decrease in absorbance across the UV-visible wavelength regime, it has become a common practice to fit experimentally measured absorbance data across a broad range of wavelengths to the equation

$$(1.1) \quad A(\lambda) = A(\lambda_{ref})e^{-S_A(\lambda-\lambda_{ref})}$$

where $A(\lambda)$ is the absorbance at wavelength λ between a given range of wavelengths, λ_{ref} is a reference wavelength, and S_A is the absorption spectral slope, which is determined from the fit.¹⁶ The spectral slope defines how quickly the absorption decays with increasing wavelength, and thus serves as a useful parameter for CDOM characterization, and is commonly used as a simple proxy for changes in CDOM composition.

Although the lack of spectral features of CDOM limits the usefulness of UV/visible absorption spectrophotometry, it is possible to infer some general conclusions regarding CDOM's overall composition when combined with other analytical techniques. For instance, Chin et al. reported that the average molecular weight (determined by high performance size exclusion chromatography) and aromaticity (determined by ¹³C nuclear

magnetic resonance spectroscopy) of humic and fulvic acids from various locations throughout the United States correlated very well ($R^2 \geq 0.90$) with their molar absorptivities at 280 nm (calculated on a per molar basis of measured total organic carbon).²³ Other studies have demonstrated that the spectral slope can be a useful indicator of major sources of DOM, with low spectral slopes being indicative of a high influx from nearby terrestrial (or riverine) sources, while high spectral slopes indicative of photobleached and/or perhaps microbially derived DOM.^{3,24,25}

Excitation/emission matrix spectroscopy (EEMS) has been used to provide more detailed information about the structure and composition of CDOM. In this technique, a sample is excited at several wavelengths (usually spaced ≤ 10 nm apart) and an emission spectrum is recorded for each excitation wavelength. This results in a 3-dimensional plot (Figure 1.4) of fluorescence intensities over a range of excitation and emission wavelengths (EX and EM, respectively) and offers two primary advantages over absorption spectroscopy. Firstly, EEMS allows for greater resolution of broad electronic spectral features by spreading them out over a second dimension. Secondly, the superior sensitivity of fluorescence to absorption spectroscopy allows for measurements to be made on natural water samples without the need for concentration (e.g. via solid phase extraction) of CDOM, which can substantially alter the composition and observed optical properties.^{3,26,27} The primary disadvantage of fluorescence spectroscopy is that only a fraction of the organic species which *absorb* UV and visible light will *emit* radiation. Therefore, EEMS measures the optical properties of only a sub-fraction of CDOM known as fluorescent dissolved organic matter (FDOM). However, it has been demonstrated that FDOM and CDOM are very closely correlated with one another.²⁸

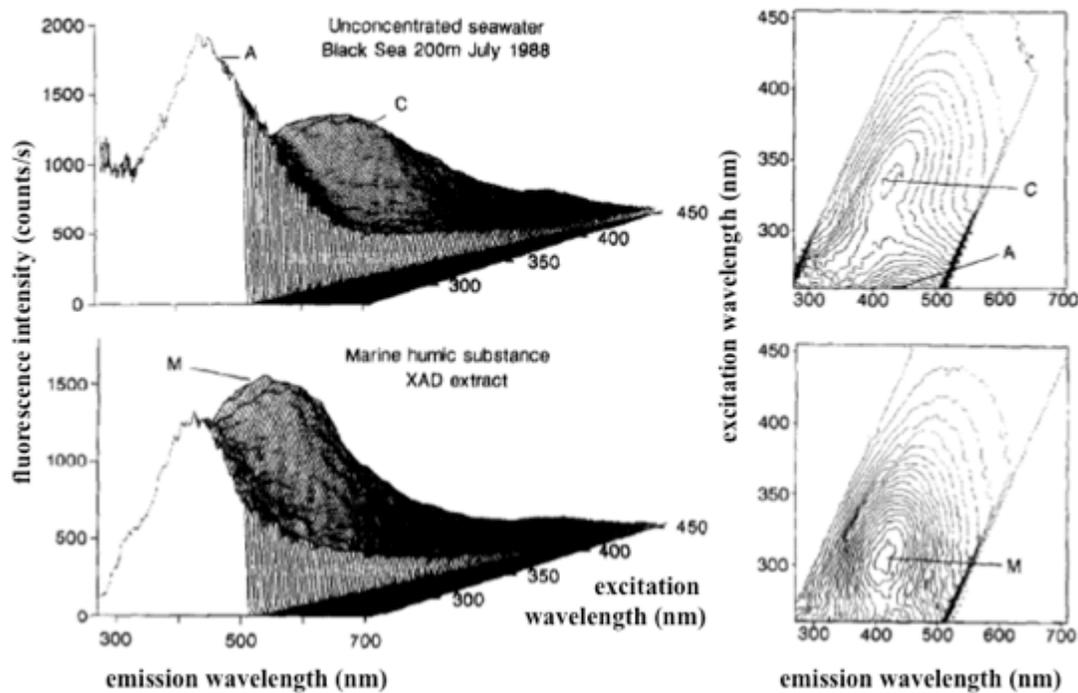


Figure 1.4. Fluorescence EEM of two CDOM samples, depicted using 3-dimensional wireframe (left) and contour plot (right) renderings. Adapted from Coble et al., 1996.²⁹

The position (i.e., EX and EM) and intensity of bands in EEM spectra can be used to discriminate differences in CDOM composition and source. For instance, EEM spectra of humic and fulvic acids display two major features: One peak at EX/EM = 260/380 – 460 nm and a lower intensity peak at EX/EM = 320 – 360/420 – 460 nm, which are generally referred to as UVC and UVA humic-like bands, respectively (Table 1.1). For CDOM samples collected from marine environments, this second band is usually shifted to shorter EX and EM ranges. In some cases, two UVA humic-like bands (marine and non-marine) can be resolved.³⁰ Also commonly present in CDOM samples are protein-like signatures at EX/EM = 275/310 – 340 nm. This feature typically presents itself as a well-defined shoulder of the higher intensity UVC humic-like band. The differences in

λ_{EX} and λ_{EM} for these bands, as well as their intensities, can thus be correlated to compositional changes within CDOM.

Table 1.1. Common fluorophores identified in CDOM.

| EX (nm) | EM (nm) | assignment | peak name ²⁹⁻³¹ |
|-----------|-----------|----------------------------|----------------------------|
| 260 | 380 - 460 | UVC humic-like | A |
| 275 | 310 | protein-like (tyrosine) | B |
| 275 | 340 | protein-like (tryptophan) | T |
| 290 – 310 | 370 – 410 | UVA marine humic-like | M |
| 320 – 360 | 420 – 460 | UVA humic-like | C |
| 400 | 660 | pigment-like (chlorophyll) | P |

1.3. Physicochemical justification of the optical properties of CDOM

An intuitive approach to rationalizing the optical properties of CDOM may be to assign all spectral features to different classes of chromophores or fluorophores, thus assuming that the optical properties can be accurately described by a superposition of independently absorbing and emitting organic compounds within a DOM sample. However, many studies have demonstrated that a superposition model has many shortcomings, and that electronic interactions (e.g. charge-transfer complexes) between different chromophores greatly influence the observed optical properties.

The long wavelength absorption of CDOM is particularly difficult to justify with a simple superposition of the spectra of many organic compounds, since most organic chromophores absorb primarily at wavelengths shorter than ~300 nm. The main exception to this general rule are highly conjugated aromatic species and/or quinones, which can have $n \rightarrow \pi^*$ transitions that occur in the visible wavelength regime. However, such absorption bands are typically rather weak, and thus, if the total absorption spectra

of CDOM was indeed a simple superposition of many spectra, the absorption at short wavelengths (200 – 300 nm) would most likely be far greater than that at visible wavelengths. Perhaps even more difficult to justify are the excitation wavelength-dependent trends in the wavelength of maximum emission and quantum yield, since most organic fluorophores which absorb and emit at long wavelengths typically have relatively high quantum yields. Therefore, the concomitant *decrease* in quantum yield with *increasing* excitation wavelength observed for CDOM is surprising. Also, since emission is largely independent of excitation wavelength, the continuous red shift in emission maxima that occurs with increasing excitation wavelength implies a continuum of *coupled* species.

These observations, along with a wide range of other experimental^{22,25,32–34} and theoretical^{35,36} considerations have led to the acceptance of a model in which the absorption and emission of CDOM – especially in the visible wavelength regime – are better explained by a coupled manifold of many inter- and intramolecular electronic interactions, rather than merely a superposition of independent chromophores. These interactions are assumed to take place largely between poly hydroxy/methoxy aromatic electron donors and aromatic carbonyl electron acceptors which are present in partially oxidized lignin, a major component of CDOM.

The suitability of this electronic interaction model has been rigorously tested by observing the optical properties of CDOM following treatment by sodium borohydride (NaBH₄), a selective reductant of ketones and aldehydes.^{3,37,38} In these experiments, the absorption of a CDOM extract across a broad range of UV and visible wavelengths decreased substantially following treatment by NaBH₄, with a greater *fractional* loss at

successively longer wavelengths (Figure 1.5). This result is difficult to rationalize by assuming a superposition of independent chromophores, since ketones generally do not absorb at wavelengths > 350 nm. Furthermore, although some quinones which have weak absorption bands at visible wavelengths and are readily reduced by borohydride to hydroquinones, these species usually rapidly reoxidize in the presence of air, whereas the majority of the loss of absorption observed in many CDOM samples is irreversible.

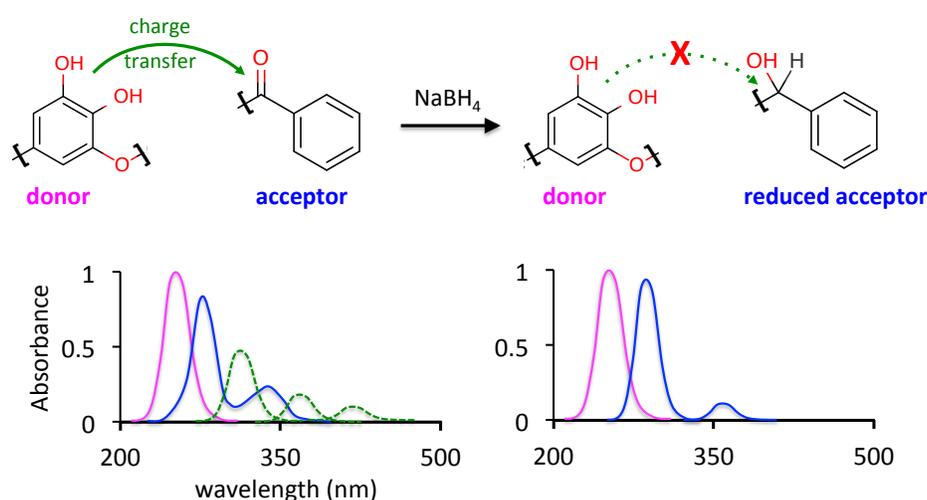


Figure 1.5. Schematic of an absorption spectrum of CDOM before (left) and after (right) treatment by NaBH_4 showing the disruption of charge transfer interactions (by reduction of acceptor moieties) and the corresponding loss of longer wavelength (>300 nm) absorption bands.

The reduction experiments described above indicate the importance of ketone and aldehydes within CDOM to the characteristic long wavelength absorption, even though such compounds would probably not absorb at long wavelengths if isolated. This interaction demonstrates, as alluded to earlier, that species within a DOM sample cannot be meaningfully separated into “chromophoric” and “non chromophoric” species. Rather, it is important to describe the role that individual compounds may play in the optical properties of a DOM sample in the presence of other “non chromophoric” species, thus

resulting in an accurate description of how the total ensemble of DOM species interact with one another.

Chapter 2: Advanced characterization of DOM using ultrahigh resolution mass spectrometry (an overview)

2.1. Analyzing the structure and composition of DOM

Most analytical techniques are capable of only measuring bulk properties of an ensemble of dissolved organic matter (DOM). Although some techniques such as excitation emission matrix spectroscopy (EEMS; described in Chapter 1) may at first seem to resolve different subsets of species, inferring the presence of even broadly defined compound classes is difficult, since electronic interactions can drastically alter the apparent composition. Other instrumental methods such as Fourier transform infrared (FTIR) and nuclear magnetic resonance (NMR) spectroscopy allow the relative abundance of specific structural features to be measured. Although all of the above methods have been and continue to be invaluable for studying DOM,^{39,40} they are limited in terms of specificity, since they only allow for the measurement of bulk properties of an entire sample.

A common practice in early DOM research was to describe a hypothetical “average DOM molecule” for DOM ensembles (Figure 2.1A,B). However the usage and description of these types of hypothetical model compounds is misleading, as it is likely that no such molecule exists in a given DOM sample. Instead, DOM can be more accurately thought of as a complex collection of a wide range of (mostly biologically derived) chemical compounds (Figure 2.1C). In fact, owing to the multitude of biological sources, an ensemble of DOM extracted from any given aquatic environment can consist of hundreds to even thousands of distinct chemical species.

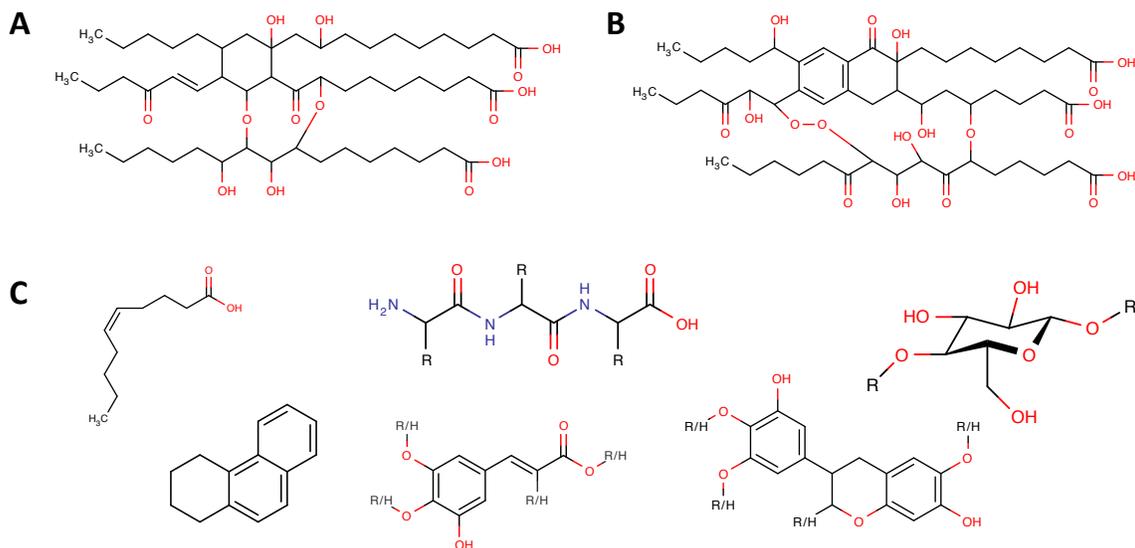


Figure 2.1. Early hypothetical structures for (A) fulvic and (B) humic acids from seawater (structures from Zafiriou et al., 1984⁴¹) and (C) representations of various classes of chemical compounds (lipids, peptides, cellulose, condensed hydrocarbons, lignin, condensed tannins) that may exist in a DOM sample.

Due to the overwhelming number of individual chemical species within DOM, it is practically impossible to identify specific compounds by using the aforementioned methods. A fundamental feature of mass spectrometry, on the other hand, is that it detects and measures the mass to charge ratio (m/z) ions produced by discrete chemical species (or fragments of species), thus serving as a complementary approach to the bulk analysis methods described above. This approach is especially useful when a “soft” ionization technique (i.e., one that does not produce fragments of compounds due to ionization) such as electrospray ionization (ESI) is coupled with ultrahigh resolution Fourier transform ion cyclotron resonance (FT-ICR) mass analysis. With this combination, a substantial subset of species within DOM can be ionized, detected, and differentiated from one another.

2.2. Instrumentation for mass spectrometric analysis of DOM

Ionization. For decades, mass spectrometry had limited use in DOM research, since most early ionization techniques required samples to be in the gas phase prior to ionization and mass analysis. Due to this limitation, mass spectrometric analysis could only be applied to small fractions of a total DOM sample. However, in the 1980s, electrospray ionization (ESI) quickly revolutionized mass spectrometry by greatly broadening the types of samples for which it can be used to study. ESI allows for the ionization of nonvolatile organic compounds containing polar functional groups. The general setup for an ESI interface is shown in Figure 2.2 (operated in negative ion mode). Although positive ion mode is more commonly used and discussed throughout the literature, negative ion mode is usually preferred for the analysis of DOM.

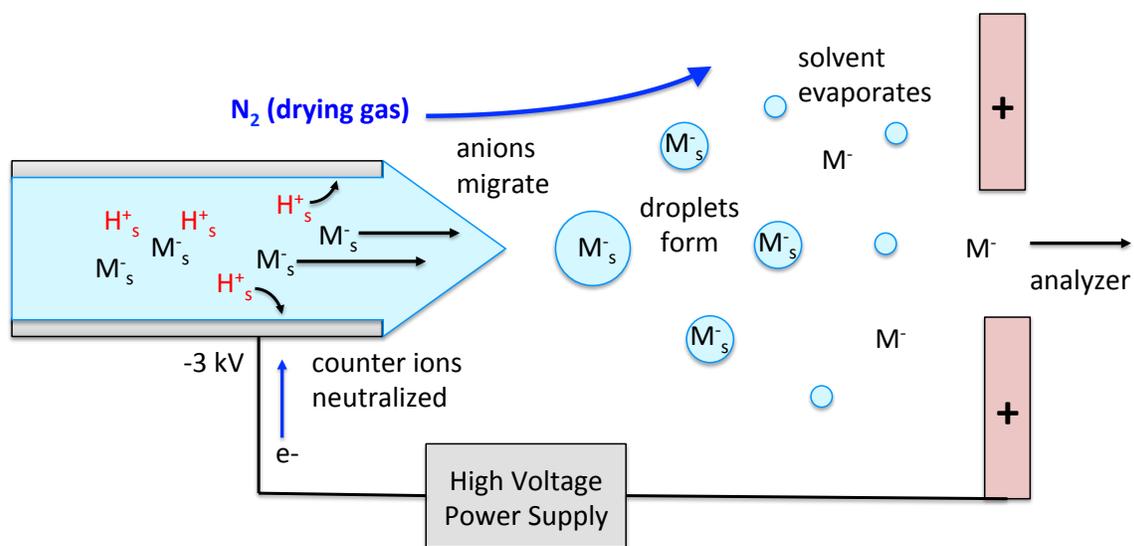


Figure 2.2. Schematic of an ESI interface operated in negative ion mode.

In an ESI interface, a sample solution is introduced at a rate of $\sim 50 \mu\text{L}/\text{min}$ through a ($\sim 0.1 \text{ mm}$ internal diameter) metal capillary tube held at a high ($3 - 5 \text{ kV}$) potential. In negative ion mode, the capillary is held at a negative potential, and as a result, positively charged counter ions are attracted to the interior walls, where they are neutralized via redox reactions. For acidic species, the counter ions are typically H^+ and Na^+ . Simultaneously, the anionic forms of the compounds within the solution drift towards the front of the tube forming a cone shape at the tip.

As anions collect at the tip of this cone shaped jet, the electrostatic repulsion of the like-charged ions eventually overcome the surface tension of the liquid. When this occurs, the jet breaks up into small electrically charged droplets, which drift apart from one another while also drifting towards the positively charged counter electrode near the instrument inlet. The combination of the anions repelling one another while travelling towards the inlet results in a spray of charged droplets. Meanwhile, the charged droplets continue to break apart into successively smaller droplets until gas phase analyte anions are left, which eventually make their way into the mass spectrometer. This process is facilitated by a flow of nitrogen gas which aids in the evaporation of the solvent.

The ESI process described above generates deprotonated, quasi-molecular anions of compounds in DOM which contain polar functional groups. In the ESI process, solvent composition greatly influences the efficiency in which species are ionized.⁴² One of the most popular solvent compositions for ESI mass spectrometric analysis of DOM is a mixture of methanol and water, or in some cases, pure methanol.⁴³ The fact that methanol works so well in ESI is perhaps serendipitous, as it is commonly used to elute chemical species from the solid phase extraction (SPE) cartridges used to isolate DOM. Therefore,

eluates from the SPE processes can be readily analyzed via ESI. Moreover, ESI can be interfaced with high performance liquid chromatography, which can further aid in the characterization of DOM.^{44,45}

Although positive ion mode is more common and generally achieves greater sensitivity, negative ion is preferential for DOM analysis. This because positive ion mode can potentially lead to the formation of not only the protonated quasi molecular ions ($M+H^+$), but also sodium adducts ($M+Na^+$), thus complicating the resulting mass spectrum.⁴⁶⁻⁴⁹ Also, chemical compounds with carboxylic acid functional groups are known to be prevalent in DOM⁴⁹⁻⁵¹, and are therefore particularly amenable to ESI. Although ESI continues to be the most popular and well established ionization method for DOM analysis, it is possible that using other soft ionization techniques such as atmospheric pressure photoionization (APPI) and atmospheric pressure chemical ionization (APCI) can be used as complementary techniques to give a more complete picture of the composition of DOM by preferentially ionizing subsets of DOM species which are not as amenable to ESI.⁵²

Mass analyzers and the need for ultrahigh resolution. In order to successfully differentiate as many compounds as possible within such as complex sample as DOM, a mass spectrometer must be capable of resolving species with similar m/z . The resolving power (RP) of a mass spectrometer is typically defined by the equation

$$(2.1) \quad RP = \frac{M}{\Delta M}$$

where M is the measured m/z of a peak and ΔM is the width of the peak at half the maximum intensity. RP is most often reported as the full width of a mass spectral peak at

half its maximum intensity (FWHM) at a particular m/z value (usually 400).⁵³ Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers can routinely achieve RP beyond 500,000 FWHM, which corresponds to peak widths of less than 0.001 m/z for ions with m/z 500.

As stated earlier, a DOM sample may contain thousands of compounds. Therefore, FT-ICR mass analyzers are needed to differentiate all ionizable compounds within a DOM sample. Such analyzers reveal that DOM can contain around 20 peaks within a single nominal mass (Figure 2.3A). When lower resolution mass analyzers are used with ESI for DOM analysis, many of the peaks at each nominal mass become broadened and unresolved (Figure 2.3B, C). However, although mass analyzers with lower (e.g. < 50,000 FWHM) would be insufficient for the detection of specific compounds within DOM, they have been used to successfully gain insights in to the general composition of DOM samples. For instance, time-of-flight and/or quadrupole mass analyzers have been used to measure molecular weight distributions of different fractions of DOM.^{48,54-56} Selectivity of specific subsets,⁵⁷ as well as overall compositional information can also be gained by using tandem mass spectrometry.^{45,49}

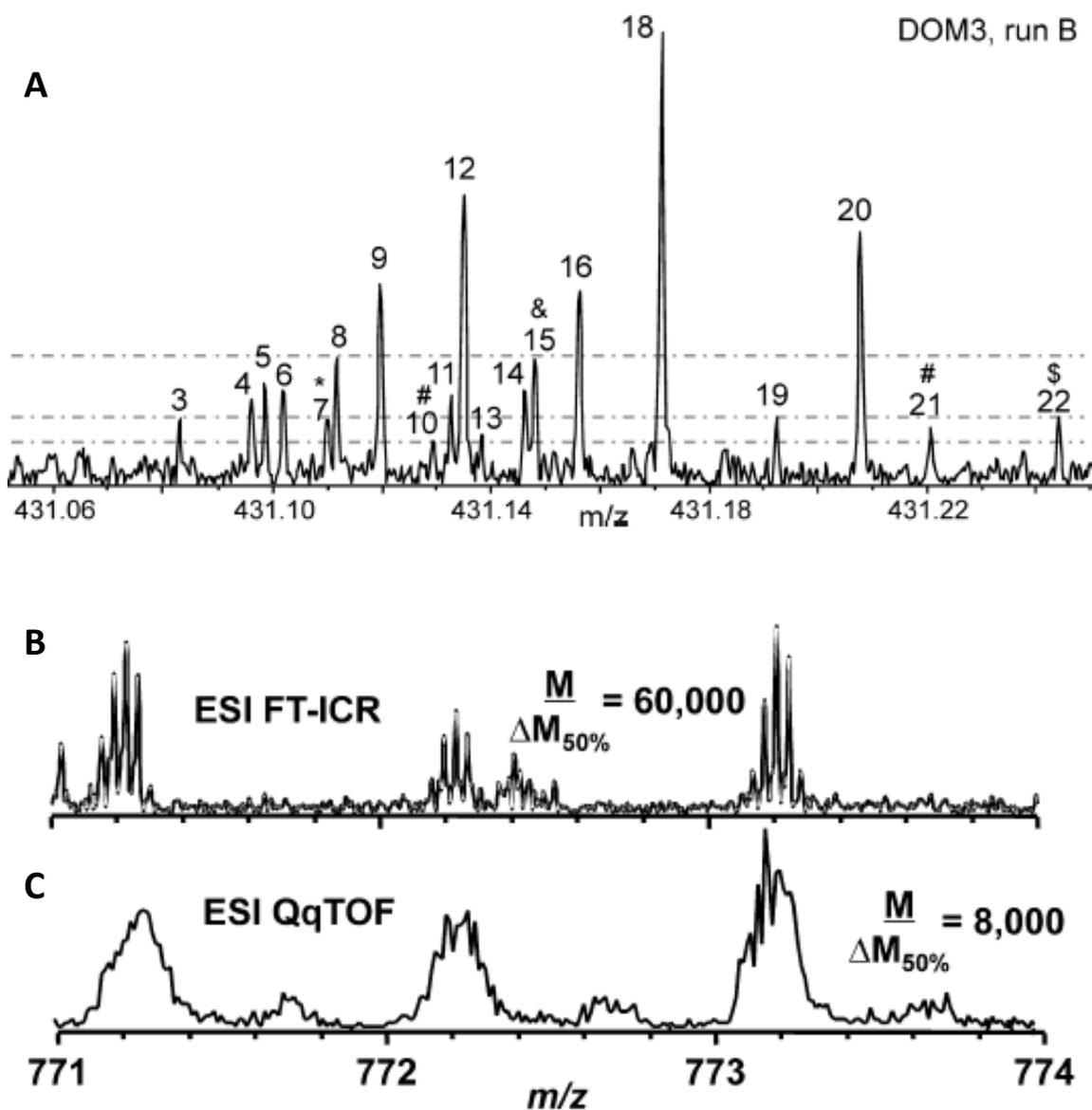


Figure 2.3. ESI mass spectra of various DOM samples collected using a range of operating resolving powers: (A) DOM extracted from the North Pacific Ocean using 12 T FT-ICR with a resolving power of $> 500,000$ (adapted from Sleighter et al., 2012⁵⁸) and humic acid from Mount Rainier using (B) 7 T FT-ICR and (C) quadrupole time-of-flight mass analyzers (adapted from Kujawinski et al., 2002⁵⁹).

The exceptional mass accuracy and resolution of FT-ICR mass spectrometry was first demonstrated by Comisarow and Marshall in 1974.⁶⁰ An FT-ICR mass analyzer consists of cell known as a Penning trap, which is placed inside a strong, uniform magnetic field. Packets of ions are introduced into the cell via radiofrequency (rf) multipole ion guides and are confined axially in the cell by an applied electrical field. A pair of electrodes excite the confined ions into coherent orbital motions. These motions are subsequently detected by a second pair of electrodes, resulting in a complex waveform signal from which the m/z and intensity of each ion can be accurately determined.

There are three main physical principles on which an FT-ICR mass analyzer is based. First, ions within a uniform magnetic field will travel in a circular pattern perpendicular to the applied field at a frequency equal to

$$(2.2) \quad \omega_c = \frac{B}{m/z}$$

Where B is the strength of the magnetic field and m/z is the mass to charge ratio of the ion, and ω_c is the ion's *cyclotron frequency*.⁶¹ Secondly, ions experiencing this cyclotron motion can absorb energy from an externally applied rf pulse, provided that the frequency of the pulse matches that of the ω_c of the ion (resonance). Finally, the circular motion of a single ion between a pair of plate electrodes (perpendicular to the plane of ion motion and parallel to one another) will create an alternating current due to its varying proximity between the two plates.

Although the cyclotron motion spontaneously occurs from trapping ions in the magnetic field, it does not by itself result in a measurable electrical signal. This is because ion cyclotron radii for ions at room temperature is prohibitively small to create a

measurable current (typically < 0.8 mm for ions less with masses less than 10,000 atomic mass units⁶²). Furthermore, initially, the ion cyclotron motions of an ensemble of ions is incoherent and random. That is, since ions begin their circular motion at random times, any charge that is induced on one detector plate by the proximity of single ion at any given moment will be balanced out by an equal charge on the opposite detector plate by induced by another ion.

Since, as stated earlier, ions of a single m/z of interest can be excited to a larger and coherent cyclotron motion by application of an rf pulse of an appropriate frequency by a pair of excitation electrodes. After all ions are excited, the ion packet's circular motion between a second set of detector electrodes will create a current which varies sinusoidally with time, with the amplitude of the signal being proportional to the number of ions in the coherent packet. While the resulting current is detected, the ion motions eventually relax back to their initial cyclotron radii, thus resulting in a dampening of the amplitude of the resulting electrical signal. This electrical signal is therefore referred to as a free induction decay (FID).

Typically it is desirable to detect a broad range of m/z ions. In this case, ions of different m/z are successively excited to coherent cyclotron radii by using an excitation pulse that sweeps through a range of frequencies (e.g., 72 to 641 kHz over 3.79 ms, corresponding to excitation of ions with m/z between 225 and 900⁶³). This type excitation pulse results in an signal which comprising a superposition of the FID waveforms created by the coherent packets of ions whose ω_c fall within the range of the frequencies swept. This superposition can then be Fourier transformed to result in a plot of signal intensity

versus ω_c . Since, as seen previously, ω_c is only dependent on m/z and B, this plot can be easily calibrated to be converted to a mass spectrum.

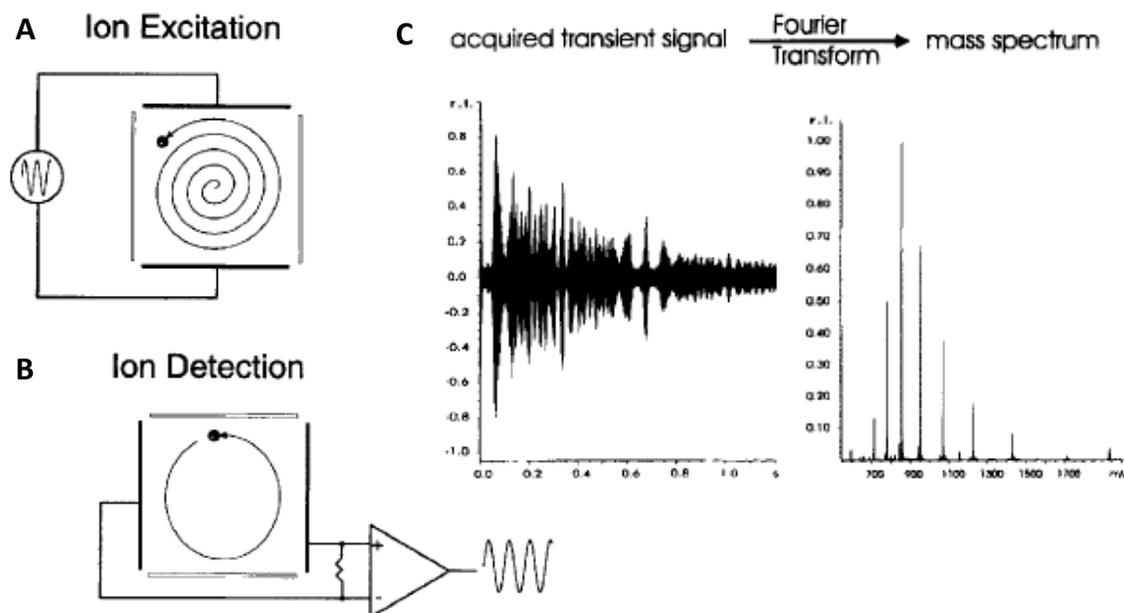


Figure 2.4. Illustrations of (A) excitation of ions with small initial incoherent cyclotron orbits to larger and more detectable coherent motions and (B) subsequent detection of the currents in the ICR cell resulting from the coherent motions. (C) A measured FID signal and the corresponding mass spectrum obtained by Fourier transformation of the FID (adapted from Amster, 2002⁶⁴).

There are several instrumental parameters that affect the quality of a mass spectrum in terms of the achieved resolving power (RP), signal-to-noise ratio (S/N), and number of detected peaks. Such parameters include the strength of the magnetic field and the length of the scan for a single acquisition. Since ω_c is directly proportional to field strength (B, Equation 2.2), the difference between the observed ω_c of two ions with different masses will be greater for instruments with high-field magnets than for

instruments with a lower field magnet. Usually, the strength of the magnetic field in modern-day FT-ICR mass spectrometers are between 7 to 18 tesla (T). The greater the change in ω_c for a given change in m/z , the greater the resolution that is achieved.⁶⁵ Although the magnitude of the field does not adversely affect any experimental parameters, such high-field magnets are prohibitively expensive.

In general, the resolution of a mass spectrometer is determined by the distance ions travel in a given scan. A fundamental reason why FT-ICR mass spectrometers are capable of substantially higher resolution than any other type of instrument is that in a the rapid, circular precession of ions within a cell allow ions to travel very large distances during the course of the scan. For instance, in a typical ICR cell, an excited ion of 100 m/z can travel a distance of about 30 km during a 1 s observation period.⁶² Thus, the cyclotron frequency of a single ion can be determined to a very high degree of precision due to the large number of revolutions that the ion makes during the single acquisition period. This period of time in which the resulting FID is measured can be experimentally controlled to maximize resolution. However, the maximum time that a single acquisition can be scanned for is limited by how quickly an ion's excited cyclotron motion decays. Furthermore, longer transient times may result in distortions in the mass distribution observed due to ion cloud stability and space charge effects.^{43,66}

2.3. Analysis, interpretation, and representation of mass spectrometric data

Although DOM samples can vary greatly in specific composition, the ultrahigh resolution mass spectra of DOM samples display remarkably similar features (Figure 2.5). Typically, ESI FT-ICR mass spectra of DOM contain distributions of up to several thousands of peaks which are centered at approximately 300 – 500 m/z , with several peaks appearing at each nominal mass. Also several specific m/z differences between peaks are frequently observed, which correspond to specific chemical changes between two species. Due to the overwhelming amount of data contained within a single FT-ICR mass spectrum of DOM, data must be simplified to efficiently and clearly highlight the differences and similarities between the composition of different DOM samples.

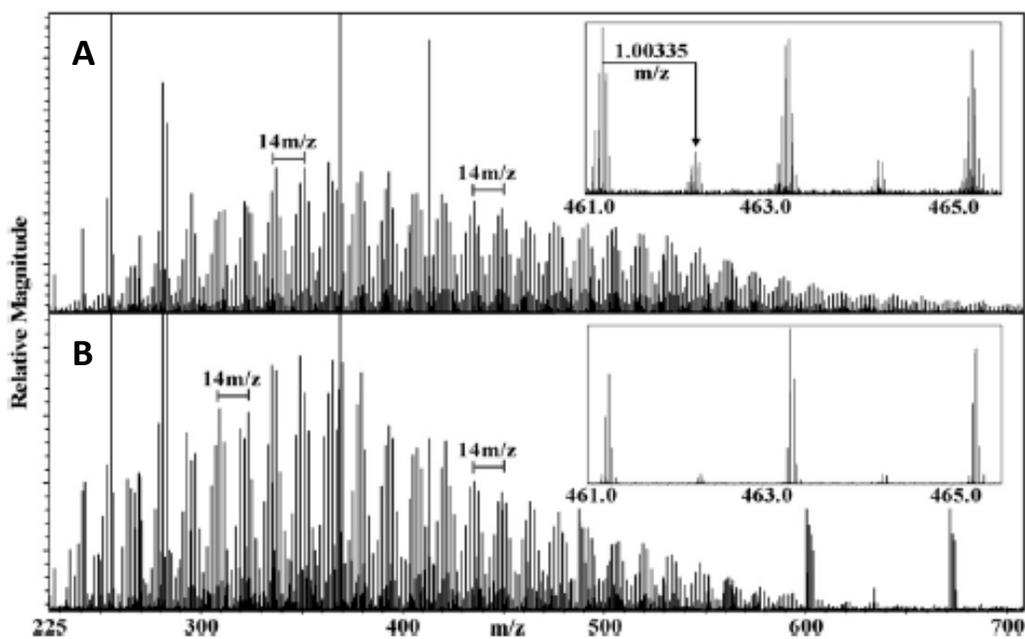


Figure 2.5. ESI FT-ICR mass spectra of DOM extracted from (A) the Dismal Swamp and (B) the Atlantic Ocean off the coast of northern Virginia (adapted from Sleighter and Hatcher, 2008⁶⁷).

Mass distributions. Typical average molecular weights of DOM that are calculated from ESI mass spectrometry typically much lower (300 – 500 Da) than those determined by other methods such as size exclusion chromatography (500 – 4000 Da⁶⁸) or small angle X-ray scattering (500 – 10,000 Da⁶⁹). The discrepancies between the molecular weight estimates of humic substances (HS) from ESI and other bulk methods was formerly attributed to the multiple charging that is commonly observed in ESI. However, as pointed out by Stenson et al.,⁷⁰ ions with high charge states would be associated with peaks corresponding ¹³C-containing isotopologues that are 1.003355/z higher in *m/z* (i.e. difference between the mass of ¹³C and ¹²C divided by the charge). However, pairs of peaks with these mass differences are rarely observed, demonstrating that multiple charging cannot sufficiently account for ESI's bias toward lower apparent molecular weights. Instead, there are two likely explanations for this difference. First, high molecular weight species may not be efficiently ionized under ESI. Second, HS are primarily composed of relatively lower molecular weight species which form aggregates. In this explanation, these HS species remain aggregated when analyzed via bulk methods such as SEC, but are disrupted during the ESI process.

In early work, *m/z* distributions observed via ESI mass spectrometry varied considerably, with some studies displaying distributions of peaks centered near 1000 – 2000 *m/z*. However, the reproducibility and signal to noise ratio (S/N) of these higher *m/z* distributions were generally poor, and were heavily dependent on the data processing methods that were used.^{71,72} Currently, parameters such as transient processing, ionization source and ion optics voltages are routinely optimized to give the greatest sensitivity, resolution, and reproducibility at the lower *m/z* range (< 1000 *m/z*).⁴³

Molecular formula assignment. One of the most common analyses done on an ESI FT-ICR mass spectrum of a DOM sample is the assignment of molecular formulae to as many of the (quasi-molecular ion) peaks as possible. Accurate molecular formula determination is useful for DOM analysis, as different classes of chemical compounds have characteristic ranges of atomic ratios. By analyzing the molecular formulae determined for a DOM sample, researchers can estimate the relative contribution of individual classes of compounds to the entirety of the ionizable species in such samples, and would thus be able to make conclusions based on the overall composition of DOM.

The most rudimentary approach to molecular formula assignment is to generate lists of all chemically reasonable formulae for the given mass range (typically using software such as MATLAB or Microsoft Excel), and to compare the masses of those formulae to a list of measured converted neutral masses (i.e., the measured m/z of a singly-charged, deprotonated species plus the exact mass of H^+). Each converted mass is then matched up to a formulae if the resulting formulae error (FE) is below a pre-defined cutoff, typically defined in parts per million (ppm) as follows

$$(2.3) \quad FE = 1,000,000 \times (M_{theor} - M_{meas}) / M_{theor}$$

Where M_{theor} is the calculated mass of the matched formula and M_{meas} is the measured converted neutral mass of the peak. To achieve reliable molecular formulae assignments, the FE cutoff is usually set at ~ 1 ppm. The lower this maximum error limit is set, the more reliable the molecular formulae assignment is. The lowest FE cutoff which can practically be defined is limited by the mass accuracy of the instrument being used.

Fortunately, however, FT-ICR and even some orbitrap mass spectrometers are capable of

routinely achieving mass accuracies less than 1 ppm at the m/z ranges that are used for DOM analysis, especially when internal calibration routines are used.⁷³⁻⁷⁶

In addition to the allowed FE cutoff that is used for formula determinations, another variable (or set of variables) that can have dramatic effects on the success of the molecular formula assignments is the number of atoms of each element that are used to calculate all possible formulae. Generally, the percentage of peaks in a given mass spectrum which are assigned molecular formula increases with the number of elements that are allowed. However, the inclusion of more elements also leads to a greater number of formulae which fall within a given FE cutoff of a single measured mass. For example, if formulae containing only C, H, and/or O are considered, approximately 50% of all peaks in a typical DOM sample can be successfully assigned formulae (Figure 2.6). Although this percentage can be dramatically increased to ~90% or more if formulae containing N, S, and/or P are included, the number of formulae which fall within 1 ppm of the measured mass increases as well (~15 or more!), especially at higher masses.⁷⁷

To offset this disadvantage, many researchers choose to allow for small numbers of non-oxygen heteroatoms (e.g. $N \leq 2$, $S \leq 2$ ^{78,79}). This typically results in a balance between minimizing equivocal assignments (i.e. assignments in which more than one formula is possible at a given FE cutoff) and analyzing the majority of chemical compounds. When higher numbers of heteroatoms are used, an extra step to choose the most reliable formula must be defined in order to minimize incorrect formula assignments. Although assigning the formulae whose mass deviates the least from the measured mass may be a possible means of choosing a single formulae, this is not necessarily reliable. A more widely accepted method selecting formulae is to use the

formula with the least number of heteroatoms.⁹ Another method of choosing the best formulae for peaks with more than one possible formulae which was recently proposed is to consider the range and frequency of DBE – O (count of double bond equivalent minus number of oxygen in a molecular formula) of all unequivocal assignments, and selecting the formulae with DBE – O values that fall within that range.⁸⁰

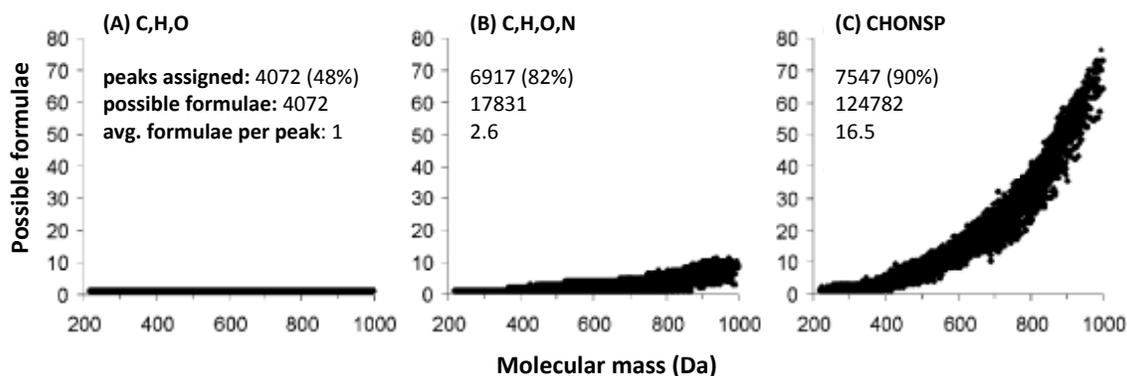


Figure 2.6. Number (and percentage) of peaks with at least possible molecular formulae, the respective sum of all possible assignments, and average number of formulae per peak (total number of possible formula divided by number of peaks assigned) for a negative ion ESI FT-ICR mass spectrum of Suwannee River fulvic acid (8429 total peaks with S/N > 3; 5029 at odd nominal masses, 3400 at even nominal masses) for three different sets of the number of elements used for formula determination: (A) $C_{0-\infty}H_{0-\infty}O_{0-\infty}$, (B) $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-30}$, and (C) $C_{0-\infty}H_{0-\infty}O_{0-\infty}N_{0-30}S_{0-2}P_{0-2}$. Mass accuracy < 1 ppm (adapted from Koch et al., 2007⁷⁷).

An alternative approach to molecular formula assignment takes advantage of the common mass spacing patterns that are inherent in the ultrahigh resolution mass spectra of DOM (Table 2.1). In this method, the molecular formulae for low mass peaks (< 500 m/z) are calculated as described previously. Next, unassigned peaks with masses differing from each assigned peak by one of the common mass spacing patterns are then searched

for and assigned an appropriately derived molecular formula. Although this “chemical building block”⁸¹ approach is less common than the one described previously, a similar methodology is oftentimes adopted to identify ¹³C and ³⁴S containing isotopologues.⁷⁸ Also, analyzing a mass spectrum for a given homologous alkylation series (i.e. peaks whose molecular formulae only differ by the numbers of CH₂ units) is useful for internal calibration routines.^{82,83}

Table 2.1. Mass spacing patterns observed in the ESI mass spectra of DOM.

| <i>m/z</i> difference | molecular formula difference |
|-----------------------|----------------------------------------------------------------|
| +14. 015 650 | + CH ₂ |
| +0. 036 385 | + CH ₄ – O |
| +2. 015 650 | + H ₂ |
| +0. 995 249 | + N – CH |
| + 1. 003 355 | + ¹³ C – ¹² C |
| + 1.995 796 | + ³⁴ S – ³² S |
| + 0.003 371 | + ³² SH ₄ – ¹² C ₃ |

Visual representation of data (Van Krevelen plots). Ultrahigh resolution ESI mass spectra of DOM can be extremely complicated, oftentimes comprising thousands of peaks from which molecular formulae can be calculated. Due to the overwhelming amount of data that is obtained from a single spectrum, mass spectral data must be visually represented in a compact manner in order to allow for any concise and meaningful comparisons between samples to be made. Perhaps the most popular method for visually representing molecular formula data of complex samples such as DOM is to construct a scatter plot comprising the O/C (x-axis) versus H/C (y-axis) molar ratios of each calculated molecular formulae, where each single point corresponds to the

molecular formula of a single resolved peak. This type of plot is known as a Van Krevelen plot, and was first used as a way to compare the elemental analyses of coal and petroleum samples.⁸⁴

The main utility of using such plots is that, as mentioned previously, major classes of biologically derived compounds have characteristic ranges of H/C and O/C molar ratios. This allows for the concise depiction of broad compositional differences between samples. For example, Figure 2.7 shows the Van Krevelen plots of molecular formulae calculated for ESI FT-ICR mass spectra of DOM collected from two local aquatic environments within the Lake Superior watershed: The Toivola Swamp and Brule River.⁸² From these plots, it is apparent that the Toivola Swamp sample has a greater abundance of species with low O/C (<0.3) and H/C (<0.5), which corresponds to compounds which are most likely condensed hydrocarbons. Conversely, a higher number of species at high O/C (>0.6) and H/C (>1.0) are present in Brule River. These O/C and H/C molar ratios are typical of a wide range of biologically derived compounds such as carbohydrates, proteins, and tannins.

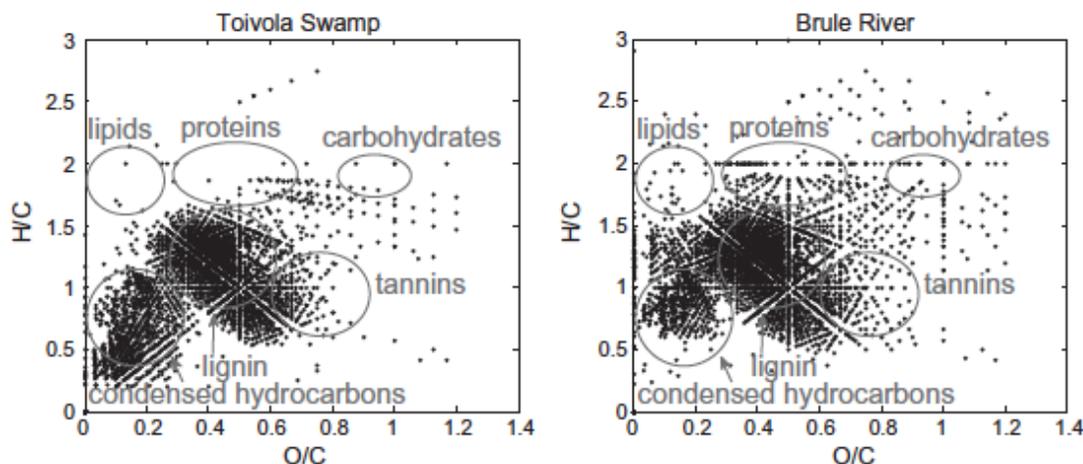


Figure 2.7. Van Krevelen plots of the molecular formulae (obtained using negative ion ESI FT-ICR mass spectrometry) within DOM collected from the Toivola Swamp and Brule River (adapted from Minor et al., 2012⁸²).

Another attractive feature of Van Krevelen plots is that, the common chemical transformations and relationships between compounds in DOM (such as the ones described in Table 2.1) are visually apparent. For example, upon close inspection, many data points can be placed on one of several imaginary lines which converge at the point $(O/C, H/C) = (0, 2)$. Here, all points which fall on one such line are species which differ from one another only by the number of CH_2 structural units (for species with the same number of non-oxygen heteroatoms). Other such series can be visually identified on horizontal lines, vertical lines, and lines which converge at the origin $(O/C = 0, H/C = 0)$. These three sets of series correspond to varying degrees of oxidation/reduction, hydrogenation/dehydrogenation, and hydration/condensation, respectively.

As with any type of reduction of a large data set to a simplified representation, some information is lost when plotting molecular formulae of DOM on a Van Krevelen plot. Namely, the m/z and relative intensity of the peaks corresponding to the plotted molecular formulae are not explicitly shown. Furthermore, it is possible for multiple formulae to have identical O/C and H/C molar ratios. However, it is possible to ameliorate some of the loss of information by adding a third dimension to represent relative abundances. Since 3-dimensional projections are oftentimes difficult to read in 2-dimensional space (i.e. in print), this third dimension is usually represented by varying the color of each data point (Figure 2.8, left). Alternatively, relative intensity can be represented relatively effectively by using varying sizes (point diameter or area) of data points (Figure 2.8, right), where larger points represent the formula of peaks with high relative intensities. An advantage of this type of representation is that the use of multiple data point colors can be used to represent differing like series of formulae. For instance, formulae containing only C, H, and O can be plotted as using one color, while formulae containing C, H, O, and N can be plotted with another color.

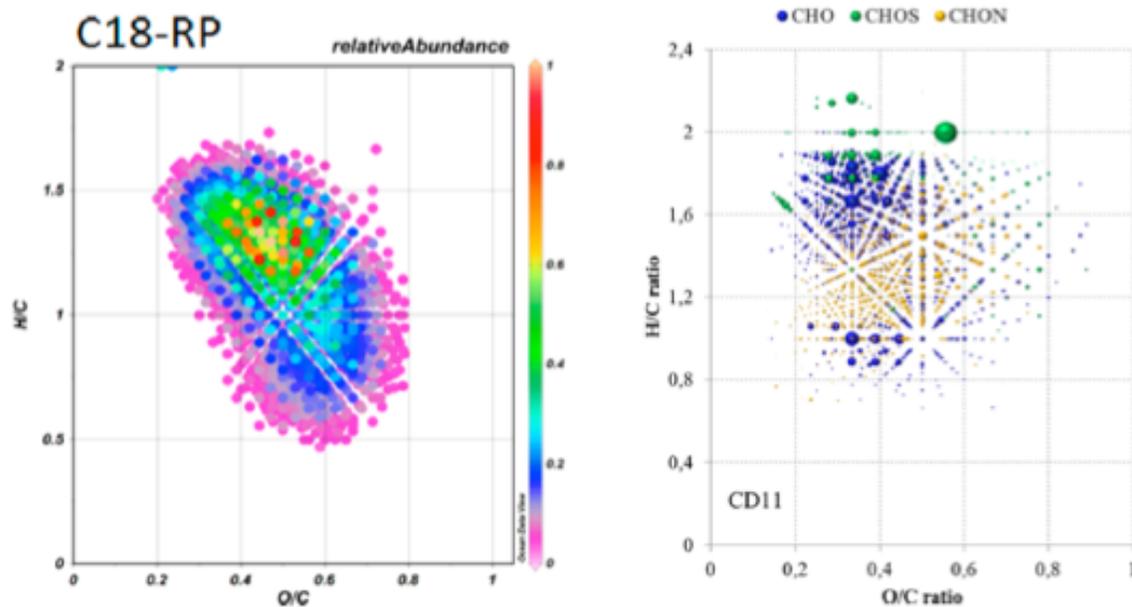


Figure 2.8. Possible improvements on the representation of Van Krevelen plots for the clarification of compositional differences: Relative intensities displayed using a color plot (left; adapted from Perminova et al., 2014⁸⁵), and different sizes of data points (right; adapted from Yekta et al., 2012⁷⁹).

In summary, ESI FT-ICR mass spectrometry has emerged as an indispensable tool for advancing the current understanding for the structure and composition of DOM.³⁹ Its major disadvantage, however, is that it cannot differentiate between isomers. This disadvantage can be ameliorated when combined with other analytical techniques such as chromatographic separations,^{86,87} tandem mass spectrometry, and mass labeling methods.^{88,89} Many new discoveries concerning the composition and dynamics of DOM continue to be made with FT-ICR mass spectrometry, and the continuation of such progress is contingent upon the development of methods which aim to resolve the isomeric complexity of the thousands of mass spectral peaks within an ESI FT-ICR mass spectrum.⁹⁰

Chapter 3: Development of a mass labeling method for identifying ketone/aldehyde-containing species in DOM

(Adapted from 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*, 9891-9897.⁹¹)

3.1. Introduction

Over the past decade, electrospray ionization (ESI) Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry has emerged as an important tool³⁹ for studying how the composition of dissolved organic matter (DOM) differs due to geographic locale^{47,82} as well as how it is affected by natural processes such as photodegradation.^{5,92,93} ESI's ability to ionize polar organic compounds with little to no fragmentation, combined with the unrivaled resolution and mass accuracy of FT-ICR, affords the capability to resolve and assign accurate molecular formulae to thousands of ionizable compounds within DOM.

As new techniques to study DOM are developed, the specific roles that certain chemical moieties play in the overall physicochemical behavior of such ensembles are becoming better understood. For example, carboxylic acids and phenols have long been known to be major constituents within aquatic and terrestrial humic substances (HS) and have been shown to be important in metal ion binding, humic aggregation, and other chemical reactions.^{15,21,41,51,94} Other prominent organic moieties that have been identified

in humic substances include (aromatic) ketones⁹⁵, quinones, and hydroquinones, of which the latter of two have been shown to be involved in the redox behavior of DOM.^{96,97}

Previous work has shown that treatment of chromophoric DOM (CDOM) with sodium borohydride, a selective reductant of ketones, aldehydes, and quinones, produces a substantial loss of absorption at the visible wavelengths and enhanced, blue-shifted fluorescence emission, which has been interpreted to result from the elimination of charge-transfer interactions between carbonyl-containing electron-acceptors and aromatic electron donors.^{3,37,38} Other studies have shown that borohydride reduction also affects many photochemical properties of these materials and has been employed to investigate the mechanisms of these reactions.^{20,21,37}

Although bulk structural techniques such as nuclear magnetic resonance (NMR) and Fourier transform infrared (FTIR) spectroscopy can be used to study the overall presence of ketones and aldehydes within CDOM ensembles⁹⁵, such techniques do not allow for identification of the specific, individual molecules that contain such moieties. However, selectively reducing ketone and aldehyde functional groups with sodium borodeuteride would allow one to readily detect and identify specific compounds via ESI mass spectrometry, owing to the unique mass defect of deuterium. In this chapter, initial experiments are described which employed ESI FT-ICR mass spectrometry to identify species in a standard fulvic acid which contain ketones and aldehydes.

3.2. Materials and Methods

Samples, reagents, and materials. Suwannee River fulvic acid (SRFA) was purchased from the International Humic Substances Society (catalog number: 2S101H). Sodium borohydride (NaBH_4), sodium borodeuteride (NaBD_4 ; 98% D, 90% purity), and Sephadex G-10 (40 – 120 μm particle size) were purchased from Sigma Aldrich. A C_{18} solid phase extract from the upper Delaware Bay was collected in December 2006 at coordinates 39.633, -75.574 (DEUB) and processed as previously reported by Boyle et al. (see Appendix 7 for details).³² Water was obtained from a Milli-Q Academic water purification system (Millipore).

Chemical reduction of DOM. A Shimadzu 2401-PC spectrophotometer was used to monitor changes in absorption during reduction. Two (3.0 mL) aliquots of an aqueous 1 mg/mL SRFA solution was adjusted to pH 7 with NaOH, placed in a 1 cm quartz cuvette capped with a vented septum, and purged with ultrapure nitrogen for 15 min. The cuvettes were opened and a 5-fold mass excess (relative to SRFA) of NaBH_4 or NaBD_4 was quickly added, dissolved by nitrogen bubbling, and allowed to react with the sample in the (resealed) cuvettes for 24 h in the dark under nitrogen, and then open to air for 2 h. The reduced samples, as well as an untreated aliquot of the original SRFA stock were passed through a 2.5 cm x 5 cm Sephadex G-10 column equilibrated with Milli-Q water to remove borate salts, thus adjusting the final pH back to ~ 7 . Eluates were stored at 0° C in the dark until further analysis by mass spectrometry. From this procedure, three aliquots were produced: untreated (UNT), borohydride-reduced (BDR), and borodeuteride-reduced (BDR). Reduction of DEUB was performed in a similar manner by Kelli Golanoski. To approximate the concentration (in mg/L), the stock

extract was diluted with water to match the absorption of the sample at 350 nm to that of a 1 mg/mL SRFA stock, as described elsewhere¹⁹.

ESI FT-ICR MS data acquisition and processing. Ultrahigh resolution mass spectra of all samples were obtained using a hybrid 7 T linear ion trap (LTQ) FT-ICR mass spectrometer (LTQ FT Ultra, Thermo Electron Corp.) at the Woods Hole Oceanographic Institution, with the assistance of Dr. Melissa Kido-Soule. Immediately prior to analysis, thawed samples were diluted in HPLC-grade methanol to a final solvent composition of 50:50 water/methanol. Data were collected in negative ionization mode from 200 – 1000 m/z using a spray voltage of 3.6 kV, capillary voltage of -17.50 V, capillary temperature of 265 °C, flow rate of 4 $\mu\text{L}/\text{min}$, and 206 co-added scans. An average resolving power of about 400,000 was achieved.

Signal processing, internal calibration, and peak alignment was done by Dr. Krista Longnecker. Signal transients were averaged together and processed using code provided by Southam et al.⁷⁶ For each sample, individual transients were included in the average if their total ion current (TIC) was at least 20% of the highest TIC in the scans. The resulting average transient was Hanning apodized, zero-filled once, and fast Fourier transformed. A peak was considered to be detected if its signal-to-noise ratio (S/N) was greater than 5, and the generated peak lists were internally calibrated using a list of known internal standards (Appendix 1, Table A1.1) to achieve a mass accuracy of < 1 parts per million (ppm). The individual peak lists were aligned using code developed by Mantini et al.⁹⁸, which created a single master list comprising m/z and intensity data for all peaks detected in the acquired spectra, where peaks within 1 ppm were considered to be the same. All other data analyses were done using a custom written MATLAB

function (Appendix 9, Section A9.1). Only the peaks below 600 m/z were used for further analysis, since very few peaks were observed above this m/z .

Molecular formula assignment. Molecular formula containing ^{12}C , ^1H , and ^{16}O were calculated for the detected masses on the basis that such formulae should account for the majority of the most intense peaks in the ESI mass spectra of SRFA (and other DOM extracts)⁹⁹ and that unequivocal identification (i.e., only one such possible formula per mass) is possible below about 600 m/z given a 1 ppm error window.⁷⁷ This formula assignment was done in two steps: First, formulae of the form $^{12}\text{C}_c^1\text{H}_h^{16}\text{O}_o$ were determined for each converted neutral mass (i.e., measured m/z + exact mass of H^+) using the following requirements: Formula error ≤ 1 ppm, $1 \leq c \leq m/z / 12$, $0.2 \leq h \leq 2c + 2$, $0 \leq o \leq c$ and number of double bond equivalents ($\text{DBE} = c + 1 - h/2 + n/2$) is an integer. Next, isotopologues containing a single ^{13}C atom were identified by finding peaks with masses 1.003355 m/z (i.e. mass of ^{13}C minus that of ^{12}C) higher than the peak to which a molecular formula has been assigned.

Expected mass shifts upon reduction. If a species of mass M in an untreated sample contains n borohydride-reducible moieties (i.e., ketones and/or aldehydes), then reduction by NaBH_4 and NaBD_4 should result in the generation of new species at masses $m+2.015650n$ (from the addition of $n\text{H}_2$) and $m+3.021927n$ (from the addition of $n\text{HD}$), respectively. Since a typical ultrahigh resolution mass spectrum of DOM contains an abundance of peaks whose m/z differ by 2.015650, unequivocal assignment of $m+2.015650n$ masses as borohydride-reduced species would be practically impossible due to the presence of isobaric species. However, $m+3.021927n$ masses should be unique to the mass spectra of the borodeuteride-reduced sample and can thus be used as reliable

mass markers for species containing ketones and aldehydes in the corresponding untreated sample owing to the negligible natural abundance (< 0.02%) of deuterium.

Although quinone moieties would also be reduced by sodium borohydride, the resulting hydroquinones that are produced are expected to rapidly re-oxidize back to their original form³⁷ and thus should not affect the mass spectrum of a reduced sample.

Furthermore, protons on hydroquinones would be readily exchangeable with water, thus making these species unidentifiable by borodeuteride reduction. As a consequence, the mass shifts described above are only expected to be observed for ketones/aldehydes.

Identification of species containing ketones and/or aldehydes. Peaks in the untreated mass spectrum of a sample (at mass m) to which a molecular formula was assigned was identified as comprising species with n ketone/aldehyde functional groups if a peak at $m+3.021927n$ was present (± 1.0 ppm) in the subset of new peaks that appeared following borodeuteride reduction, where $n = 1$ or 2 . As an additional constraint, the molecular formulae to that peak was required to have a least n O atoms and DBE. Since most peaks at even m/z were of low relative intensity, only peaks at odd m/z were considered to be identifiable as ketones/aldehydes. This, however, is not expected to be a major limitation, since many of the peaks at even m/z comprise ¹³C isotopologues of the higher intensity monoisotopic peaks at odd m/z .

3.3. Results and Discussion

Changes in optical properties upon borohydride and borodeuteride reduction. Reduction of SRFA by NaBH₄ and NaBD₄ produced a similar loss of absorption throughout the UV-vis spectrum with the greatest fractional loss occurring at

long (>400 nm) wavelengths (Figure 3.1). Most of the loss in absorption occurred within 3 h after the addition of NaBH₄ or NaBD₄ and remained after samples were open to air. These observations are consistent with earlier results^{3,19,37} and provide evidence that ketones/aldehydes were reduced selectively and irreversibly. Also, the similar changes in absorption upon treatment by both reducing agents support the assumption that no significant chemical difference between the two reduced samples occurred, aside from the α -H/D substitution on newly generated alcohols, which should not affect the optical properties of these species.

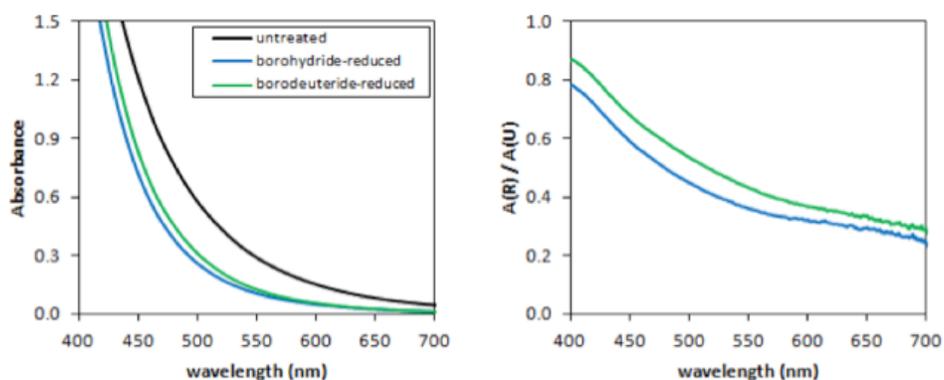


Figure 3.1. UV-vis absorption spectra (left) and fraction of absorbance remaining after reduction (right) of UNT (black), BHR (blue), and BDR (green) SRFA. Spectra were taken prior to purification.

General mass spectrometric features and changes following reduction.

Although all mass spectra were recorded using a 200 – 1000 m/z range, only the range from 200 – 600 m/z were used for further data analysis. This range was chosen for three reasons. Firstly, as described above, molecular formula assignment becomes equivocal (i.e., more than one possibility) at $m/z > 600$ given a 1 ppm error. Secondly, the use of the S/N cutoff of 5 resulted in an apparent truncation of peaks at $m/z > 600$ in BDR,

which was due to the intensities these of peaks falling below this cutoff (Figure 3.2). Therefore, it would be impossible to identify ketone/aldehyde-containing species above this m/z . Thirdly, the majority of peaks and TIC in all spectra resided in this region (Appendix 2, Table A2.1), thus demonstrating that this mass spectral region is representative of most major ionizable species within the sample.

The mass spectra of untreated (UNT), borohydride-reduced (BHR), and borodeuteride-reduced (BDR) SRFA all displayed features that have been observed in previous studies,^{45,67,70,100,101} such as an overall distribution of peaks centered near 400 – 500 m/z , periodic ~ 14 m/z -wide intensity distributions (from species which differ only by the number of CH_2 units), and several peaks per nominal mass (Figure 3.3). Following borohydride reduction, the repeating 14 m/z wide intensity distributions were shifted 2 m/z higher relative to that of the untreated sample, thus suggesting a reduction of major peaks throughout the entire m/z range. In contrast, this intensity pattern was shifted 3 m/z higher following borodeuteride reduction. Additionally, the alternation of intensities between the clusters of peaks at odd and even m/z observed in UNT and BHR was not observed in BDR. These changes are consistent with reducible species with large relative intensities (mostly at odd m/z) being shifted to even m/z following borodeuteride-reduction via the expected $m+3.021927$ mass shifts.

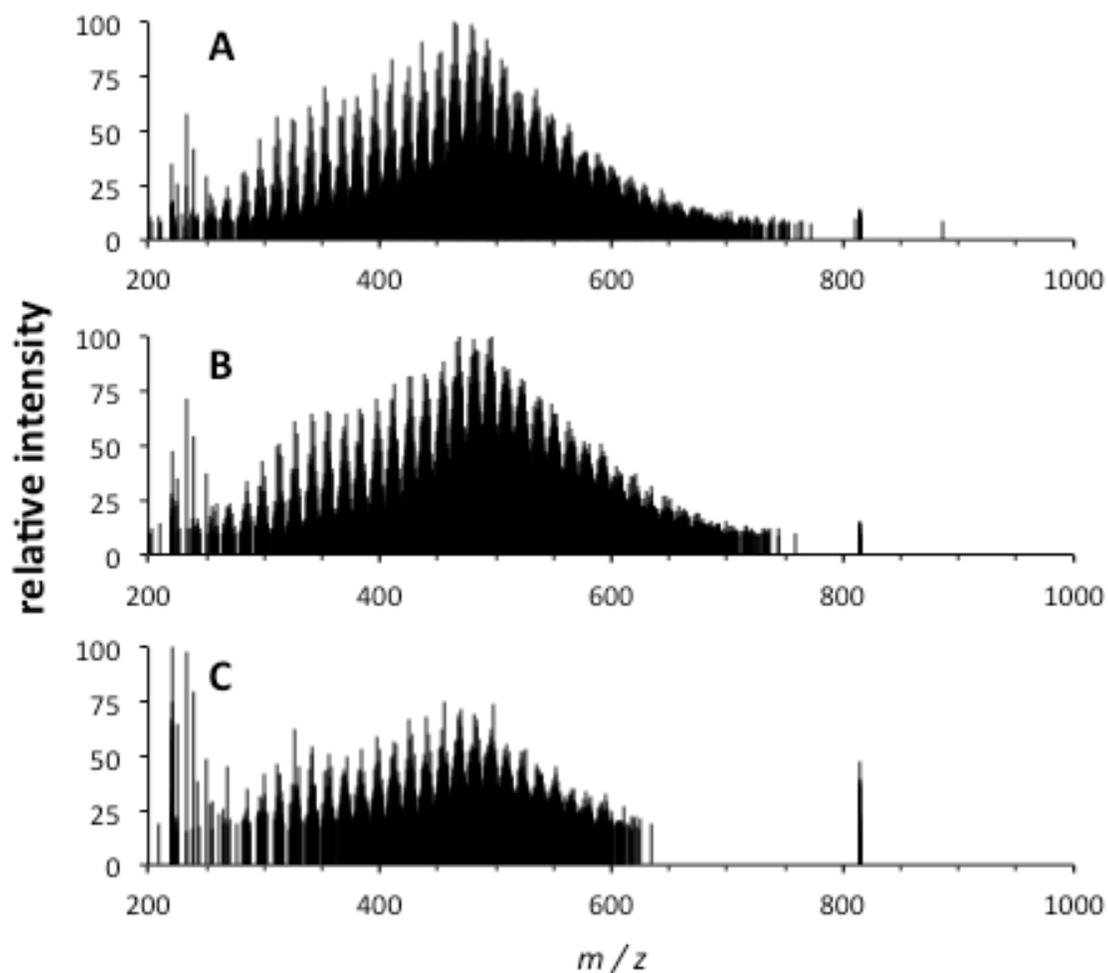


Figure 3.2. Full scan negative ion ESI mass spectra of (A) UNT, (B) BHR, and (C) BDR SRFA. The apparent truncation of peaks at about 620 m/z is due to most peaks at higher m/z falling below the S/N cutoff. The peak cluster at 814 m/z was of unknown origin and were present in all spectra in this study acquired using the 7 T ESI FT-ICR MS (at the Woods Hole Oceanographic Institution).

The substantial loss of ion intensities across the entire BDR spectrum relative to UNT and BHR can be partially explained by these expected mass shifts, assuming that the ionization efficiency of most species are not greatly effected by the reduction of an aldehyde or ketone moiety to a corresponding alcohol. During borohydride reduction, reduced species are simply shifted to masses that are the same as other non-reducible

species. If this transformation is assumed to occur throughout the entire mass range, it is expected that much of the loss of intensity at a given mass m from reduction will be counteracted by the addition of reduced species that initially had a mass $2.015650 m/z$ lower in the untreated sample. During borodeuteride reduction, reduced species are spread out onto new masses, rather than being superimposed onto peaks that are already present in the mass spectrum of the untreated sample. As a result, the loss of intensity at any given mass m from reduction would *not* be counteracted by the formation of new isobaric (reduced) species, and hence the relative intensity of each peak, and thus the entire mass spectrum, would decrease.

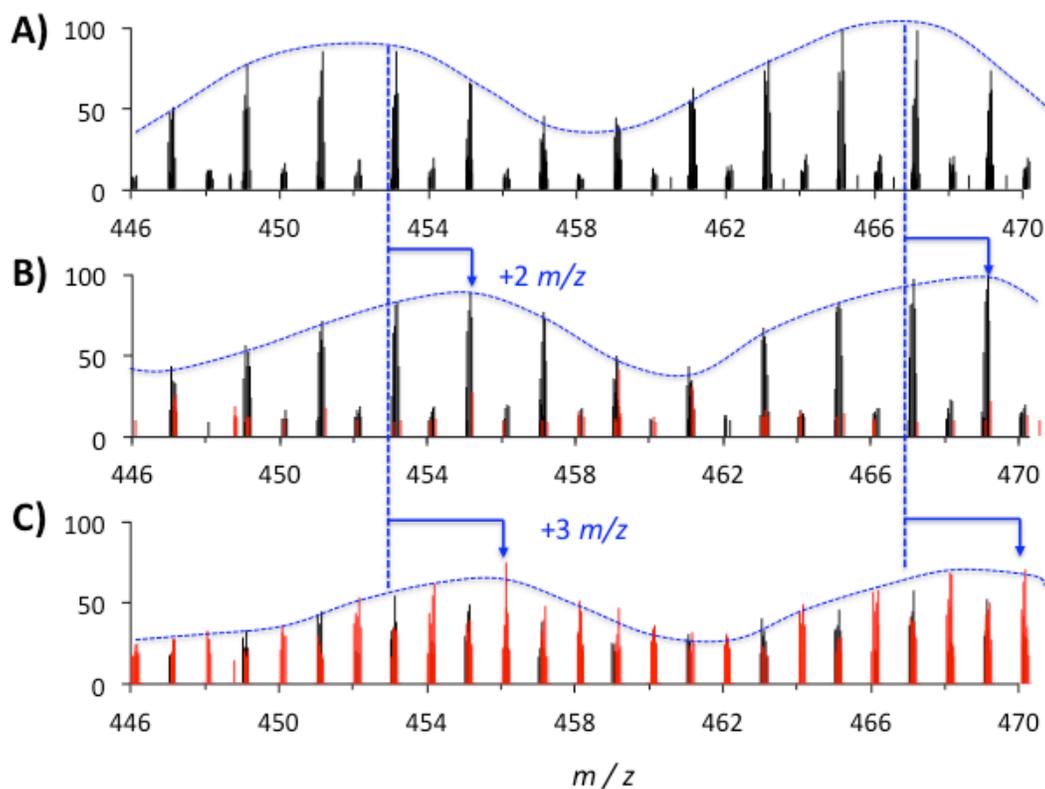


Figure 3.3. Expanded region from 446 – 470 m/z of the negative ion ESI 7 T FT-ICR mass spectra of (A) UNT, (B) BHR, and (C) BDR SRFA. Peaks in red are those which were not present in UNT SRFA.

To show more clearly the occurrence of the aforementioned exact mass shifts, the five most intense peaks at 467 m/z in UNT were considered (Figure 3.4; peaks A – E; see Appendix 3 for peak lists). In UNT, masses corresponding to the addition of H₂ (+2.015650 m/z ; peaks A' – E') to these precursor (M) masses were readily observed at 469 m/z , whereas no peaks corresponding to the addition of HD (+3.021927 m/z) were found at 470 m/z within the 1 ppm error limit, as expected. However, after borodeuteride reduction, new peaks corresponding to the addition of HD for all five peaks appeared at 470 m/z (peaks A'' – E'') in BDR, with intensities similar to parent masses. Based on the appearance of peaks at $m+3.021927$ following borodeuteride reduction, and the changes in relative intensity of peaks at $m+2.015650$ following borohydride reduction, it is evident that all five major peaks at 467 m/z arise (at least in part, since several isomers are possible for each detected peak) from species containing at least one reducible group. Indeed, the lack of complete losses of peaks A – E following reduction implies the presence of isobaric species which do not contain reducible groups and/or incomplete reduction.

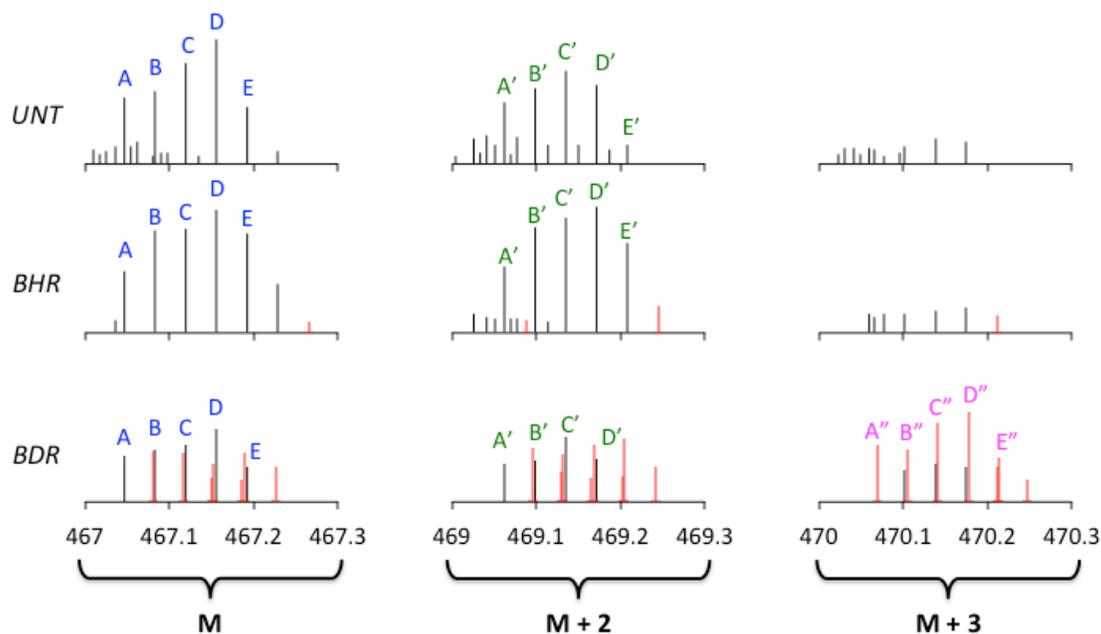


Figure 3.4. Mass spectra of UNT, BHR, and BDR SRFA at 467 (left; M), 469 (middle), and 470 (right) m/z . Red peaks are those that were not present in UNT.

Peak reproducibility and molecular formula assignments. Figure 3.5 shows three Venn diagrams comparing the peaks detected in UNT to those detected in the replicate mass spectrum (UNT-2) and to the two reduced samples, along with the percentages of peaks to which a molecular formula was assigned. Of all peaks detected in the two untreated spectra, 73% (2138 peaks divided by 2930 peaks in one *or* both replicates) were identified in both replicates, which is similar to the reproducibility of mass spectra of identical DOM samples reported previously.^{43,58} The commonality of peaks in UNT and BHR was 49%, which can be expected from a slight chemical alternation of species in DOM, whereas the commonality of peaks in UNT and BDR was much lower (17%), thus demonstrating that borodeuteride reduction results in a large number of *unique* masses. Similar relationships between the mass spectra for DEUB (Figure 3.6).

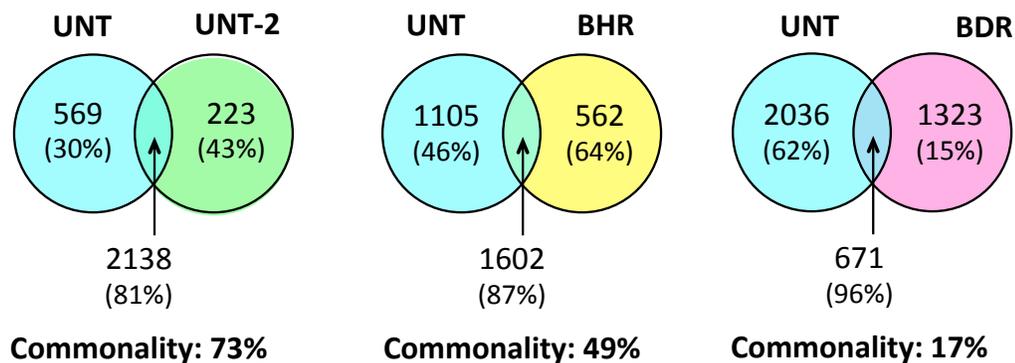


Figure 3.5. Venn diagrams comparing the number of peaks detected in the mass spectrum of SRFA UNT to those in an untreated replicate (UNT-2), BHR, and BDR mass spectra. Percentages of peaks belonging to each subset to which molecular formulae were assigned are given in parenthesis. Below each Venn diagram is the percentage of all species in the two spectra that were in *both* spectra.

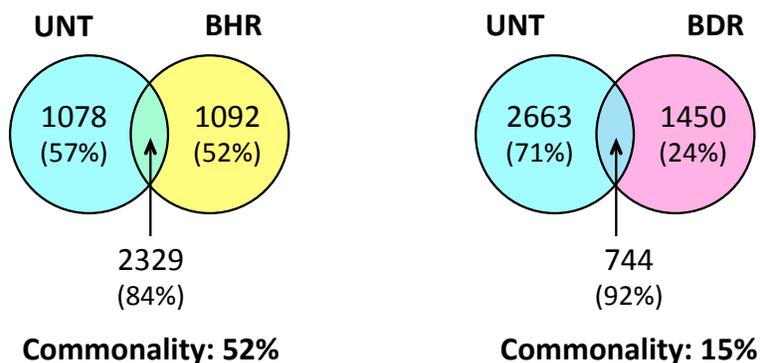


Figure 3.6. Venn diagrams comparing the number of peaks detected in the mass spectrum of DEUB UNT to those in the corresponding BHR, and BDR mass spectra. Percentages of peaks belonging to each subset to which molecular formulae were assigned are given in parenthesis. Below each Venn diagram is the percentage of all species in the two spectra that were in *both* spectra.

Molecular formulae were successfully assigned to the majority of peaks in the untreated replicates and the borohydride-reduced mass spectra (70 – 80%). Moreover, the percentage of peaks which were assigned molecular formulae were greater for peaks which were present in more than one spectrum. In contrast, only 42% of all peaks in SRFA-BDR were successfully assigned CHO-containing molecular formulae, which is to be expected since many species in this sample were expected to contain deuterium atoms. In fact, when the molecular formula algorithm used for this work was altered slightly to allow for up to two deuterium atoms, the percentage of peaks in BDR to which a molecular formulae assigned increased dramatically to 97%, with very little changes in the percentage of assigned peaks for UNT and BHR (data not shown).

Reliability of $m+3.021927n$ mass shifts as unique markers of ketones/aldehydes. In order for $m+3.021927n$ masses to be used as reliable markers the identification of ketone/aldehyde-containing species in DOM samples, such masses must not be observed in the untreated sample (i.e., are unequivocal mass markers) *and* all masses that appear following borodeuteride reduction must be identifiable as $m+3.021927n$ masses (where n is an integer), thus demonstrating that no other chemical change occurred. To test this, all possible $m+3.021927n$ masses were calculated from the mass of each peak in the untreated samples at odd m/z to which a molecular formula was assigned using values of n ranging from 1 to 3. The peak lists of the untreated and borodeuteride-reduced mass spectra were then searched for these expected masses within a 1.0 ppm mass window.

When $n = 1$, less than half (48.3% and 38.6% for SRFA and DEUB, respectively) of the new peaks appearing after borodeuteride reduction were identified as $m+3.021927$ masses (Table 3.1). When $n \leq 2$, these percentages increased nearly two-fold (83.5% and 71.4% for SRFA and DEUB, respectively). While including $n = 3$ mass shifts in this search resulted in slightly larger percentages of the new peaks identified, a substantial percentage ($\geq 15\%$) of masses in the untreated sample were also identified as $m+3.021927n$ masses. Although these data show that $m+3.021927n$ masses are only reliable markers for species with one or two reducible groups (with the currently achieved instrumental resolution and mass accuracy), the majority of peaks appearing after borodeuteride reduction were identified as being either $m+3.021927$ or $m+6.042854$ masses, suggesting that species with more than two ketones/aldehydes are relatively uncommon in these samples. Also, it is possible that some of the unidentified peaks in the borodeuteride-reduced mass spectra are indeed $m+3.021927$ and/or $m+6.042854$ masses from species in the untreated sample which had S/N ratios lower than the employed cutoff, and were therefore not present in that mass spectrum.

Table 3.1. Percentage of peaks in the mass spectra of untreated and borodeuteride-reduced SRFA and DEUB identified as $m+3.021927n$ ($n = 1, 2$, and/or 3) masses.

| subset | sample | $m+3.021927n$ (% of subset) | | |
|------------------|--------|-----------------------------|------------|------------|
| | | $n = 1$ | $n \leq 2$ | $n \leq 3$ |
| UNT (all peaks) | SRFA | 0 | 1.7 | 15.3 |
| | DEUB | 1.8 | 2.6 | 20.4 |
| BDR (not in UNT) | SRFA | 48.3 | 83.5 | 91.5 |
| | DEUB | 38.6 | 71.4 | 81.8 |

Compositional analysis of reducible species in SRFA and possible relations to its bulk physicochemical properties. The relatively high degree of peak overlap between the untreated and borohydride reduced mass spectra, the similar total numbers of peaks, ion counts, and mass spectral patterns of these two samples, as well as the uniqueness of $M+3.0219n$ masses in the borodeuteride-reduced samples all suggest that chemical modification by NaBH_4 and/or NaBD_4 does not result in any artifacts (e.g. from contamination or non-uniform changes in ionization efficiency) that would interfere significantly with detection and identification of ketone/aldehyde-containing species. To further demonstrate that mass spectral artifacts from sodium borohydride was not an issue, ESI mass spectra of an unreduced, borohydride-reduced, and borodeuteride-reduced model compound (acetovanillone) were collected using orbitrap mass spectrometry. No evidence for large differences in ionization efficiency of the unreduced and reduced forms nor substantial contamination that would be present in the employed mass range was observed (Appendix 4). Therefore, the search method employed here is expected to be quite reliable.

This search identified 48.5% and 34.4% of all peaks with molecular formulae at odd m/z in the untreated SRFA and DEUB, respectively, as arising (at least partially) from species with one and/or two ketones/aldehydes (Table 3.2). Most peaks that were identified as arising from species containing two ketones/aldehydes were also identified as arising from species containing one such moiety. That is, following borodeuteride reduction, many peaks which gave rise to a peak at mass $M+6.0438$ from a di-ketone/aldehyde also gave rise to a peak at mass $M+3.0219$ from an isobaric mono-

ketone/aldehyde, which is consistent with either incomplete reduction of di-ketones and/or the isobaric/isomeric complexity of HS samples reported by others.^{52,86,90}

Table 3.2. Percentage of peaks in the mass spectrum of untreated SRFA and DEUB identified as comprising ketone and/or aldehyde-containing species (based on a search for M+3.0219 and M+6.0438 masses in the corresponding borodeuteride-reduced mass spectrum).

| <i>m</i> +3.021927 <i>n</i> in BDR? | | peak composition (of <i>m</i> in UNT) | SRFA | DEUB |
|----------------------------------------|--------------|------------------------------------------|----------------|-----------------|
| <i>n</i> = 1 | <i>n</i> = 2 | | | |
| no | no | non reducible species | 699 (51.5%) | 1190 (65.6%) |
| YES | no | singly-reduced species only | 2 (13.2%) | 143 (7.9%) |
| no | YES | doubly-reduced species only | 22 (1.6%) | 50 (2.8%) |
| YES | YES | singly- & doubly-reduced species | 464 (32.2%) | 437 (24.1%) |

Plots of H/C vs. O/C molar ratios (“Van Krevelen plots”; Figure 3.7) and DBE vs. *m/z* (Figure 3.8) are shown for all formulae identified in the mass spectra of both samples in this study. In DEUB, the number of ketone/aldehyde moieties identified for a given peak clearly increased with increasing *m/z* and DBE – a trend that is not as clear in SRFA. That is, while most species with only one ketone/aldehyde moiety had low *m/z* and DBE, the species identified as having two such moieties resided mostly at high *m/z* and DBE, which would be an expected trend if most ketone/aldehyde-containing species belonged to the same general class of compounds. A more obvious difference between the two samples that is revealed by this method is observed in the species at low H/C and high O/C molar ratios: While the species in this region of the Van Krevelen plot in the

Upper Bay are not identified as containing ketone/aldehyde moieties, they are identified as such in SRFA. Although a detailed characterization of these subsets of molecules is beyond the scope of this study, this specific structural difference between the two samples is only apparent using this mass labeling technique and could not have been deduced from molecular formulae assignments alone.

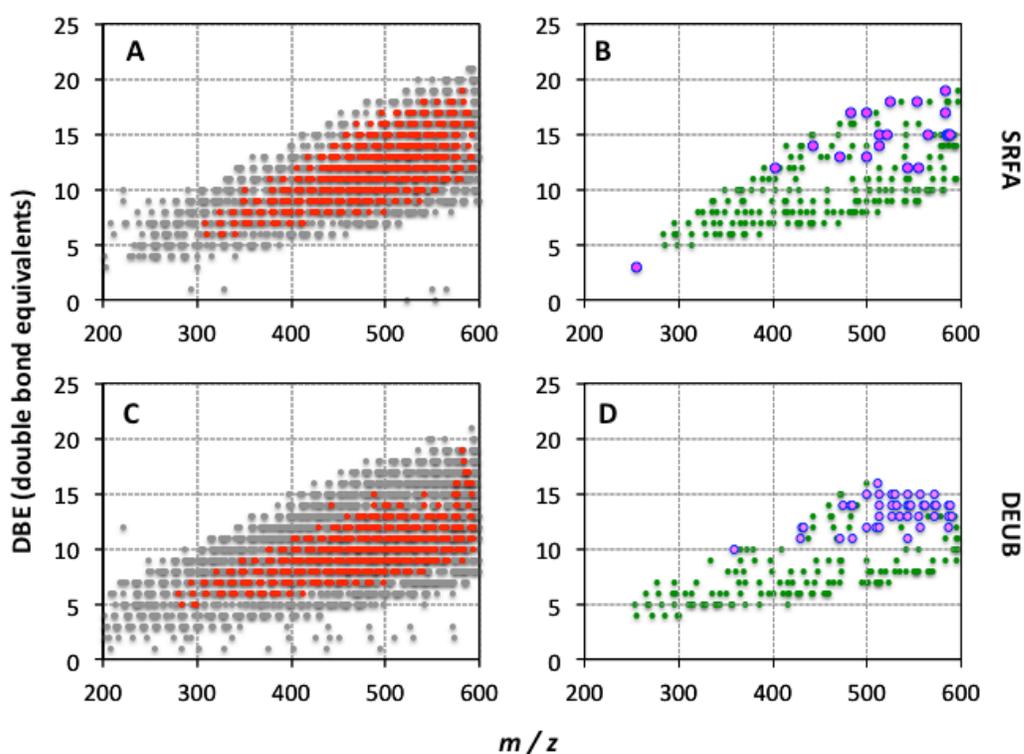


Figure 3.7. Plots of DBE versus m/z of all molecular formulae in untreated SRFA (top) and DEUB (bottom) showing peaks comprising species with no reducible groups (grey), both one and two reducible groups (red, A & C), one group only (green), and two reducible groups only (purple, B & D).

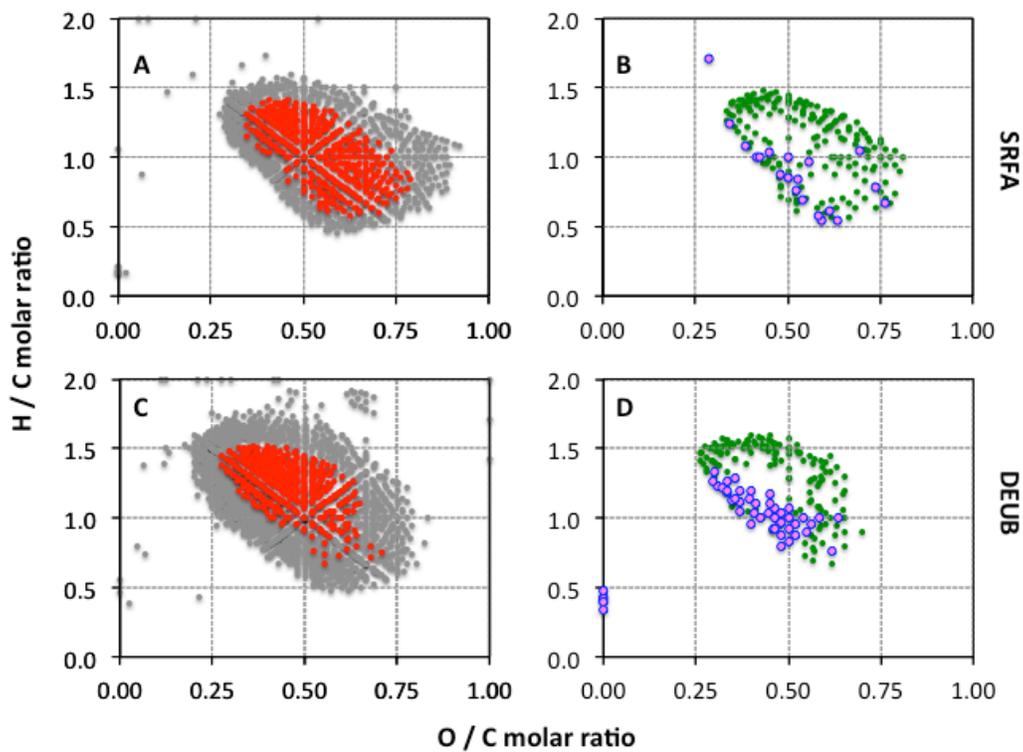


Figure 3.8. Plots of H/C versus O/C molar ratios of all molecular formulae in untreated SRFA (top) and DEUB (bottom) showing peaks comprising species with no reducible groups (grey), both one and two reducible groups (red, A & C), one group only (green), and two reducible groups only (purple, B & D).

The majority of the species identified as containing at least one ketone/aldehyde moiety had H/C and O/C molar ratios, DBE, and molecular weights, which are consistent with and thus commonly attributed to two major compound classes: lignin-derived species and/or carboxyl-rich alicyclic molecules (CRAM).^{5,7,39,67,100} However, it is highly unlikely that CRAM could comprise the majority of the carbonyl-containing species being reduced here. The substantial changes in absorption/emission and photochemical properties^{3,19,21,37,38} that are observed in CDOM/HS cannot be explained by borohydride

reduction of CRAM, since the reduction of the isolated, aliphatic ketones in such species would be expected to cause only slight changes in absorption due to the disappearance of weak $n \rightarrow \pi^*$ transitions at ~ 280 nm (Figure 3.9A). In contrast, lignin-derived structures are likely to contain reducible groups in conjugation with alkoxy- or hydroxy- substituted aromatics, which have relatively strong transitions in the 280 to 330 nm wavelength region (Figure 3.9B). Reduction of the carbonyl would disrupt the extended conjugation and thus result in significant losses in absorption in the UVA. Moreover, the aliphatic ketones within CRAM cannot account for the previously observed wavelength dependence of the photosensitized oxidation of trimethylphenol, nor for the loss of this sensitization under borohydride reduction.^{19,102}

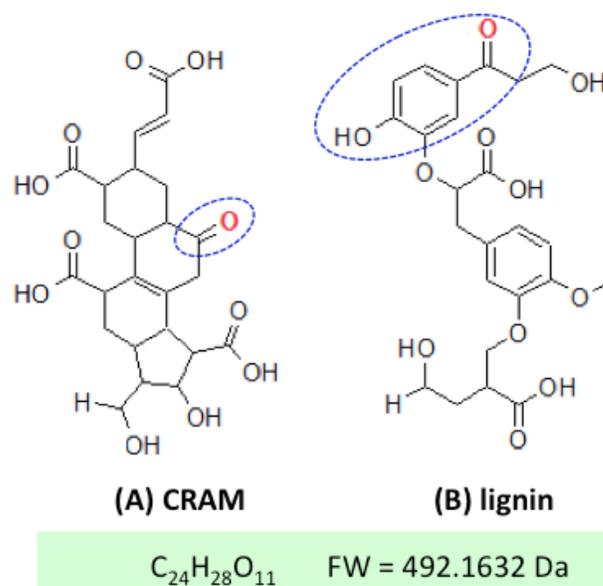


Figure 3.9. Two hypothetical molecular structures for the molecular formulae $C_{24}H_{28}O_{11}$ ($O/C = 0.46$, $H/C = 1.17$) containing one borohydride-reducible group: (A) carboxyl-rich alicyclic molecule (CRAM) and (B) lignin-derived structure.

Although visible absorption (>400 nm) still cannot be explained directly from either lignin derivatives or CRAM corresponding to these low mass (< 600 Da) compounds, aggregation could facilitate charge transfer interactions between electron acceptors (e.g. aromatic ketones/aldehydes and quinones) and donors (e.g. alkoxy aromatics), both of which are known to be present within lignin derivatives.^{22,33} Indeed, previous work^{45,70} has indicated that the ESI process may cause disaggregation of high molecular weight species. The view that humic substances are composed primarily of aggregates of relatively low molecular weight components^{12,13,15,48} is compatible with a model in which inter and intra-molecular charge transfer interactions between lignin-derived molecules of these aggregates give rise to the unique absorption and emission properties of these materials.

3.4. Conclusions

In this chapter, a novel method for discriminating among peaks in the ultrahigh resolution mass spectrum of a DOM sample which comprise species containing zero, one, and/or two reducible moieties was described and validated. In this method, species that contain up to two ketone and/or aldehyde functional groups are identified in the mass spectrum of an untreated DOM sample by searching for $m+3.021927$ and $m+6.043854$ masses in the mass spectrum of a corresponding borodeuteride-reduced DOM aliquot. As the mass accuracy and resolution that can be achieved by mass spectrometry continues to improve, it may be possible to extend this mass labeling technique to accurately identify species containing three borodeuteride-reducible moieties. However, species with $m/z <$

1000 which contain more than two such moieties are expected to be quite rare, and such an extension may not be necessary.

This method allows specific structural information to be gained from ultrahigh resolution mass spectra of DOM, thus extending the technique's overall usefulness in DOM research beyond providing general compositional information based solely on m/z and molecular formula assignments. Furthermore, the isomeric/isobaric complexity of peaks in the mass spectra of DOM, as well as the reactivity of reducible species can be studied by using this technique. In summary, this method shows great promise as a means to further characterize the structure and composition of DOM, especially in identifying specific subsets of chemical species that contribute significantly to the optical and photochemical properties of such samples.

Chapter 4: Reactivity of ketone/aldehyde-containing species in Suwannee River fulvic acid and their contribution to bulk optical properties

4.1. Introduction

In the previous chapter, a method to mass label ketone and aldehyde-containing species in dissolved organic matter (DOM) using sodium borodeuteride (NaBD_4) was developed. One experimental variable that was held constant in the initial demonstration was the extent to which the original DOM sample was reduced. If a sample being analyzed is only partially reduced as a result of insufficient amounts of NaBD_4 , then the mass labeling method would result in an underestimation of the abundance of ketone/aldehyde-containing species within the untreated sample. To address this issue, aliquots of an aqueous stock solution of Suwannee River fulvic acid (SRFA) were reduced with increasing amounts of NaBD_4 and their ultrahigh resolution electrospray ionization (ESI) mass spectra were obtained.

Several experimental and instrumental improvements were incorporated. Firstly, the gel permeation used for the removal of borate salts (which would interfere greatly with the ESI analysis) was replaced with a solid phase extraction (SPE) protocol, a well accepted method of DOM purification and enrichment.¹¹ Secondly, higher resolution, mass accuracy, and sensitivity was achieved using a 12 tesla (T) Fourier Transform ion cyclotron resonance (FT-ICR) mass spectrometer, as opposed to the 7 T instrument that was used in the previous chapter. Thirdly, the major mass spectral peaks that were

present in an extraction blank were subtracted from all sample spectra prior to data analyses, thus minimizing interferences from common contaminants and solvent peaks. Finally, the MATLAB code used in the previous chapter was written only to accommodate the aligned peak list which was used for the specific study. All MATLAB code was substantially revised and written in the form of several discrete functions. These new functions are intended to allow for the automation of this mass labeling technique on other ultrahigh resolution mass spectral datasets of an untreated and borodeuteride-reduced sample.

In addition to implementing these technical improvements, a second goal of this study was to qualitatively investigate how the bulk optical properties of chromophoric dissolved organic matter (CDOM) change with respect to the reduction of the ketone and aldehyde-containing species which are detected by ESI FT-ICR MS by way of the newly developed mass labeling method. This relationship between optical properties, structure, and reactivity was investigated by identifying the subset of ketone/aldehyde-containing species which were reduced with various amounts of borodeuteride and relating those changes to the observed optical properties.

4.2. Materials and Methods

Samples, reagents, and materials. Suwannee River fulvic acid (SRFA) was purchased from the International Humic Substances Society (catalog number: 2S101H). Sodium borodeuteride (NaBD_4 ; 98% D, 90% purity), hydrochloric acid (HCl, trace select) and sodium hydroxide (NaOH) was purchased from Sigma Aldrich. Water was obtained from a Milli-Q Academic water purification system (Millipore). LC grade

methanol (MeOH) was purchased from Fisher, and potassium phosphate mono- and dibasic (KH_2PO_4 and K_2HPO_4) was purchased from Baker. Bond Elut PPL solid phase extraction (SPE) cartridges (100 mg functionalized styrene-divinylbenzene polymer solid phase; part # 12105004) were purchased from Agilent.

Sample preparation. A 0.10 mg/mL stock of SRFA was prepared in Milli-Q water and adjusted to pH 10 with NaOH. This stock was divided into six (12 mL) aliquots within glass vials. To all but one vial, solid NaBD_4 was added to different concentrations, thus achieving varying degrees of reduction. As a blank, 1.9 mg/mL NaBD_4 (the same concentration used for the 20-fold mass equivalent reduction) in pH 10 Milli-Q water was prepared in an additional vial. All vials were loosely capped and placed in the dark at room temperature for 24 h. To all aliquots, 10 M and/or 4 M HCl was added incrementally ($\sim 100 \mu\text{L}$ total) to a final pH of 2.0 ± 0.1 to decompose any residual NaBD_4 , with the aliquots then stored in the dark for 24 h. Descriptions of all samples and their corresponding abbreviations used throughout this chapter are given in Table 4.1.

After the 24 h period, the pH of each aliquot was measured. 10.0 mL of the pH 2 aliquot was then loaded on a PPL cartridge (preconditioned with 2 mL MeOH and 2 mL of 10 mM HCl), and allowed to elute by gravity. The cartridges were rinsed with an additional 3 mL of 10 mM HCl and dried using a steady flow of nitrogen (~ 10 mins). The organic species were then eluted into glass vials using 1.5 mL MeOH, which was capped and then stored at 0°C until mass spectrometric analysis. A small fraction of the pH 2 sample (pre-extraction) was reserved and stored in the dark at 4°C for absorbance and fluorescence analysis.

Table 4.1. Descriptions, pH, and absorption data for untreated (UNT) and all borodeuteride-reduced (BDR) samples.

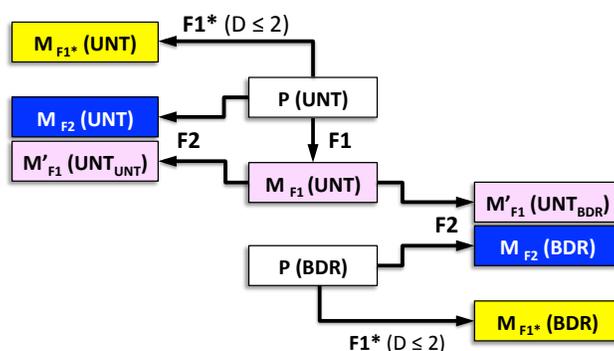
| sample | ME ^a | pH (pre-SPE) ^b | pH (optics) ^c | A(λ) _{BL} ($\times 10^{-3}$) ^d | S _A ^e | res ^f |
|--------|-----------------|---------------------------|--------------------------|----------------------------------------------------------------|-----------------------------|------------------|
| UNT | 0 | 2.09 | 6.75 | 1.5 \pm 0.2 | 0.0152 | 0.0040 |
| BDR-1 | 2.3 | 2.10 | 6.77 | 1.9 \pm 0.2 | 0.0166 | 0.0028 |
| BDR-2 | 6.0 | 2.08 | 6.75 | 0.9 \pm 0.2 | 0.0171 | 0.0020 |
| BDR-3 | 20 | 2.11 | 6.73 | 2.0 \pm 0.7 | 0.0182 | 0.0024 |
| BDR-4 | 58 | 2.08 | 6.67 | 5.3 \pm 0.4 | 0.0185 | 0.0016 |
| BDR-5 | 155 | 1.99 | 6.56 | 4.1 \pm 0.4 | 0.0190 | 0.0011 |

^a Mass equivalents of NaBD₄ (i.e. ratio of [NaBD₄] / [SRFA]); ^b sample pH immediately before SPE; ^c sample pH immediately before absorbance and emission measurements ([SRFA] = 25 mg/L, [phosphate] = 24 mM); ^d average \pm standard deviation of the absorbance between 650 and 700 nm prior to baseline correction; ^e spectral slope for the 300 to 650 nm wavelength range; ^f sum of the square of residuals between the measured and fitted absorbance data.

Mass spectrometric data acquisition and pre-processing. Ultrahigh resolution mass spectra of the untreated (UNT) and borodeuteride-reduced (BDR) aliquots were obtained by Dr. Mourad Harir using a Bruker Apex QE 12 T FT-ICR mass spectrometer at the Helmholtz Center for Environmental Science (Munich, Germany). An Apollo II electrospray ionization source was operated in negative ion mode using a spray voltage of -3.6 kV and 0.3 μ L/min flow rate. Data were recorded from 100 – 2000 m/z , 500 individual scans were averaged, and the resulting average resolution was >500,000. For each spectrum, lists of m/z and intensities of peaks whose signal-to-noise ratio (S/N) was greater than 10 were generated for further analysis and internally calibrated using a list of

common species found within DOM samples (Appendix 1, Table A1.2). Peak list generation and post calibration was done using the Bruker Daltonics Data Analysis 4.0 software by Dr. Michael Gonsior to achieve a mass accuracy of < 0.2 ppm.

Peak list analyses (molecular formula assignment and identification of reducible species). All subsequent data processing and analyses were done using custom written MATLAB functions (Appendix 9), illustrated in Figure 4.1 and described below.



| | |
|------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| P (UNT/BDR) | Peak list of singly charged ions in UNT or BDR. |
| M_{F1} (UNT) | Molecular formulae of peaks in UNT (calculated by F1). |
| M'_{F2} (UNT_{UNT/BDR}) | Molecular formulae of peaks in UNT identified as comprising non, singly and/or doubly-reduced species based on a search for $m+3.021927$ and $m+6.043854$ peaks in UNT or BDR (calculated by F2). |
| M_{F2} (UNT/BDR) | Molecular formulae of peaks in UNT or BDR with D-containing formulae for peaks that are 3.021927 or 6.043854 m/z higher than peaks in UNT to which a molecular formula was assigned (calculated by F2). |
| M_{F1*} (UNT/BDR) | Molecular formulae of peaks in UNT or BDR (calculated by F1*). |

Figure 4.1. Analysis of mass spectral peak lists (white boxes labeled P(UNT) and P(BDR)) by custom-written MATLAB functions (black arrows). Colored boxes represent matrices of molecular formulae calculated by the indicated functions **F1**, **F1***, or **F2**.

Excision of multiply-charged peaks and common contaminants: For each raw peak list, doubly and triply-charged species were identified by searching for peaks in each mass list that varied by 0.501678 and 0.334452 m/z , respectively (i.e., the expected mass difference between doubly and triply-charged species and their ^{13}C -containing isotopologues), within a 0.2 ppm error window. If such a mass difference was found, then the two peaks were excluded from the processed peak list. Also excluded from the sample peak lists were those that resided within 0.2 ppm of the most intense peaks in the mass spectrum of the extraction blank (see Appendix 1, Table A1.3 for peaks removed from the sample mass spectra).

Molecular formula assignment of UNT (F1): Molecular formula assignment of peaks in UNT was accomplished using a custom MATLAB function based on the “low-mass moiety” approach described by Perdue and Green.^{103,104} This algorithm performs hundreds of times faster than a conventional “brute-force” algorithm that cycles through all chemically reasonable combinations of C, H, O, N, and S until a matching formulae is found for each measured m/z value. The new low-mass moiety approach yields identical formulae to those given by a brute force approach with the added benefit of completing molecular formula assignment of a single spectrum in a few minutes rather than in several hours.

Briefly, each measured m/z was converted to its neutral mass (by addition of the mass of H^+) and a base molecular formula of a hydrocarbon with the lowest possible H/C molar ratio was calculated (e.g. C_{28}H_6 for a measured m/z of 341.0878). This base formula was altered by adding or subtracting C_4O_{-3} and/or CH_4O_{-1} subunits until the following requirements were satisfied: (1) calculated mass of the formula was within 0.2

ppm of the measured mass, (2) $c > 0$, (3) $2 \leq h \leq 2c+2$, (4) $o \leq c$, and (5) $o + n + s > 0$ (i.e., final formula is not a hydrocarbon), where c , h , o , n , and s are the numbers of C, H, O, N, and S, respectively. Additionally, non-oxygen heteroatom (N and S)-containing formulae were considered by altering the base hydrocarbon with varying numbers of $N(CH_2)_{-1}$ and/or SH_4C_{-3} subunits, where the maximum numbers of N and S were user-specified inputs. If more than one molecular formula was within this error limit, the formula with the least number of heteroatoms ($N + S$) was used. Afterwards, isotopologues containing a single ^{13}C atom were identified by finding peaks that were $1.003355 m/z$ (i.e., the mass of ^{13}C minus that of ^{12}C) higher than a peak which was assigned a molecular formula. Similarly, ^{34}S -containing isotopologues were identified by searching for peaks that were $1.995980 m/z$ (i.e., the mass of ^{34}S minus that of ^{32}S) higher than a peak that was assigned a ^{32}S -containing formula in the first step.

“Mass shift-based” identification of ketone/aldehyde species in UNT and of deuterium-containing formulae in BDR (F2): For each peak (at mass m) in UNT to which a molecular formula was assigned, the peak list of a borodeuteride-reduced sample was searched for peaks at $m+3.021927$ and $m+6.043854$ within a 0.2 ppm window, corresponding to singly and doubly-reduced species, respectively, where mass m was calculated from the molecular formula at that peak. As done previously,⁹¹ if only the $m+3.021927$ or $m+6.043854$ peak was found in the borodeuteride-reduced spectrum, then the species at mass m in UNT was identified as comprising at least one singly-reducible or at least one doubly-reducible species, respectively. Alternatively, if *both* the $m+3.021927$ and $m+6.043854$ peaks were found, then the peak at mass m was identified as comprising either a combination of species with one *and* two reducible moieties, or an

individual (or set of isobaric) species with two reducible moieties that was only partially reduced. Peaks in the searched (reduced) peak list that were identified as $m+3.021927$ and $m+6.043854$ (reduced) species were then assigned formulae equivalent to that of the peak at m with the addition of HD and H₂D₂, respectively. Molecular formulae of any remaining, unassigned peaks in the searched peak list were calculated using function **F1** as described above. As a control, the peak list of UNT, rather than that of a reduced sample, was searched in a similar manner for $m+3.021927$ and $m+6.043854$ mass markers.

Molecular formula assignment allowing up to 2 deuterium atoms (“F1*”): As a further check on the reliability of the “mass shift-based” assignment of deuterium (D) containing formulae in the peak lists of BDR, a modified version of function **F1** was used to calculate molecular formulae for all samples. This modification allowed D to be included in the possible combinations of atoms by allowing up to two HD subunits to be included in the base hydrocarbon molecular formulae. D-containing formulae calculated in this manner were then compared to those that were calculated based on mass shifts from **F2**.

Optical measurements. For all optical measurements, a portion of all aliquots of SRFA (initially at pH 2) was diluted with phosphate buffer to final concentrations of 25 mg/L SRFA and 32 mM phosphate. This preparation served two purposes: Firstly, to ensure that the final pH of all samples were constant (6.7 ± 0.2) to minimize pH-dependent optical changes, and secondly, to keep the absorbance of each sample below 0.1 for all excitation wavelengths used in fluorescence measurements to minimize inner filter effects. Prior to analysis, all samples were filtered through 0.2 μm polyethersulfone

(VWR International, part no. 28145-501) sterile membrane filters to minimize interference (e.g. scatter) from particulate matter.

All optical measurements were done in a 1 cm quartz cuvette. Absorption spectra were measured against air from 200 – 700 nm using a Shimadzu 2401-UV spectrophotometer. Emission spectra were recorded using an Aminco-Bowman AB2 luminescence spectrophotometer. Using excitation wavelengths spaced 5 nm apart from 250 to 450 nm, forty-one emission spectra were recorded every 2 nm from 250 to 650 nm at a rate of 2 data points per second. Excitation and emission monochromator band passes were set a 4 nm, and spectra were corrected for the instrument response using factors supplied by the manufacturer. Absorption and emission (excitation wavelength = 350 nm) spectra were acquired for solutions of 10 ppm and 100 ppb quinine sulfate in 0.1 M H₂SO₄, respectively.

The absorption and emission spectra of a blank (32 mM phosphate buffer in Milli-Q water) were recorded and subtracted from all sample spectra using a custom MATLAB function (Appendix 9, section A9.8). For each absorption spectra, the average absorbance from 650 to 700 nm was subtracted from all absorbance values for the entire wavelength range to correct for slight baseline offsets.³ The absorption spectral slope (S_A) was calculated from 300 to 650 nm using a non-linear least squares fitting of the following equation,

$$(4.1) \quad A(\lambda) = A(\lambda_{ref})e^{-S_A (\lambda - \lambda_{ref})}$$

where $A(\lambda)$ is the (baseline-corrected and blank-subtracted) absorbance at wavelength λ and λ_{ref} is the reference wavelength (300 nm).

Each individual emission spectrum was blank subtracted, smoothed using a 3-point moving average, and corrected for primary and secondary inner filter effects as described by McKnight et al.²⁶ The emission intensities at wavelengths within 10 nm of the Rayleigh and Raman scatter peaks were excised and interpolated using a procedure similar to that of Zepp et al.²⁷ All corrected intensities were converted to quinine sulfate units (QSU), where 1 QSU is equal to the emission intensity of a 1 ppb quinine sulfate solution at excitation and emission wavelengths 350 and 450 nm, respectively. Fluorescence quantum yields were calculated for each excitation wavelength as described previously.²⁸

4.3. Results and Discussion

General mass spectrometric features. The mass spectra of all samples contained about 8000 – 9000 resolved peaks, most of which were [M-H]⁻ ions (Table 4.2; see Appendix 2, Figure A2.2 for full mass spectra of all samples in this study). For all subsequent analysis, only peaks within the 200 to 600 *m/z* range were analyzed. This range comprised the majority of peaks (79% to 94%) and total ion count (TIC; 94% to 98%) in the full scan range. Many spectra contained a few peaks with very high intensities relative to the ions of interest within SRFA and were also present in the mass spectrum of the blank. The peaks with the highest intensities were tentatively identified as series of organosulfonate contaminants. Also present in substantial intensities in the blank spectrum were two clusters of unidentified species that were most likely chloride adducts, as well as two other unknown organic species (Appendix 1, Figure A1.1 and

Table A1.3). These suspected common contaminant peaks, as well as any peaks comprising doubly or triply charged ions, were removed from the sample peak lists.

The number and weight average molecular weights (AMW_N and AMW_W) for all samples were around 400 m/z and were similar to values reported in previous studies using negative ion ESI mass spectrometry.⁴⁸ These average molecular weights decreased slightly (average percent change \pm standard deviation relative to UNT: $-6.5 \pm 0.7\%$ and $-7.8 \pm 1.0\%$ for AMW_N and AMW_W , respectively) following reduction. This may tentatively suggest that there are more species that were reduced at high m/z than at low m/z . Although counterintuitive because borodeuteride reduction of a single species would result in a mass *increase*, it is likely that reduction of many species produces a net loss of ionization efficiency. Therefore, if more species at high m/z are reduced, then there would be a loss of intensity at the high m/z range, and a subsequent decrease in average molecular weight would be observed.

In the 200 to 600 m/z range, the TIC of the mass spectra of the reduced samples were significantly lower than that of the mass spectrum of the untreated sample ($-20.7 \pm 4.7\%$), which is consistent with what was observed in the previous chapter (i.e., a loss of TIC following reduction). In contrast, the numbers of peaks in the reduced samples were *higher* ($21.5 \pm 5.3\%$) relative to that of the untreated, which is the opposite of what was observed in Chapter 3. However, the previously observed decrease in the number of peaks following reduction was most likely an artifact of relatively low instrument sensitivity, and the slight *increase* in number of peaks is more reasonable, since borodeuteride reduction would generate many peaks which were not in the untreated sample itself.

Table 4.2. Distribution of peaks in the full (50 – 2000) and analyzed (200 – 600) m/z ranges.

| sample | ME ^a | peaks | | TIC ^b (x10 ¹¹) | | AMW _N ^c | AMW _W ^d |
|--------|-----------------|-------|----------|---------------------------------------|----------|-------------------------------|-------------------------------|
| | | full | analyzed | full | analyzed | | |
| UNT | 0 | 8507 | 6753 | 2.88 | 2.70 | 408 | 436 |
| BDR-1 | 2.3 | 9030 | 8021 | 2.34 | 2.26 | 386 | 409 |
| BDR-2 | 6.0 | 8481 | 7716 | 2.06 | 2.01 | 380 | 402 |
| BDR-3 | 20 | 9095 | 8432 | 2.24 | 2.19 | 383 | 403 |
| BDR-4 | 58 | 9269 | 8632 | 2.28 | 2.24 | 379 | 399 |
| BDR-5 | 155 | 8734 | 8220 | 2.04 | 2.00 | 379 | 398 |

^a Mass equivalents of NaBD₄ (i.e. ratio of [NaBD₄] / [SRFA]) prior to PPL extraction; ^b total ion count; ^c number and ^d weight averaged molecular weight calculated using the full m/z range (see Appendix 9 and section A9.9 for calculations).

Molecular formula assignment of UNT and BDR-3. Prior to molecular formula assignment, the maximum number(s) of non-oxygen heteroatoms (e.g. N, S) must be chosen. While allowing large numbers of heteroatoms can be beneficial for assigning formulae to as many peaks as possible in a mass spectrum, doing so may lead to a greater chance for false and/or *equivocal* formula assignments, especially at high m/z . Here, molecular formulae were calculated from the peak list of BDR-3 using two algorithms employing various maximum allowed numbers of N and S. In algorithm **F1***, formulae were calculated directly using a modified “low-mass moiety” approach which allowed the addition of up to two HD moieties. In algorithm **F2**, the mass list of BDR-3 was searched for peaks at $m+3.021927$ and $m+6.043854$, where m is the mass of a peak in UNT to which a formula was previously assigned. These peaks were then assigned

formulae identical to that of the original peak at m with the addition of HD or (HD)₂. The D-containing formulae calculated by algorithms **F1*** and **F2** within BDR-3 were then compared to one another at four 100 m/z -wide ranges. To test the reliability of the D-containing assignments, the two algorithms were used on the peak list of the UNT sample itself as a control.

For all N and S allowances, virtually all D-containing formula identified by **F2** were also identified by **F1*** (Figure 4.2A). However, the converse was not necessarily true, especially in the 500 to 600 m/z range. That is, there were some peaks that were assigned D-containing formulae by **F1***, yet were not identified as $m+3.021927$ or $m+6.043854$ derivatives (yellow bars). The presence of formulae identified by **F1*** but not **F2** and the relatively small number of formulae identified by **F2** at the highest m/z range might suggest that the mass shift-based search fails to identify some subset of D-containing species. However, the numbers of D-containing formulae falsely identified by **F1*** in UNT were very similar to those identified in BDR (Figure 4.2B), suggesting that those peaks identified by in BDR-3 by only **F1*** are incorrect formulae.

A second possible reason for the lower number of D-containing species identified by **F2** than by **F1*** is that a 0.2 ppm formula error window is inappropriately small at the high m/z range, since formula error generally increases with m/z (see Appendix 5). To test this, all analyses were carried out a second time using an increased error tolerance of 0.3 ppm. This resulted in substantial increases in the number of D-containing formula identified at high m/z ranges by both functions, with smaller changes occurring in the low m/z ranges (Figure 4.2C). Also, in all cases, *only* **F1*** and *not* **F2** returned significant numbers of false assignments in UNT (Figure 4.2D), suggesting that by including this

additional nucleus (D) in **F1*** more false assignments were obtained due to the increased degrees of freedom in all possible elemental combinations, particular at the higher m/z range. These results demonstrate that a mass shift-based identification of D-containing species (as in **F2**) is far more reliable than a direct formula assignment, even at high m/z and N and S allowances. However, it is important to note that **F1*** served as an independent verification of all D-containing formulae assigned by **F2**. Based on the above results, a 0.2 ppm error tolerance and up to one N and one S was allowed for the molecular formulae assignment of peaks in UNT. Using these parameters, 54.4% (3674 of 6753) of the peaks in the 200 to 600 m/z were assigned molecular formulae (see Appendix 5 for details). These assigned peaks accounted for 86.2% of the total spectral peak magnitude, thus showing that the unassigned peaks were of very low relative intensities. For the mass shift-based identification of ketones/aldehydes (via **F2**), a 0.3 ppm error tolerance was used.

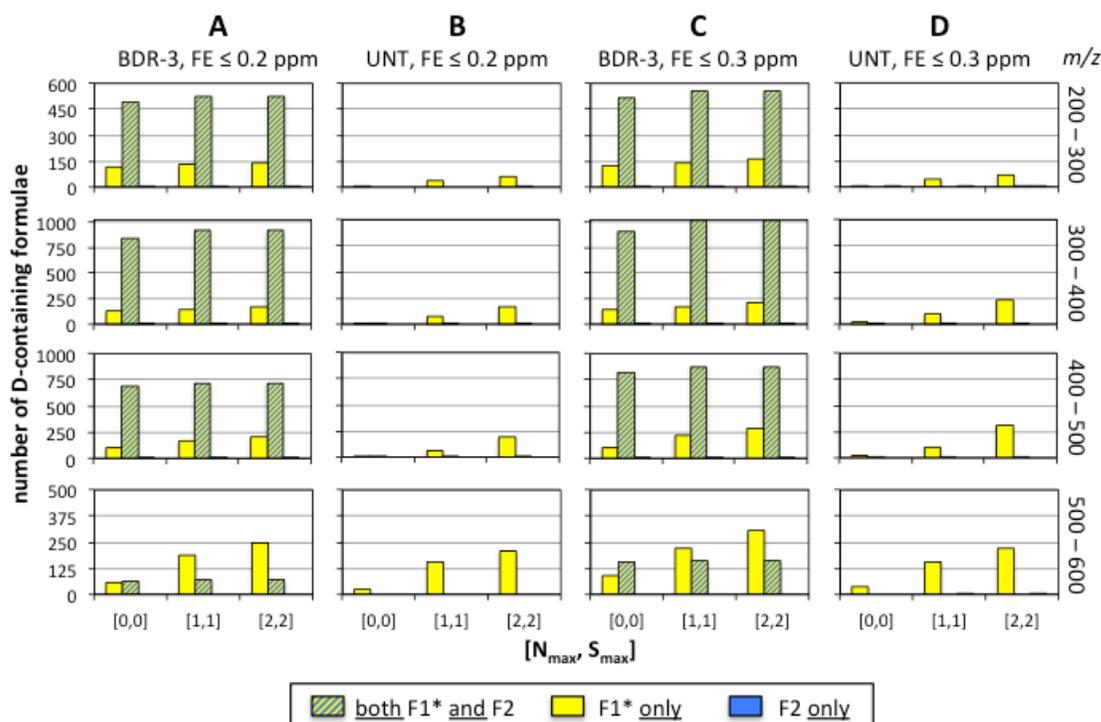


Figure 4.2. Number of D-containing molecular formulae identified in BDR-3 (A and C) and UNT (B and D) in four m/z ranges by algorithms F1* and/or F2 with various maximum numbers of N and S using maximum formula error (FE) tolerances of 0.2 (A and B) and 0.3 (C and D) ppm.

Identification and reactivity of ketone/aldehyde-containing species in UNT.

Species that were reduced by varying amounts of borodeuteride were identified by searching the mass lists of BDR-1 through BDR-5 for peaks at $m+3.021927$ and $m+6.043854$ (within 0.3 ppm), where m is the m/z of a peak in UNT to which a molecular formula was assigned. Additionally, a search in the peak list of UNT for these mass markers was performed as a control search. There were very few false assignments in this control search, which is consistent with the above results for the mass shift-based formula assignments. From a search in BDR-1 (i.e., the sample reduced with the lowest $[\text{NaBD}_4]$),

45% (1647 out of 3674) of all assigned peaks in UNT were identified as comprising reduced species, 64% of which were identified as comprising species that were reduced only a single time (Figure 4.3A). In general, using higher $[\text{NaBD}_4]$ had minor effects on the *total* number of species reduced, but more drastic effects on the number of species that were reduced a second time. For instance, the *total* number of peaks reduced after treatment with 20 mass equivalents of NaBD_4 (BDR-3) was only 2.4% higher than that which was reduced after treatment with 2.3 mass equivalents (BDR-1; 1686 versus 1647), whereas the number of *doubly-reduced* species increased by 65% (994 versus 601).

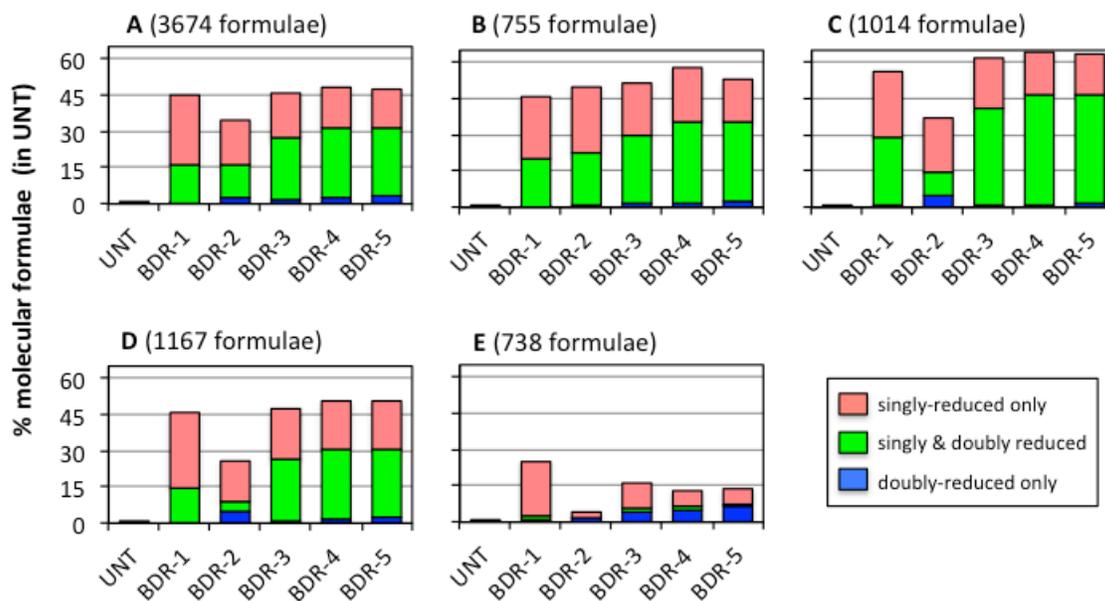


Figure 4.3. Percentage of peaks in UNT with an assigned molecular formula identified as comprising singly- and/or doubly- reducible species for m/z ranges (A) 200 – 600, (B) 200 – 300, (C) 300 – 400, (D) 400 – 500, and (E) 500 – 600.

The percentages of peaks identified as comprising ketones/aldehydes within the first three m/z ranges were similar to one another (Figure 4.3B – E), with the 300 to 400 m/z range containing the highest percentages of reduced species. In these ranges, there were considerably more peaks identified as comprising singly- *and* doubly-reduced species than there were identified as comprising *only* doubly-reduced species. In all m/z ranges, the percentage of peaks that were identified as reduced from a search in BDR-2 was much lower than expected. The cause of this deviation is unknown and is perhaps due to contamination, electrospray instability, and/or issues with the post-calibration of the peak list of BDR-2, which would lead to an artificially low number of $m+3.021927$ and $m+6.043854$ mass markers. The highest m/z range was relatively devoid of reduced species, which may be due to suppression of low intensity peaks at this range for all BDR samples. The unexpected decrease in the number of identified ketones/aldehydes with increasing $[\text{NaBD}_4]$ confirms this explanation, since signal suppression would increase with successively higher $[\text{NaBD}_4]$.

Based on the observed positive monotonic relationship between the doubly-reduced species and $[\text{NaBD}_4]$, it seems that the majority of species that contain *at least* one ketone/aldehyde moiety are reduced once under low $[\text{NaBD}_4]$, while higher $[\text{NaBD}_4]$ results mostly in the subsequent reduction of other reducible moieties within the same species. This phenomenon was readily confirmed in Table 4.3, which shows the number of species identified as non, singly, and/or doubly-reduced from a search in BDR-1 (treatment with low $[\text{NaBD}_4]$) and BDR-3 (treatment with high $[\text{NaBD}_4]$). Most peaks (87%; 1768 of 2027) that were identified as non-reduced following treatment with low $[\text{NaBD}_4]$ were also identified as such following treatment with high $[\text{NaBD}_4]$.

Conversely, of the 1046 peaks identified as comprising singly-reduced species upon treatment with low [NaBD₄], about 35% were reduced a second time upon treatment with high [NaBD₄], while about 45% remained singly-reduced, and about 20% went from singly-reduced to non-reduced. This phenomenon is useful, as it suggests that treatment of SRFA (and perhaps other DOM samples) with relatively low [NaBD₄] can be used to confidently identify most mass spectral peaks which comprise species with *at least* one ketone/aldehyde moiety, whereas multiple reductions would be required to differentiate between peaks comprising singly- and/or doubly-reduced species.

Table 4.3. Numbers of non, singly, and/or doubly reduced species in UNT (between 200 and 600 *m/z*) following a 2.3-fold (BDR-1) and 20-fold (BDR-3) addition of NaBD₄. Regions above and below the (shaded) diagonal correspond to species which decreased or increased in reduction, respectively, from BDR-1 to BDR-3.^a

| peak composition vs. BDR-3 | peak composition vs. BDR-1 | | | |
|----------------------------|----------------------------|---------------|------------|-------------|
| | non (2027) | singly (1046) | both (587) | doubly (14) |
| non (1988) | 1768 | 210 | 6 | 4 |
| singly (693) | 202 | 474 | 16 | 1 |
| both (929) | 25 | 338 | 563 | 3 |
| doubly (64) | 32 | 24 | 2 | 6 |

^a In parenthesis: Total number of species identified as non-reduced (“non”), singly-reduced only (“singly”), singly- and doubly-reduced (“both”), and doubly-reduced only (“doubly”) for BDR-1 (topmost row) and BDR-3 (leftmost column).

Molecular formulae of ketone/aldehyde-containing species. To further understand the molecular composition of the reduced species within the UNT sample, plots of H/C versus O/C molar ratios (Van Krevelen plots) were constructed for each of the four m/z ranges, shown in Figure 4.4 (see Table 4.4 for numbers of formulae). These plots distinguish between species identified as non-, singly-, and/or doubly-reduced upon treatment with low (top) and high (bottom) $[\text{NaBD}_4]$. The H/C and O/C molar ratios of the majority of all formulae are centered around H/C = 1 and O/C = 0.5, which has been observed previously for SRFA and other terrestrially influenced DOM samples, and are consistent with an abundance of lignin-derived material and/or carboxyl-rich alicyclic molecules (CRAM). This distribution is spread out at the low m/z range, and becomes tightly distributed and shifted to slightly lower H/C and higher O/C molar ratios at the higher m/z range. In addition to the major distribution of formulae at mid H/C and O/C, there was a small distribution of formula at low O/C (<0.25) and low H/C (< 1) at the high m/z range, which has previously been identified as carboxylated aromatic species derived from biochar.^{105,106} These species were devoid of ketone/aldehyde functional groups, even when high $[\text{NaBD}_4]$ is used, thus consistent with the common assumption that carboxyl moieties are the only dominant functional groups within these species.

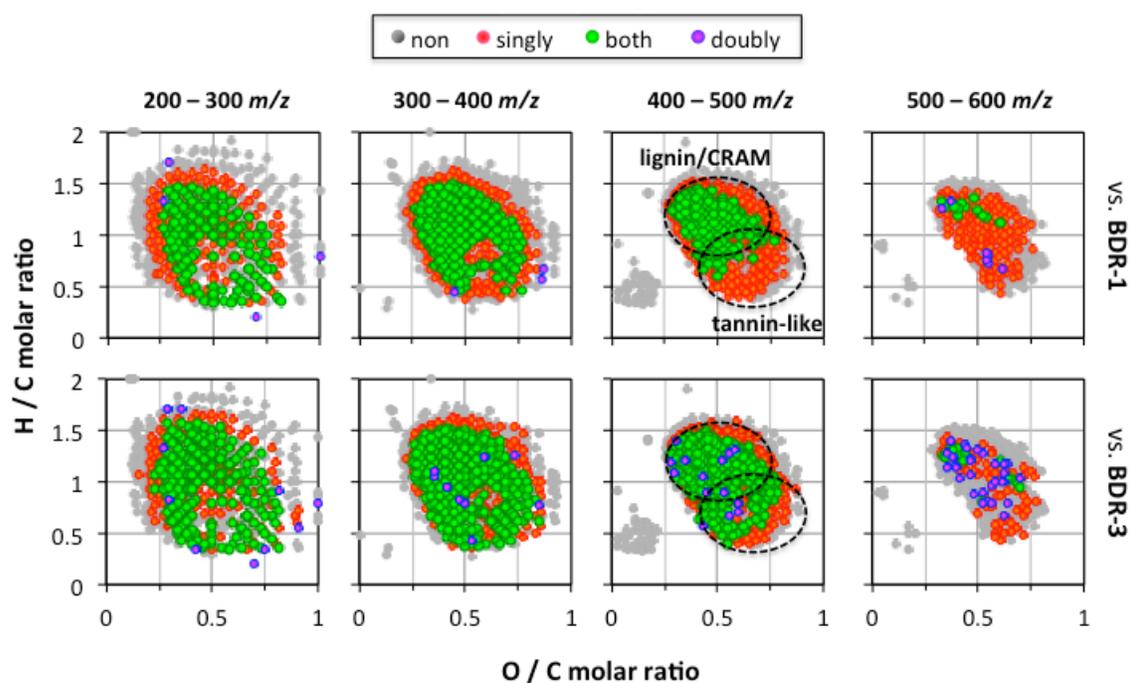


Figure 4.4. Van Krevelen plots of formulae of peaks in UNT identified as non-, singly-, and/or doubly-reduced based on a search for $m+3.021927$ and $m+6.043854$ mass markers in (top) BDR-1 and (bottom) BDR-3 in the four m/z ranges analyzed (see Table 4.4 for numbers of formulae).

Table 4.4 Number of formulae in UNT at each m/z range identified as non-, singly-, and/or doubly-reduced.

| ME ^a | peak composition | 200 – 300 m/z | 300 – 400 m/z | 400 – 500 m/z | 500 – 600 m/z | total |
|-----------------|------------------|-----------------|-----------------|-----------------|-----------------|-------|
| 2.3 (BDR-1) | non | 408 | 443 | 634 | 553 | 2038 |
| | singly | 197 | 275 | 363 | 0 | 835 |
| | both | 146 | 289 | 170 | 0 | 605 |
| | doubly | 4 | 7 | 0 | 1 | 12 |
| 20 (BDR-3) | non | 368 | 390 | 612 | 618 | 1988 |
| | singly | 162 | 212 | 244 | 75 | 693 |
| | both | 214 | 403 | 297 | 15 | 929 |
| | doubly | 11 | 9 | 14 | 30 | 64 |

At high m/z , the main distribution of formulae was slightly shifted to lower H/C and higher O/C, which may be either indicative of either oxidized lignin-derived species or condensed and/or hydrolysable tannins composed of flavone, polyphenolic, and/or sugar moieties.^{107,108} In contrast to the more saturated (i.e., higher H/C) and less oxygenated (lower O/C) lignin/CRAM-like compounds, the peaks with tannin-like formulae were reduced only a single time with low [NaBD₄] and required much higher [NaBD₄] to become reduced a second time. This suggests a greater structural heterogeneity for the tannin-like compounds. That is, while most ketones/aldehyde moieties within lignin/CRAM-like compounds reacted equally with both high and low [NaBD₄], there were a few moieties in the tannin-like compounds that did not react except under very high [NaBD₄].

Effects of borodeuteride reduction on absorption and fluorescence emission of SRFA. To study the effect of borodeuteride reduction on the optical properties of SRFA, absorption and emission spectra were taken for each aliquot prior to SPE. These spectra are thus representative of the optical properties of the entire sample, whereas all mass spectra were collected from post-SPE material. Therefore, direct comparisons between the compositional information obtained by mass spectrometry and the optical properties must be done with caution, since the species detected by ESI may exclude some subset of species within the original ensemble. However, this is true for all ESI analyses of DOM, as ESI is understood to only ionize a fraction of the total ensemble of a sample of DOM. Furthermore, preliminary experiments have demonstrated that SPE has relatively small effects on the bulk optical properties (Appendix 6).

Borodeuteride reduction resulted in a loss of absorption at wavelengths higher than about 250 nm, with a preferential fractional loss of absorption at the longer wavelengths (Figure 4.5). This behavior has been observed in many other humic substances following borohydride reduction, and has been explained by the disruption of charge-transfer complexes formed between electron donors (e.g. phenols) and reducible acceptor moieties (e.g. aromatic ketones).³⁷ From 300 to 450 nm, the observed fractional loss of absorbance increased logarithmically (see Appendix 6, Figure A6.2) with the amount of borodeuteride added, which is again in agreement with what has been observed for several terrestrial and aquatic humic and fulvic acids.¹⁰⁹ For example, while reduction with 2.3 mass equivalents of NaBD₄ (i.e., the lowest [NaBD₄] used in this study) resulted in a ~30% loss of absorption at 400 nm, 155 mass equivalents of NaBD₄ were required to double this loss of absorption!

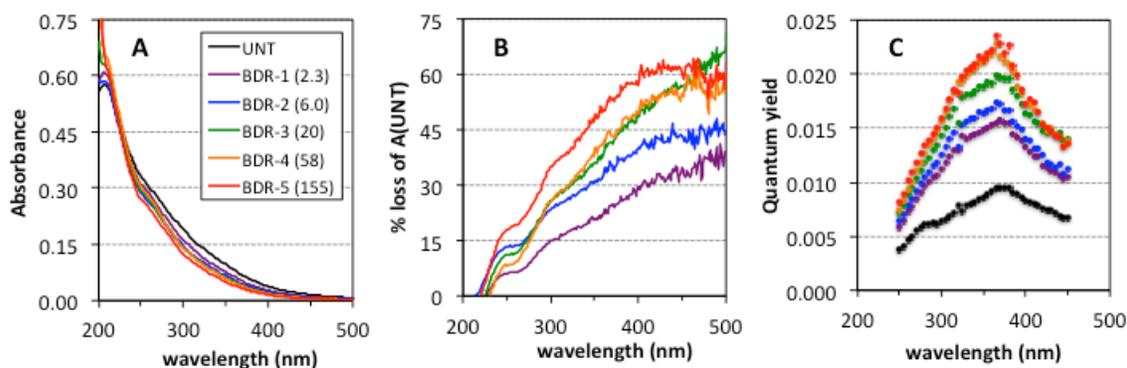


Figure 4.5. Absorption and emission of all SRFA samples. **A:** Blank-subtracted absorption spectra, **B:** Percent loss of absorption (relative to UNT) following reduction, and **C:** quantum yield.

Although the loss of long-wavelength absorption increased with successively higher amounts of NaBD₄ for all tested mass equivalents may suggest that complete reduction of all reducible moieties within a DOM sample would be achieved using 100 or more mass equivalents of NaBD₄, doing so may not be advisable for a number of practical reasons. Firstly, using such high concentrations of NaBD₄ (i.e. ≥ 10 mg/mL NaBD₄ for a 0.1 mg/mL DOM sample) would necessitate the use of equally large amounts of acid such as HCl for the neutralization of samples after reduction. This can be especially problematic, since high concentrations of HCl would increase the likelihood of residual chloride remaining in the solid phase extracted sample, thus interfering with the ESI process. Secondly, introducing such large amounts of NaBD₄ produces a vigorous evolution of hydrogen gas from the reaction of borodeuteride with water. Furthermore, previous work has suggested that amounts of borohydride in excess of approximately 5-fold mass equivalents of HS contribute to the reduction of quinones, which re-oxidize once the excess borohydride (or borodeuteride) is consumed.¹⁰⁹ Finally, the additional changes in absorption and emission from a 20-fold mass equivalent to a 155-fold mass equivalent were relatively small. Based on these considerations, a 20-fold mass equivalent of NaBD₄ to DOM should be sufficient to convert nearly all (irreversibly reducible) ketones and aldehydes in other DOM extracts.

Shown in Figure 4.6 are fluorescence emission spectra collected at excitation wavelengths (EX) spaced over 15 nm from 260 to 440 nm for all six samples. In all samples, the emission spectral envelope from successively higher EX fell within the envelope from lower EX, with steadily increasing (i.e. red shifted) emission maxima – a feature that has been interpreted as arising from a continuum of coupled charge-transfer

states.²² Upon reduction, the wavelength of maximum emission (EM_{MAX}) for all EX shifted to shorter wavelengths, with the greatest shifts occurring primarily for lower EX (see Appendix 6, Figure A6.1). Most of these shifts occurred from the untreated to 20-fold reduced sample, and very little additional shifts were observed for successively higher amounts of borodeuteride. Furthermore, emission intensities generally increased with increasing mass equivalents of $NaBD_4$, with the exception of BDR-5 (i.e., 155-fold reduction).

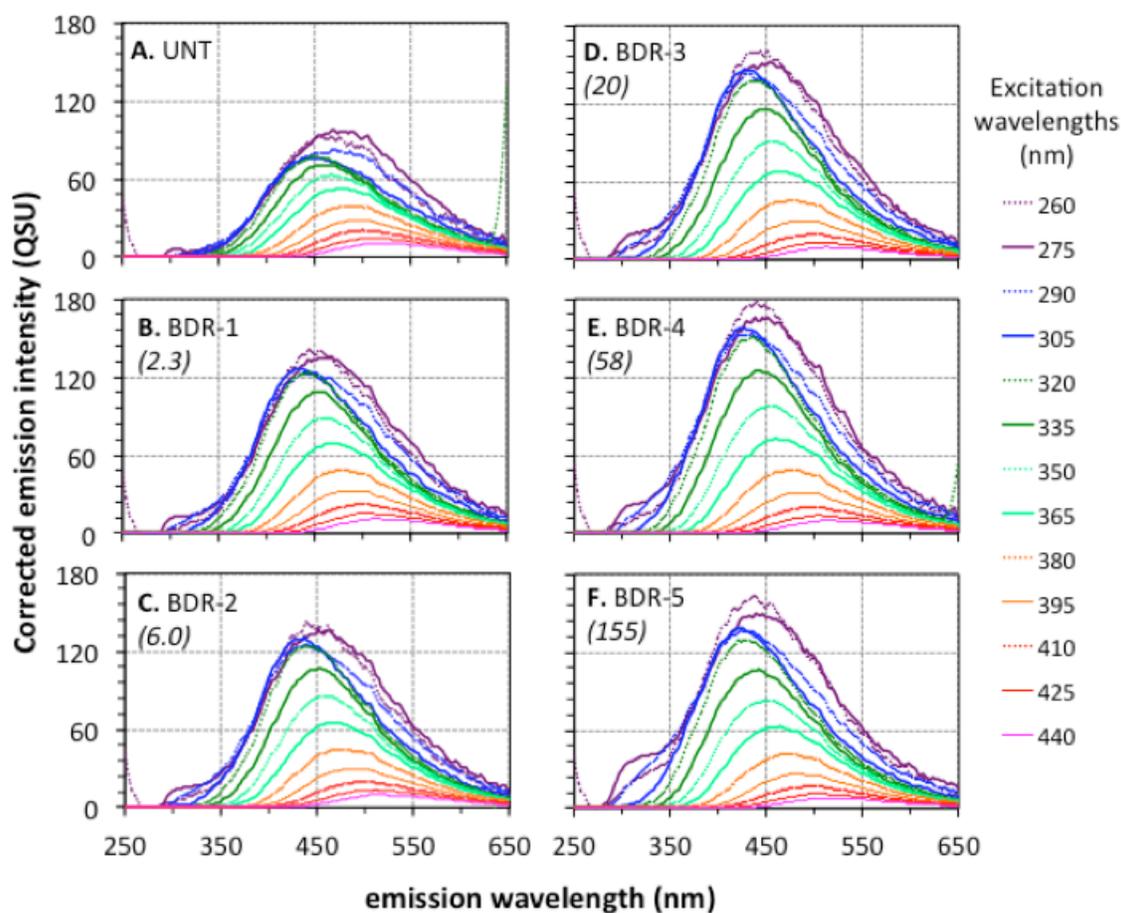


Figure 4.6. Corrected emission spectra of untreated and reduced SRFA (25 mg/L), in quinine units (QSU), where 1 QSU = emission intensity of a 1 ppb quinine sulfate solution at (EX, EM) = (350, 450). Mass equivalents of $NaBD_4$ are given in parentheses (see Appendix 6 for difference spectra).

An additional spectral feature that appeared following reduction was a shoulder from emission wavelengths (EM) 300 – 350 nm for $EX \leq 275$ nm. This spectral region is typically assumed to be indicative of proteinaceous material arising from the emission of tryptophan and tyrosine amino acid residues.^{29,110} However, there was no significant evidence of protein or peptide-like material in the ESI mass spectral data, which would produce N-containing species at high H/C molar ratios. Therefore, it is more likely that this spectral feature is due to the emission of other phenolic species (e.g. lignins and tannins) whose emission is originally quenched in the untreated sample via charge transfer from an excited-state phenolic moiety to a reducible acceptor such as an aromatic ketone. This result highlights how electronic interactions between reducible chemical moieties and other species within DOM can greatly influence the bulk excitation-emission spectra of such materials, and that assignment of emission spectral features to a single compound class must be done with great caution.

Relating the bulk absorption and emission of SRFA to number of identified reduced species. The numbers of peaks identified as comprising ketone/aldehyde-containing species were compared to several measured optical properties of each BDR sample (Figure 4.7). Since the number of ketone/aldehyde-containing species identified in sample BDR-2 was much lower than expected, data from this sample were excluded. The correlation between the number of ketone/aldehyde-containing species and the percent loss in absorption of BDR to UNT generally increased with increasing wavelength, thus supporting the premise that ketone/aldehyde-containing species influence primarily the long wavelength absorption of SRFA. The number of singly-reduced species did not correlate strongly to the observed optical properties, but the

number of doubly-reduced species as well as the total number of reductions did. The correlations between the number of reduced species and the observed emission intensities were very weak when BDR-5 was included. However, upon exclusion of the data for this sample, very strong linear correlations ($R^2 \geq 0.98$) were observed for the percent change in emission versus the number of doubly-reduced species and total number of reductions. The lack of a correlation between all observed optical properties and the number of singly-reduced species implies that species with two or more ketone/aldehyde influence the bulk optical properties more than species with only one such moiety.

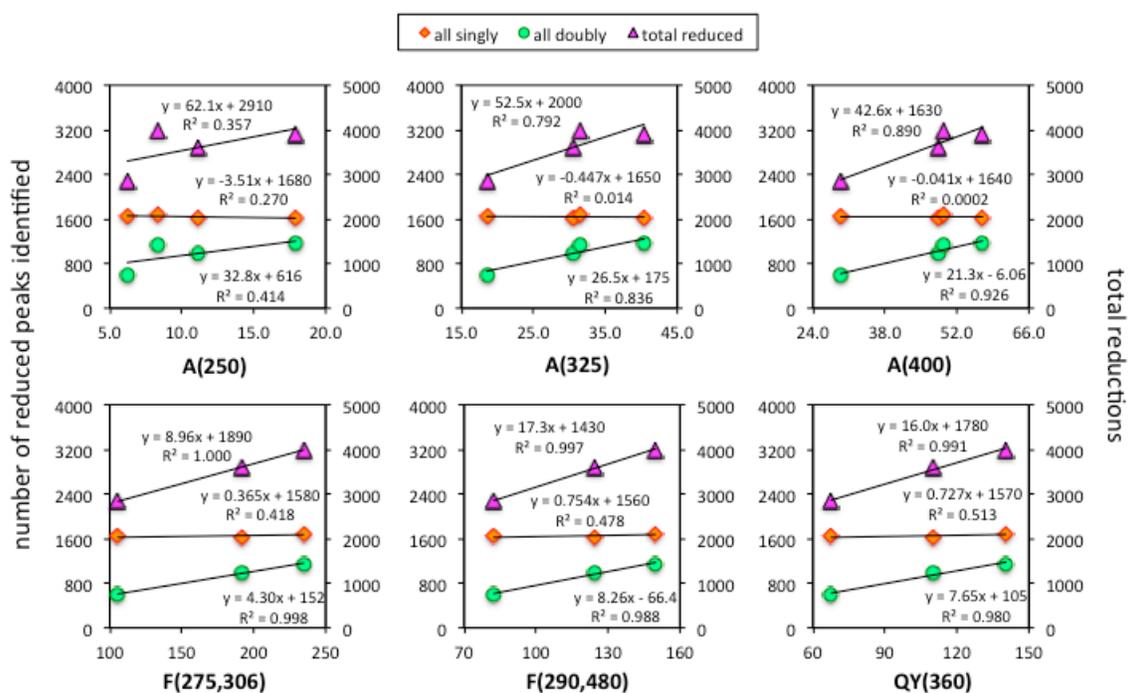


Figure 4.7. Number of reduced species identified via ESI FT-ICR MS versus percent change in $A(\lambda)$, $F(\text{EX,EM})$, and $QY(360)$ relative to UNT. “All singly reduced” = singly reduced only + singly & doubly reduced; “all doubly reduced” = doubly reduced only + singly & doubly reduced; and “total reductions” = singly reduced only + 2(doubly reduced only) + 3(singly & doubly reduced). Total reductions (purple triangles) are given on the right-hand vertical axes. BDR-5 data are excluded for percent change in $F(275,306)$, $F(290,480)$, and $QY(360)$. Percent changes are absolute values.

4.4. Conclusions

The work in this chapter addressed and expanded upon several major aspects of the reduction and mass labeling of DOM with NaBD₄ that was developed previously. Firstly, it established that most of the major optical and mass spectrometric changes occur with a 20-fold mass equivalent of NaBD₄ to SRFA. Although some additional structural and optical changes occurred at higher concentrations, 20 to 30-fold mass equivalents should be a practical standard amount of NaBD₄ used to reduce nearly all ketones and/or aldehydes within an aqueous humic/fulvic acids as well as other CDOM samples. For analyzing the most reactive reducible species, considerably smaller amounts of NaBD₄ (< 1 mass equivalent) should be used. Secondly, since this analysis requires direct comparisons between two mass spectra, it is recommended that duplicate or triplicate mass spectra of untreated and reduced (low and high [NaBD₄]) be acquired for routine characterization of DOM in this manner. Finally, several measured bulk optical properties were semi-quantitatively linked to the number of species which were identified as reduced following treatment by various [NaBD₄].

Chapter 5: Comparison of the presence and composition of ketone/aldehyde-containing species in DOM from various aquatic environments

5.1. Introduction

In this chapter, the fully developed and tested method of using sodium borodeuteride (NaBD_4) to mass label ketone/aldehyde-containing species is employed to compare reducible species in dissolved organic matter (DOM) extracted from a variety of aquatic environments. Since ketone/aldehyde-containing species within DOM have been demonstrated to heavily influence the optical properties and photochemistry of these materials, the characterization of this subset of compounds would be of great use to better understand the relationship between the structure and composition of DOM and many bulk physicochemical properties such as the absorption and emission of ultraviolet and visible light by these materials.^{3,19–21,37,38,111}

Throughout the literature, there have been many studies which employ ultrahigh resolution electrospray ionization (ESI) Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry to achieve similar objectives.^{9,47,79,82,83,112,113} That is, to extensively characterize and compare the composition of DOM collected from a suite of geographic locations based on the molecular formulae that are assigned to the many species observed by ESI FT-ICR mass spectrometry.

Many of the same analysis techniques that were used in those previous studies to compare the compositional information are utilized here. In addition to these analyses, the newly developed mass labeling method is demonstrated to be a relatively easy method to employ in the routine mass spectrometric characterization of DOM, which can be used to gain additional molecular level details that would not otherwise be observed. As done in many other studies involving the characterization of DOM, Suwannee River fulvic acid (SRFA) is used as a standard with which to compare all other samples and to validate the analytical methodology being used. However, while SRFA is heavily influenced by the vast amount of decaying vegetation that is found within the environment of the blackwater Suwannee River, many of the samples in this study were extracted from remote marine samples whose compositions may be substantially different from that of SRFA.

Recently, Pony Lake fulvic acid (PLFA), another DOM reference standard, has been characterized using ESI FT-ICR mass spectrometry.³⁰ While SRFA comprises primarily terrestrially-derived material, PLFA is believed to comprise primarily (*in situ*) microbially-produced material, owing to its origin in a eutrophic, saline coastal pond in Antarctica which is devoid of high order plants.¹¹⁴ Since the origin of DOM in marine environments is still relatively poorly understood, using SRFA and PLFA as terrestrial and microbial end-members, respectively, may be useful for drawing conclusions regarding the origins of such samples. However, PLFA may not be ideal for detailed comparisons, since the microbes that are thought to be responsible for the production of PLFA (primarily *Chlamydomonas intermedia*¹¹⁵) are drastically different than those

which are found in oceanic environments. However, the general types of compounds that are produced by microbial sources may share similar compositional traits.

The overall goal of the work described in this chapter is two-fold. First, the ketone/aldehyde identification method developed in the two previous chapters is applied to the ultrahigh resolution ESI mass spectra of DOM samples collected from diverse aquatic environments in an attempt to better understand the composition and location of ketone/aldehyde-containing species within DOM. Secondly, the usefulness of SRFA and PLFA as riverine- and *in situ*-produced end-members is tested.

5.2. Materials and Methods

Reagents, and materials. Sodium borodeuteride (NaBD_4 ; 98% D, 90% purity), hydrochloric acid (HCl, trace select) and sodium hydroxide (NaOH) was purchased from Sigma Aldrich. Water was obtained from a Milli-Q Academic water purification system (Millipore). LC grade methanol (MeOH) was purchased from Fisher, and potassium phosphate monobasic and dibasic (KH_2PO_4 and K_2HPO_4 , respectively) was purchased from Baker. Bond Elut PPL solid phase extraction (SPE) cartridges (100 mg functionalized polystyrene-divinylbenzene solid phase; part #12105004) were purchased from Agilent.

DOM sample collection and predicted compositional relationships. Suwannee River fulvic acid (SRFA; catalog number: 2S101H) and Pony Lake fulvic acid (PLFA; catalog number: 1R109F) were purchased from the International Humic Substances Society. C_{18} extracts from the Lower Delaware Bay (DELB), Delaware River (DERV), Congo River Plume (EACR), equatorial Atlantic Ocean (EAUW) and northern Atlantic

Ocean at 0 m (EASO) and 1000 m depths (EADO) were collected on two separate cruises (see Appendix 7 for locations and extraction protocol). For this study, the relative contributions of terrestrial and *in situ* sources to all samples were approximately predicted based on the proximity of each sample collection location to a possible terrestrial source (e.g. riverine outputs), and SRFA and PLFA were assumed to have the largest relative contributions by terrestrial and *in situ* sources, respectively. Thus, the order in which samples were considered most terrestrially-produced to most *in situ*-produced were predicted to be as follows: SRFA, DERV, DELB, EACR, EAUW, EASO, EADO, PLFA.

Sample preparation. Working stocks of DERV and DELB (Delaware Bay extracts) were prepared by thawing the concentrated extracts (originally at pH 7), diluting them with Milli-Q water and adjusting to pH 10 with an aqueous solution of NaOH to roughly match the absorbance at 350 nm to that of a 100 mg/L SRFA solution at pH 10. Additionally, a 0.10 mg/mL stock of PLFA was prepared by dissolving solid sample in Milli-Q water and adjusting the pH to 10 with NaOH while stirring. The stocks were then divided into two (15.0 mL) aliquots in 20 mL glass vials. Approximately 75 mg of solid NaBD₄ was added to one aliquot of each sample. All vials were loosely capped and placed in the dark at room temperature for 24 h. 6 M HCl was added to a final pH of 2.0 ± 0.1 (to decompose any residual NaBD₄) and all aliquots were stored in the dark for an additional 3 h. Untreated (UNT) and borodeuteride-reduced (BDR) aliquots of SRFA were prepared as described in Chapter 4.

A similar process for EACR, EAUW, EADO, and EASO (Equatorial Atlantic extracts) was carried out as follows. Working stocks of all four samples were prepared by thawing concentrated C₁₈ extracts (originally at pH 7) diluting them by a factor of about 1:10 to 1:25 with Milli-Q water, and adjusting them to pH 10 with NaOH. These stocks were divided into two (12 mL) aliquots in 20 mL glass vials and approximately 20 mg of solid NaBD₄ was added to one aliquot of each stock. All vials were loosely capped and placed in the dark at room temperature for 24 h. 6 M HCl was added to a final pH of 2.0 ± 0.1 and all aliquots were stored in the dark for an additional 24 h.

SPE cartridges were preconditioned with 3 mL MeOH, and 3 mL of 10 mM HCl (pH 2). 9.0 mL of the pH 2 aliquot was then loaded on the SPE cartridge and allowed to elute by gravity (about 3 min). The eluate from this process was clear, showing qualitatively that the colored material was retained by the solid phase. The cartridge was rinsed with an additional 3 mL of 10 mM HCl and dried using a steady flow of nitrogen (~10 mins). The colored, (borate-free) material was eluted using 1.5 mL MeOH into a 1.5 mL LC vial, which was capped and then stored in 0° C until mass spectrometric analysis.

Mass spectrometric data acquisition and pre-processing. Ultrahigh resolution mass spectra of all samples were obtained using a Bruker Apex QE 12 T FT-ICR mass spectrometer at the Helmholtz Center for Environmental Science (Munich, Germany). An Apollo II electrospray ionization (ESI) source was operated in negative ion mode (see previous chapter for experimental details). Mass spectra for SRFA, EACR, EAUW, EASO, and EADO were acquired by Dr. Mourad Harir, and mass spectra for PLFA, DELB, and DERV were acquired by Dr. Michael Gonsior. Both batches of samples had an associated extraction blank. An average resolution of >500,000 was obtained for all

spectra. For each spectrum, lists of m/z and intensities of peaks whose signal-to-noise ratio (S/N) was greater than 10 were generated and internally calibrated, doubly and triply-charged species as well as peaks which were present in high abundance in the associated extraction blanks were removed (see Chapter 4 for instrumental details and Appendix 1 for calibrant and blank subtraction details).

In the raw mass lists of DERV and DELB, there was a very high intensity peak observed at 495.3034 m/z , which was also observed in the extraction blank. However, the difference between the measured mass in the DERV and DELB samples and in the extraction blank was 0.24 ppm – slightly above the 0.2 ppm error limit that was used for all analyses in the previous chapter. Therefore, the error threshold used for this study was set to 0.3 ppm to allow for the subtraction of this problematic peak.

Molecular formula assignment and identification of ketone/aldehyde-containing species. Molecular formula assignment of peaks in the untreated samples (UNT) was accomplished using the same custom MATLAB function described in Chapter 4. The molecular formula assignment algorithm was based on the “low-mass moiety” approach described by Perdue and Green^{103,104}, and allowed up to 2 N and 1 S to be included in the calculated molecular formulae. Next, the mass spectrum of the corresponding BDR sample was searched for peaks at $m+3.021927$ and $m+6.043854$ masses, thereby identifying peaks (at mass m) in UNT as comprising non-, singly-, and/or doubly-reduced species. As a control, the mass spectrum of the UNT sample itself was searched for these mass markers. All above analyses were carried out with a 0.3 ppm error tolerance.

5.3. Results and Discussion

Comparison of mass spectrometric peak distributions. The ultrahigh resolution, negative ion ESI mass spectra of all samples contained between 7500 and 10,000 resolved peaks within the full scanned (50 – 2000) m/z range (Figure 5.1 and Table 5.1). With the exception of PLFA, reduction with borodeuteride resulted in an increase in the number of peaks observed in the full-scan mass spectrum. This increase arises from the generation of deuterium (D)-containing species that can be differentiated from all other species in the mass spectrum. In most cases, only a slight concomitant loss of the peaks corresponding to the precursors of these reduced species are observed, since all mass spectral peaks likely comprise many isomeric species which do not contain any reducible species. Conversely, there was a considerably greater number of peaks in the untreated PLFA sample than in any other sample. Also, the total ion count (TIC) for this sample was substantially less than all other samples, suggesting that the majority of peaks in PLFA have low relative intensities. In this case, if a peak in the untreated sample partially comprised reducible species and was of very low intensity, then the peak intensity would probably fall below S/N cutoff following reduction, and would therefore not be observed in the mass spectrum of the borodeuteride-reduced sample. Similarly, the intensity of the resulting $m+3.021927$ and/or $m+6.043854$ peak(s) may also not be sufficiently above the noise level and, ultimately, a net loss in the number of peaks would be observed.

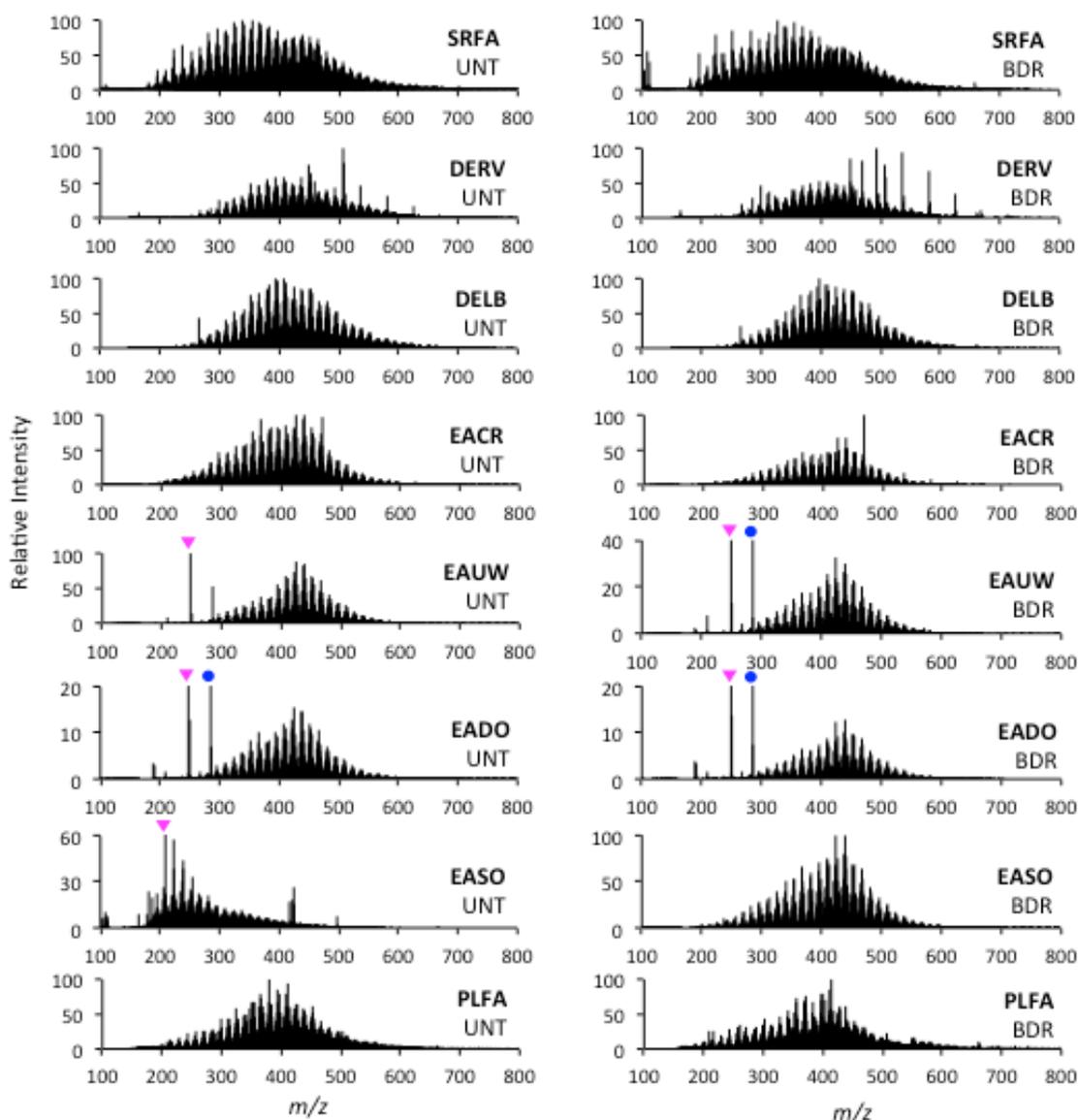


Figure 5.1. Full scan negative ion 12 T FT-ICR mass spectra for all samples included in this study. Purple triangles indicate the base peak in spectra with rescaled relative intensity axes ($249.03222\ m/z$ in EAUW and EADO; $209.00919\ m/z$ in EASO-UNT). Blue circles indicate secondary peaks that were also off-scale. All spectra exclude peaks which were doubly or triply charged, were present in the corresponding blank spectrum at high intensity, and/or had $S/N < 10$. No peaks with $m/z > 800$ or $m/z < 100$ were observed in any sample.

Table 5.1. Mass spectral peak distribution of all samples in the full (50 – 2000) and analyzed (250 – 550) m/z ranges.

| sample | [NaBD ₄] ^a | peaks | | TIC ^b (x10 ¹¹) | | AMW _N ^c | AMW _w ^d |
|----------|-----------------------------------|-------|----------|---------------------------------------|----------|-------------------------------|-------------------------------|
| | | full | analyzed | full | analyzed | | |
| SRFA-UNT | 0 | 8303 | 5353 | 2.82 | 2.40 | 408.4 | 436.9 |
| SRFA-BDR | 2.35 | 8974 | 6708 | 2.33 | 2.07 | 386.5 | 409.1 |
| DERV-UNT | 0 | 9286 | 6371 | 0.874 | 0.730 | 459.0 | 477.3 |
| DERV-BDR | 5.24 | 9675 | 8086 | 0.557 | 0.510 | 435.0 | 449.2 |
| DELB-UNT | 0 | 9710 | 7039 | 0.919 | 0.802 | 447.1 | 464.2 |
| DELB-BDR | 4.82 | 9923 | 8516 | 0.691 | 0.648 | 431.6 | 444.5 |
| EACR-UNT | 0 | 8431 | 5928 | 3.10 | 2.82 | 410.8 | 430.4 |
| EACR-BDR | 2.41 | 9452 | 7534 | 3.06 | 2.87 | 410.0 | 425.9 |
| EAUW-UNT | 0 | 8397 | 5216 | 5.38 | 4.89 | 436.8 | 453.0 |
| EAUW-BDR | 2.28 | 8048 | 6103 | 3.52 | 3.26 | 420.7 | 435.9 |
| EADO-UNT | 0 | 7505 | 4891 | 3.41 | 2.98 | 420.2 | 441.9 |
| EADO-BDR | 2.33 | 7858 | 5940 | 3.29 | 2.94 | 419.4 | 437.6 |
| EASO-UNT | 0 | 7213 | 5545 | 2.09 | 1.43 | 301.7 | 323.5 |
| EASO-BDR | 2.49 | 8030 | 6309 | 2.69 | 2.52 | 418.0 | 433.3 |
| PLFA-UNT | 0 | 9563 | 7715 | 0.660 | 0.605 | 418.4 | 434.8 |
| PLFA-BDR | 4.81 | 8917 | 7892 | 0.407 | 0.385 | 403.8 | 418.1 |

^a concentration in mg/mL; ^b total ion count; ^c number and ^d weight averaged molecular weight (see Appendix 9, section A9.9 for calculations).

Two other phenomena that were observed in all samples except for EASO were decreases in both TIC and average molecular weight following borodeuteride reduction. The loss of TIC for all samples ranged from 1.3% to 38% of the TIC of the untreated samples (for EACR and PLFA, respectively). In contrast, the TIC and average molecular weight for EASO both *increased* substantially following reduction. While low m/z distributions has been observed by others in DOM samples that have been extensively

photodegraded^{5,93} (as surface marine samples are expected to be), a similarly low m/z distribution is *not* observed for the corresponding borodeuteride reduced sample. In fact, the reduced sample displayed a distribution of peaks very similar to those of all other samples used in this study. Such a dramatic shift following reduction is very unlikely, and this extreme observation calls into question the spectral quality of the EASO samples, especially since the ion optics of the FT-ICR instrument was optimized to give a maximum signal in the 300 – 400 m/z region. Therefore, EASO was excluded in all further analyses. Since there were very few peaks below 150 m/z for all samples, and since molecular formulae assignments typically become less definitive at higher m/z , the 250 – 550 m/z range was chosen for all subsequent analysis. This m/z region contained an average of 75% of all peaks and 90% of the total ion count (TIC) in the measured mass spectra (Table 5.1).

Comparison of molecular formula assignments. Molecular formulae were assigned to $67 \pm 4\%$ (average \pm standard deviation) of all peaks detected in the mass spectra of samples. Although this percentage is lower than previous studies,^{9,83} the average percentage of the total ion count (TIC) that was assigned was $84 \pm 4\%$, indicating that most unassigned peaks were of low relative intensity (Table 5.2). The majority (72.1%) of formulae identified in SRFA contained only C, H, and O atoms, whereas this percentage was substantially lower for PLFA (58.7%) and all other DOM extracts ($53.5 \pm 6.2\%$), indicating that there were substantially fewer heteroatom-containing species in SRFA than in all other samples. This difference between the two fulvic acid standards has been observed by D'Andrilli et al.,³⁰ who identified 84.5% and 25.9% of all formula as containing only C, H, and O for SRFA and PLFA, respectively. The significant

differences between the composition of the fulvic acid standards reported here and the previous study by D'Andrilli et al. are likely due to slight differences in the detection of low intensity peaks (see Appendix 8), and thus do not necessarily indicate a major discrepancy between the analyses in the two studies.

Table 5.2. Summary of molecular formulae identified in all untreated samples in the 250 to 550 m/z range.

| | SRFA | DERV | DELB | EACR | EAUW | EADO | PLFA |
|----------------------------------|-------|-------|-------|-------|-------|-------|-------|
| num. formulae | 3329 | 4528 | 5219 | 4152 | 3364 | 3265 | 5039 |
| % peaks ^a | 62.2 | 71.1 | 74.1 | 70.0 | 64.5 | 66.8 | 65.3 |
| % TIC ^a | 89.4 | 79.7 | 83.4 | 85.2 | 82.9 | 85.3 | 78.8 |
| % CHO | 72.1 | 58.9 | 44.7 | 59.7 | 53.1 | 51.0 | 58.7 |
| % CHON | 20.6 | 22.9 | 22.7 | 29.2 | 33.2 | 31.4 | 35.0 |
| % CHONS | 1.9 | 0.5 | 5.3 | 0.6 | 1.9 | 2.0 | 0.8 |
| % CHOS | 5.4 | 17.7 | 27.3 | 10.6 | 11.8 | 15.6 | 5.4 |
| % C _{mass} ^b | 55.3 | 57.0 | 56.1 | 57.3 | 57.0 | 57.2 | 56.5 |
| % H _{mass} ^b | 5.1 | 5.7 | 6.0 | 6.0 | 6.2 | 6.2 | 6.4 |
| % O _{mass} ^b | 39.3 | 36.3 | 35.9 | 36.0 | 36.0 | 35.7 | 33.8 |
| % N _{mass} ^b | 0.2 | 0.33 | 0.47 | 0.44 | 0.44 | 0.45 | 1.52 |
| % S _{mass} ^b | 0.1 | 0.69 | 1.55 | 0.34 | 0.37 | 0.50 | 1.77 |
| H/C _{avg.} ^b | 1.091 | 1.188 | 1.276 | 1.238 | 1.287 | 1.292 | 1.338 |
| O/C _{avg.} ^b | 0.542 | 0.487 | 0.487 | 0.475 | 0.478 | 0.471 | 0.459 |
| N/C _{avg.} ^b | 0.003 | 0.005 | 0.008 | 0.007 | 0.007 | 0.007 | 0.025 |
| DBE _{avg.} ^b | 9.52 | 9.43 | 8.39 | 8.45 | 8.34 | 8.26 | 7.64 |
| AI _{avg.} ^b | 0.073 | 0.063 | 0.039 | 0.055 | 0.022 | 0.026 | 0.035 |
| error (ppm) ^c | 0.102 | 0.107 | 0.112 | 0.105 | 0.127 | 0.116 | 0.116 |

^a percentage of peaks in range to which a formula was assigned; ^b relative intensity weighted; ^c root mean square error.

The intensity-weighted average H/C and O/C molar ratios, aromaticity index (AI_{avg} ; calculated as described by Koch et al.,¹¹⁶ see Appendix 9, Section A9.9), and double-bond equivalent (DBE_{avg}) were calculated for all DOM samples. While H/C_{avg} generally increased with increasing distance from terrestrial sources, O/C_{avg} and DBE_{avg} generally decreased. These trends suggest that, in general, there were more aliphatic species present in marine DOM than in riverine DOM. One possible interpretation is that riverine DOM contains primarily lignin/tannin-derived material, while marine DOM contains species such as carboxyl-rich alicyclic molecules (CRAM) and other sterol-derived materials, which would have formulae with slightly higher H/C and lower O/C molar ratios. In addition to these trends, AI_{avg} generally decreased from riverine to marine-like DOM samples with two notable irregularities. First, AI_{avg} for DELB was substantially lower than expected. This may suggest that a unique source of aliphatic material is partially responsible for the DOM within the Lower Delaware Bay, perhaps a local microbial or algal source. Second, PLFA had a higher AI_{avg} than what would be expected for primarily microbially-produced DOM. Clearly, use of these bulk measurements as unequivocal indicators for relative contributions of riverine or marine sources to a series of DOM samples would be an oversimplification and more detailed analyses are needed. However, the trends with location and H/C_{avg} , O/C_{avg} , DBE_{avg} , and AI_{avg} can still be useful for general comparisons among DOM samples.

Comparison of reduced and non-reduced species within samples. Using the previously developed algorithms, peaks in each sample to which molecular formulae were assigned were identified as comprising species with zero, one, and/or two reducible moieties. Figure 5.2A shows the percentage of those peaks which were falsely identified

as reduced based on a search for borodeuteride-reduced ($m+3.021927$ or $m+6.043854$) species in the untreated sample itself (control search), and Figure 5.2B shows those based on a search in the corresponding borodeuteride-reduced sample (experimental search). In this study, the $[\text{NaBD}_4]$ used in the reduction of PLFA, DERV, and DELB was about twice of $[\text{NaBD}_4]$ used in the reduction of SRFA, EACR, EAUW, and EADO. However, it was demonstrated in the previous chapter that most species with one or more reducible groups are reduced *at least* once under low $[\text{NaBD}_4]$ while higher $[\text{NaBD}_4]$ primarily induces multiple reductions on the same species. Therefore, while the number of species which were identified as containing *multiple* reducible groups within all samples may not be comparable, the *total* number of identified reducible species is expected to be reasonably comparable.

In the control search, only DELB, EAUW, and PLFA gave an appreciable number of peaks falsely identified as comprising reducible species, the majority of which resided in the highest (450 – 550) m/z range. The relatively high percentages (1.1% to 4.3%) of false assignments in these three samples are most likely due to the large numbers of peaks detected in the untreated mass spectra, which would increase the chance that the m/z of a peak would coincide (within the employed mass tolerance window) with the expected m/z of a reduced (deuterated) species. In the experimental searches, the percentage of peaks in the middle and high m/z ranges that were identified as comprising ketone/aldehyde-containing species decreased with increasing distance from possible terrestrial sources, with PLFA being an exception. This decrease suggests that the majority of ketone/aldehyde-containing species are terrestrially derived, which have been attributed to the long wavelength absorption and emission of CDOM.³ While this trend is

not as clear at the low m/z range for all seven samples, it does hold among the two Delaware River extracts and three Equatorial Atlantic extracts. Assuming that terrestrial sources are a primary source of ketone/aldehyde-containing species, the high number of species identified as such in PLFA is surprising. However, the very low absorption of PLFA relative to more terrestrially-influenced ensembles of CDOM^{114,115} suggests that not all of these reducible species contribute significantly or in the same manner to the bulk optical properties and thus may differ chemically from terrestrially-derived ketones/aldehydes.

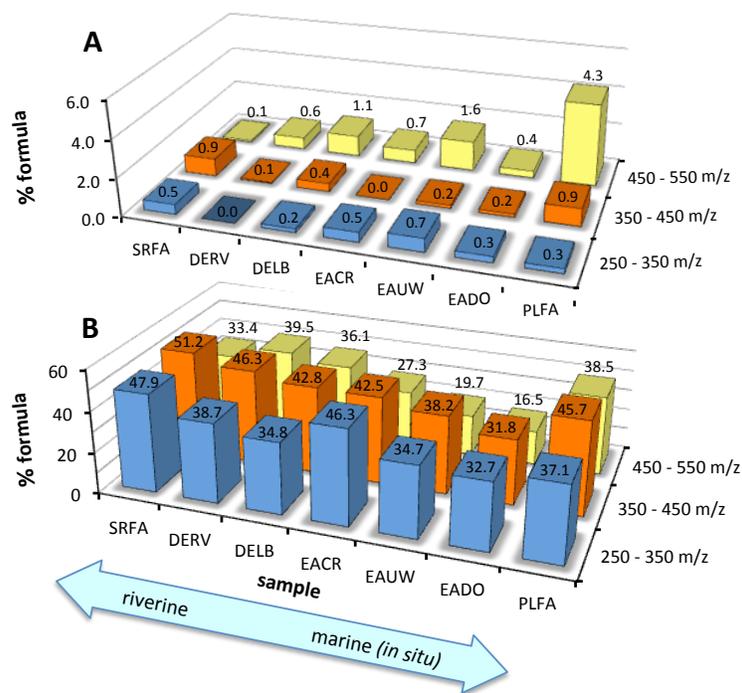


Figure 5.2. Percentage of peaks with assigned molecular formulae in all 7 samples in the three m/z ranges identified as comprising at least one reducible species (**A**: control search, **B**: experimental search).

Plots of H/C versus O/C molar ratios of the identified molecular formulae (Van Krevelen plots) reveal that the terrestrially-influenced samples (SRFA, DERV, DELB, and EACR) were rich in formulae with relatively high O/C (>0.4) and low H/C (<1.0), while the more isolated, marine-like samples (EAUW, EADO, and PLFA) were largely devoid of such formulae. Similarly, the presence of formulae with similarly high O/C and low H/C molar ratios in terrestrial/riverine DOM samples has been reported by others, while many marine DOM samples have been shown to lack such formulae.^{44,85,113,117} It is possible, therefore, to expect that a distribution of peaks at in this O/C and H/C region can be attributed to tannins and/or oxidized lignin-derived compounds from high-order plant source material and can thus serve as a useful marker for terrigenous organic material, especially if the mass spectral peaks assigned to such formulae are identified as ketone/aldehyde-containing species. However, species identified as “doubly-reduced only” are rare in this region. As in the previous chapter, this can be attributed to a high degrees of isobaric complexity of peaks in this region and a high degree of chemical heterogeneity for those reducible species.

It should be stressed, that while high O/C (>0.4) and low H/C (<1) may be useful as markers for terrestrial plant-derived material, the formulae which occur in the middle O/C and H/C molar ratios (e.g. near O/C = 0.5 and H/C = 1.0) is not necessarily attributable to lignin species, despite the frequent allocation of these formulae as such by others.^{113,118,119} As noted by these and other researchers, classifying formulae in this Van Krevelen region may be misleading, since a very large number of compositional isomers are possible for the elemental combinations with middle O/C and H/C ratios, and

therefore peaks in this region may be indicative of other classes of compounds such as CRAM.^{67,90}

At low m/z , a wide distribution of H/C and O/C molar ratios were observed for all samples. At higher m/z , H/C and O/C distribution becomes more tightly distributed, and more dramatic differences are observed among samples. In contrast to the largely terrigenous samples, EAUW, EADO, and PLFA contained predominantly formulae with high H/C. Some species at low O/C and H/C molar ratios were present in all samples except for PLFA, which have been previously attributed to condensed aromatic species derived from biochar. There was a large number of species identified as doubly-reduced only in PLFA, especially at high m/z . This is unlike the reducible species in the more terrigenous samples, which primarily comprised species which were singly-reduced only or *both* singly and doubly-reduced. This apparent difference in isobaric complexity and/or chemical heterogeneity supports the presence of a microbially-derived pool of ketones/aldehydes within the marine samples.

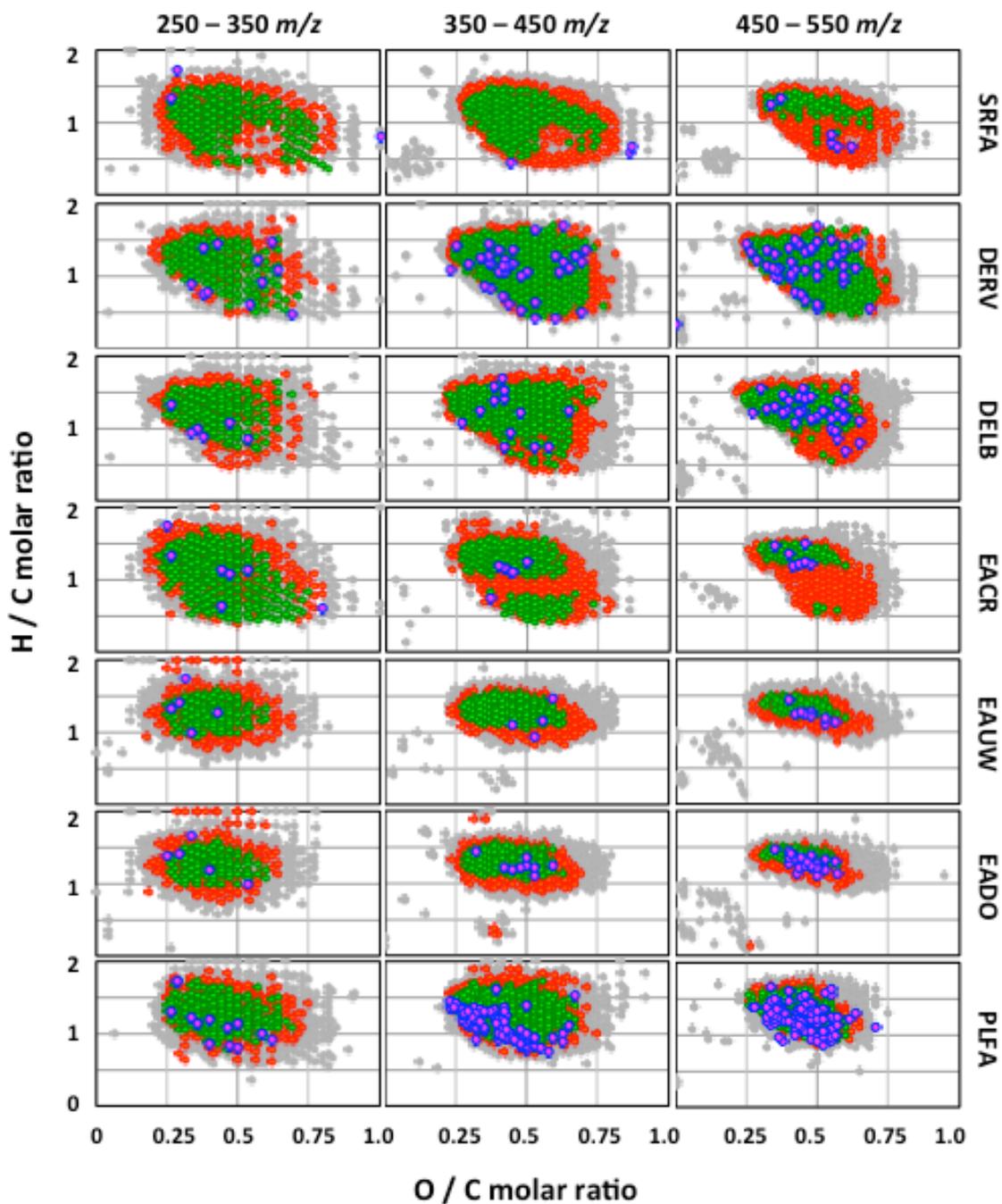


Figure 5.3. Van Krevelen plots of all formulae identified in all mass spectra within three m/z ranges (grey: non-reduced; red: singly-reduced only; green: doubly- and singly-reduced; blue/purple: doubly-reduced only).

Commonality among samples. In order for SRFA and PLFA to be suitable end-members for riverine and *in situ* sources, respectively, DOM samples collected near riverine outputs should have many compounds (or molecular formulae) in common with SRFA, while those collected in more isolated marine environments should have more compounds in common with PLFA than with SRFA. Depicted in Figure 5.4A are the percentages of formulae identified in each of the eight samples that were also identified in SRFA (red solid lines/points) and PLFA (blue dashed lines/points).

EACR had a slightly greater percentage of formulae (65%) in common with SRFA than both Delaware River/Bay extracts. Perhaps more surprisingly, the two remote oceanic samples (EAUW and EADO) had higher percentages of formulae in common with SRFA than DELB! These observations are further evidence of a secondary source of DOM within the Delaware River/Bay samples, especially for DELB. Also, the similarities and differences among SRFA, EACR, and the two Delaware River/Bay samples may be related to watershed characteristics. That is, while the Delaware River traverses through hardwood forest regions with significant industrial, agricultural, and commercial developments, the Suwannee River and Congo River watersheds reside within tropical climates and are characterized by an abundance of mangrove swamps with little or no anthropogenic influences.

The percentage of formulae in each sample that was also identified in PLFA increased monotonically as the proximity of the sample collection locations to riverine sources decreased. In fact, 78.5% of formulae identified in EADO was also within PLFA, which represented the greatest degree of similarity between all samples in this study and the fulvic acid end-members. This high commonality is similar to the peak

reproducibility of replicate measurements of a single DOM sample reported elsewhere.^{43,58,82,91} Although PLFA was extracted from a eutrophic coastal pond which is considerably different from oceanic environments, the aforementioned observations suggests that PLFA may be useful for estimations of the relative amounts of *in situ*-produced material in the ultrahigh resolution mass spectra of a suite of DOM samples.

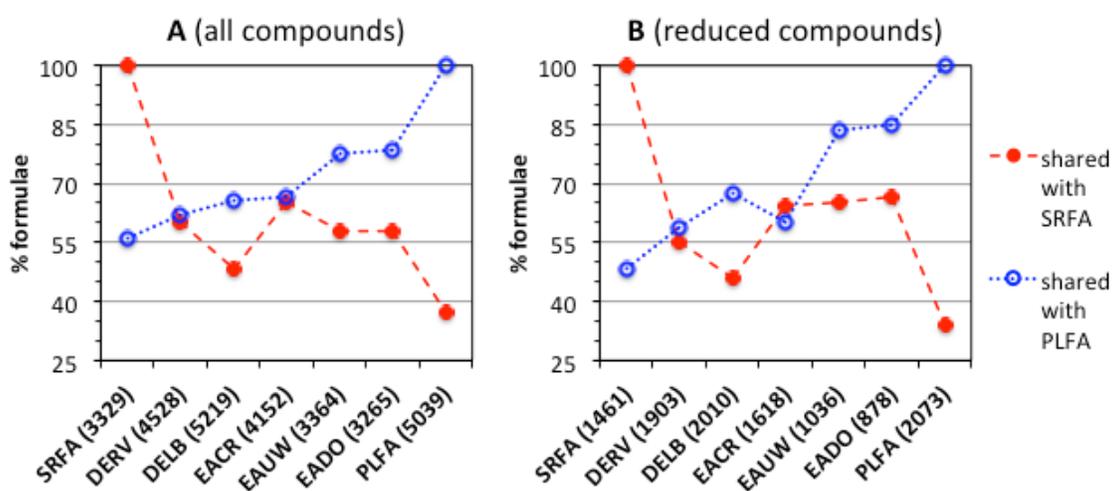


Figure 5.4. Percentage of formulae identified in each sample which were also identified in SRFA and PLFA. **A:** all formulae, **B:** formulae identified as peaks comprising reducible species. In parenthesis are the total numbers of formulae in each sample within the studied 250 – 550 m/z range.

The 37.1% of formulae in PLFA that was also found in SRFA was the lowest observed similarity. This small but significant commonality between the two fulvic acid end-members suggests that there may exist a subset of formulae that are likely to be present in nearly all similarly complex mixtures of organic compounds that would be detected by ESI mass spectrometry. Likewise, a set of stable combinations of elements and molecular masses was observed in many marine DOM samples by Lechtenfeld and

coworkers, and were thus referred to as an “island of stability.”^{120,121} Indeed, Hertkorn et al.⁵² have explained that since complex samples such as DOM typically comprise a large portion of the total possible formula compositional space (i.e., all chemically relevant combinations of C, H, and O), signal overlap by dissimilar DOM samples is likely. As a consequence, rigorous comparisons between the ESI mass spectra of several DOM samples is likely to be insufficient for revealing detailed structural differences. As a result, DOM samples are likely to appear more similar to one another by such comparisons than they are in reality.

Depicted in Figure 5.4B are the percentages of formulae identified as reducible species in each sample that were also identified as such in SRFA and/or PLFA. These plots were similar to that which considered all formulae (Figure 5.4A) with some notable differences. First, the percentage of reducible compounds in common with PLFA increased more sharply with proximity to riverine outputs than did the percentage of all compounds. Second, the percentage of reducible compounds in EACR that were also identified as such in PLFA was lower than expected, based on the observed trend. Finally, the percentage of reducible species in common with SRFA was relatively constant for all three Equatorial Atlantic samples. It appears that the reducible compounds within SRFA and PLFA can act as end-members for the reducible species in this series of DOM compounds. Therefore, it is reasonable to assume that there are two terrestrially-derived and microbially-produced pools of reducible species. As demonstrated here, this mass labeling method offers a relatively easy way to gain detailed structural information by ultrahigh resolution ESI mass spectrometry, and can thus be helpful in distinguishing among sets of DOM samples that have extensive signal overlap

within their unaltered mass spectra. It is important, however, that replicate measurements be taken in studies such as these in order to assess the uncertainty in all analyses prior to drawing conclusions regarding the compositional changes that occur among DOM samples.

5.4. Conclusions

In this study, ketone/aldehyde-containing species within eight DOM extracts were identified and characterized via mass labeling by sodium borodeuteride and analysis by ultrahigh resolution ESI FT-ICR mass spectrometry. The results suggest that many ketone/aldehyde-containing species were riverine derived lignin/tannin-like compounds in most samples, but also that a second pool of more unsaturated and less oxygenated ketones/aldehydes are produced from microbial sources. This compositional relationship between riverine and marine DOM is apparent for the mass labeled (reduced) species, but less so for the entire ensemble of compounds identified solely by molecular formula assignment. This demonstrates that the mass labeling method reveals compositional details that would be not otherwise observed by considering only the ensemble molecular formulae identified via ESI FT-ICR mass spectrometry.

An important distinction to make between the two pools of ketone/aldehyde-containing species is that reduction of the terrestrially-produced pool has been known to result in substantial changes in the bulk optical properties due to disruption of inter/intra-molecular charge transfer interactions, whereas reduction of ketones/aldehydes within microbially-produced species would not (see discussion in Chapter 3). Future comparisons of the changes in optical properties that occur under borohydride reduction

of CDOM from riverine and oceanic samples would therefore prove very useful to test this hypothesis. It is expected that the use of other bulk methods of structural analyses such as infrared spectroscopy and/or nuclear magnetic resonance, combined with optical methods as well as this mass labeling method, can be of further use to confirm the relative presence of these two major pools of ketone/aldehyde-containing species within CDOM.

Chapter 6: Conclusions and Future Work

Summary of the method. This dissertation describes the development and application of a novel method to mass label ketone and aldehyde-containing compounds in various ensembles of natural dissolved organic matter (DOM). In this technique, permanently reducible moieties are reduced by sodium borodeuteride (NaBD_4) and purified via solid phase extraction. Untreated and reduced aliquots of the sample of interest are then analyzed via ultrahigh resolution electrospray ionization mass spectrometry. Peaks in the mass spectrum of the untreated sample to which molecular formulae are assigned are identified by searching the mass spectrum of the borodeuteride-reduced sample for peaks which are 3.0219 and/or 6.0438 m/z higher, which corresponds to singly and doubly reduced derivatives, respectively. This data analysis and interpretation has been automated via custom MATLAB function code provided in the appendices of this dissertation. This code was written with the intention of allowing other researchers to readily apply this mass labeling method in future work involving the characterization of natural DOM.

Future work. Implementation of this technique is intended for a current project involving the molecular characterization and comparison of DOM extracts collected from the Southern Pacific Ocean. In this project, this mass labeling method will be used to compare the structure and composition of ketone/aldehyde-containing species within DOM which reside at various depths. Moreover, attempts can be made to semi-quantitatively link the presence of ketone/aldehyde-containing species detected by this method to the bulk optical properties of these materials.

Although the development of this analytical methodology was done using electrospray ionization (ESI) exclusively, this mass labeling method can be combined with other soft ionization methods such as atmospheric chemical ionization and atmospheric pressure photoionization, which have been used by some researchers to provide additional information by ionizing species which are not as readily ionized by electrospray ionization. Such work would be expected to offer even further insight into the structure and composition of ketone and aldehyde-containing species within DOM.

Perhaps the most unresolved phenomenon that was observed many times in these initial studies is the effect of the ionization efficiency of compounds following borodeuteride reduction. During these experiments, it was observed that the mass spectrum of the reduced samples had substantially lower signal intensities than that of the corresponding untreated sample. It may be possible to investigate this phenomenon by comparing the relative ionization efficiencies of a suite of untreated and borodeuteride-reduced model compounds. Efforts to investigate this phenomenon could possibly be accomplished by comparing the instrumental response of individual model compounds to that of a standard reference compound, similar to the work done by Lieto et al.^{122,123} Furthermore, it is possible to adapt the identification algorithms developed here to give semi-quantitative information about the degree of reduction, rather than discrete assignments of reduced species, in a method similar to what has been done for deuterium labeling of acidic species in DOM,⁸⁹ thus allowing for an approximation of the abundance of irreversibly reducible species relative to that of the non reducible species which comprise a single mass spectral peak.

In summary, it is expected that this mass labeling method will be of great use for future efforts involving the characterization of DOM, especially when a multi-instrumental approach is used.¹²⁴⁻¹²⁶ Specifically, the identification and analysis of the ketone and aldehyde-containing species DOM which is demonstrated here should be able to provide a more detailed view of how they contribute to the optical properties and photochemical reactivity of these ensembles.

Appendix 1: Internal calibrants and extraction blanks

Table A1.1. m/z and corresponding molecular formula (of $[M-H]^-$ ions) used for post-acquisition calibration of the 7 T ESI FT-ICR mass spectra (calibration performed by Dr. Krista Longnecker, Woods Hole Oceanographic Institution).

| m/z | molecular formula |
|----------|----------------------|
| 337.1657 | $C_{18}H_{25}O_6$ |
| 393.1555 | $C_{20}H_{25}O_8$ |
| 407.1711 | $C_{21}H_{27}O_8$ |
| 421.1868 | $C_{22}H_{31}O_8$ |
| 435.2024 | $C_{24}H_{33}O_8$ |
| 463.2337 | $C_{25}H_{35}O_8$ |
| 477.2494 | $C_{26}H_{37}O_8$ |
| 491.2650 | $C_{27}H_{39}O_8$ |
| 505.2807 | $C_{28}H_{41}O_8$ |
| 609.2189 | $C_{29}H_{37}O_{14}$ |

Table A1.2. *m/z* and corresponding molecular formula (of [M-H]⁻ ions) used for post-acquisition calibration of the 12 T ESI FT-ICR mass spectra (calibration performed by Dr. Michael Gonsior, University of Maryland Center for Environmental Sciences).

| <i>m/z</i> | molecular formula |
|-------------|-------------------|
| 201.0404617 | C8 H9 O6 |
| 225.0040764 | C9 H5 O7 |
| 251.0197264 | C11 H7 O7 |
| 275.0561117 | C14 H11 O6 |
| 301.0353764 | C15 H9 O7 |
| 325.0928910 | C15 H17 O8 |
| 351.1085410 | C17 H19 O8 |
| 375.0721556 | C18 H15 O9 |
| 401.0514202 | C19 H13 O10 |
| 425.1089348 | C19 H21 O11 |
| 451.1245848 | C21 H23 O11 |
| 475.0881995 | C22 H19 O12 |
| 501.0674641 | C23 H17 O13 |
| 525.0674641 | C25 H17 O13 |
| 551.0831141 | C27 H19 O13 |
| 575.1042433 | C26 H23 O15 |
| 601.0835080 | C27 H21 O16 |
| 625.0835080 | C29 H21 O16 |
| 651.0991580 | C31 H23 O16 |
| 675.0991580 | C33 H23 O16 |
| 701.0995518 | C31 H25 O19 |

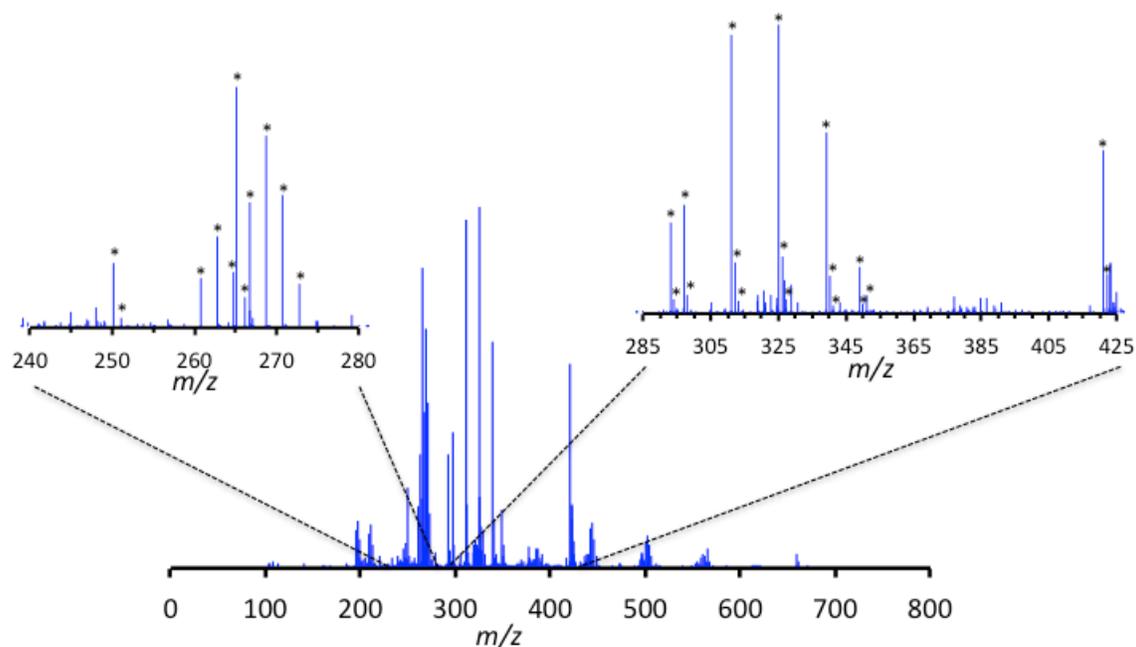


Figure A1.1. Negative ion ESI FT-ICR mass spectrum of the extraction blank associated with SRFA, EACR, EAUW, EADO, and EASO. Asterisks (*) denote m/z values which were excluded from sample peak lists.

Table A1.3. m/z and tentative assignments of peaks present in the extraction blank shown in Figure A1.1 which were removed from the mass spectra of SRFA, EACR, EAUW, EADO, and EASO (untreated and reduced).

| m/z | assignment | m/z | assignment |
|-----------|----------------------------------------------|-----------|------------------------------------------------|
| 250.14489 | unknown 1 (monoisotopic) | 297.15301 | sulfonate 1 (monoisotopic) |
| 251.14825 | unknown 1 ($+^{13}\text{C}-^{12}\text{C}$) | 298.15637 | sulfonate 1 ($+^{13}\text{C}-^{12}\text{C}$) |
| 260.78697 | chloride adduct 1a | 299.14881 | sulfonate 1 ($+^{34}\text{S}-^{32}\text{S}$) |
| 262.78401 | chloride adduct 1b | 311.16866 | sulfonate 2 (monoisotopic) |
| 264.78106 | chloride adduct 1c | 312.17202 | sulfonate 2 ($+^{13}\text{C}-^{12}\text{C}$) |
| 265.14792 | unknown 2 (monoisotopic) | 313.16446 | sulfonate 2 ($+^{34}\text{S}-^{32}\text{S}$) |
| 266.15128 | unknown 2 ($+^{13}\text{C}-^{12}\text{C}$) | 325.18431 | sulfonate 3 (monoisotopic) |
| 266.80391 | chloride adduct 2a | 326.18767 | sulfonate 3 ($+^{13}\text{C}-^{12}\text{C}$) |
| 268.80096 | chloride adduct 2b | 327.18011 | sulfonate 3 ($+^{34}\text{S}-^{32}\text{S}$) |
| 270.79801 | chloride adduct 2c | 339.19997 | sulfonate 4 (monoisotopic) |
| 272.79506 | chloride adduct 2d | 340.20333 | sulfonate 4 ($+^{13}\text{C}-^{12}\text{C}$) |
| 293.17922 | unknown 3 (monoisotopic) | 341.19577 | sulfonate 4 ($+^{34}\text{S}-^{32}\text{S}$) |
| 294.18258 | unknown 3 ($+^{13}\text{C}-^{12}\text{C}$) | 421.14288 | unknown 4 (monoisotopic) |
| | | 422.14624 | unknown 4 ($+^{13}\text{C}-^{12}\text{C}$) |

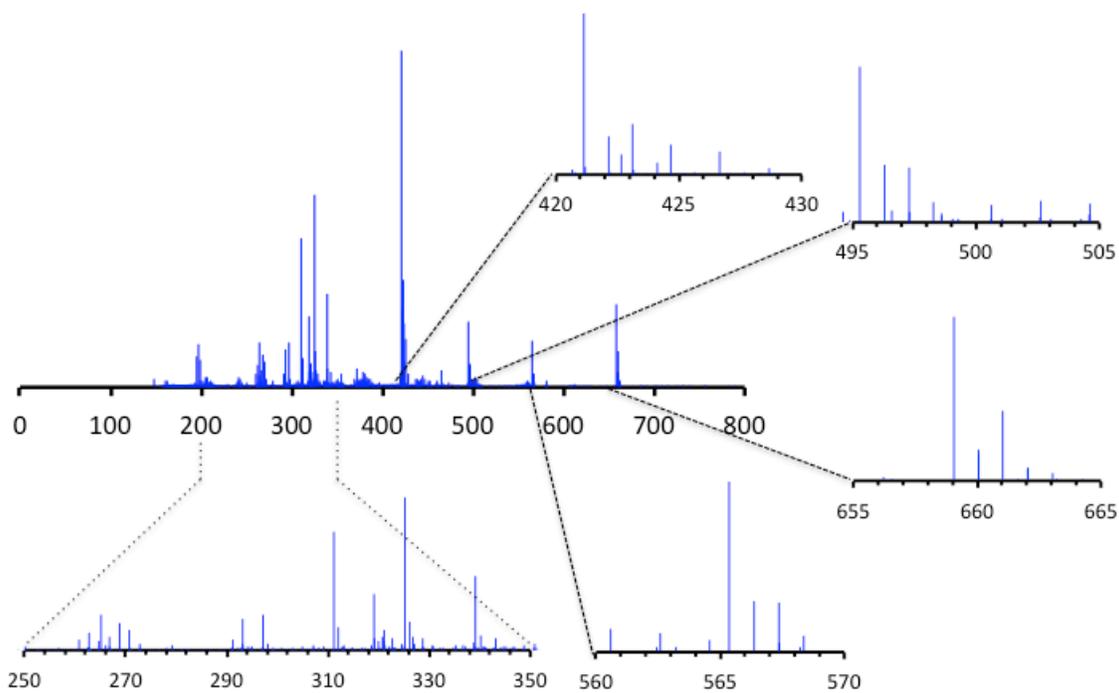


Figure A1.2. Negative ion ESI FT-ICR mass spectrum of the extraction blank associated with PLFA, DELB, and DERV.

Table A1.4. *m/z* and tentative assignments of peaks present in the extraction blank shown in Figure A1.2 which were removed from the mass spectra of PLFA, DELB, and DERV (untreated and reduced).

| <i>m/z</i> | assignment | <i>m/z</i> | assignment |
|------------|--------------------------------------------------|------------|--------------------|
| 260.78686 | Cl adduct 1a | 422.14555 | unknown cluster 1b |
| 262.78390 | Cl adduct 1b | 422.66936 | unknown cluster 1c |
| 264.78096 | Cl adduct 1c | 423.13924 | unknown cluster 1d |
| 265.14782 | unknown 1 | 424.66733 | unknown cluster 1e |
| 266.15118 | unknown 1 (+ ¹³ C- ¹² C) | 426.66528 | unknown cluster 1f |
| 266.80381 | chloride adduct 2a | 428.66323 | unknown cluster 1g |
| 268.80086 | chloride adduct 2b | 495.30340 | unknown cluster 2a |
| 270.79790 | chloride adduct 2c | 496.30675 | unknown cluster 2b |
| 272.79497 | chloride adduct 2d | 496.61539 | unknown cluster 2c |
| 291.20956 | unknown 2 | 497.30046 | unknown cluster 2d |
| 293.17913 | unknown 3 | 498.30381 | unknown cluster 2e |
| 294.18249 | unknown 3 (+ ¹³ C- ¹² C) | 498.61248 | unknown cluster 2f |
| 297.15292 | sulfonate 1 | 500.63527 | unknown cluster 2g |
| 298.15628 | sulfonate 1 (+ ¹³ C- ¹² C) | 502.63234 | unknown cluster 2h |
| 299.14872 | sulfonate 1 (+ ³⁴ S- ³² S) | 504.59342 | unknown cluster 2i |
| 311.16856 | sulfonate 2 | 565.36978 | unknown cluster 3a |
| 312.17192 | sulfonate 2 (+ ¹³ C- ¹² C) | 566.37314 | unknown cluster 3b |
| 313.16436 | sulfonate 2 (+ ³⁴ S- ³² S) | 567.36688 | unknown cluster 3c |
| 319.09531 | unknown 4 | 659.05664 | unknown cluster 4a |
| 325.18421 | sulfonate 3 | 660.05622 | unknown cluster 4b |
| 326.18757 | sulfonate 3 (+ ¹³ C- ¹² C) | 660.05994 | unknown cluster 4c |
| 327.18001 | sulfonate 3 (+ ³⁴ S- ³² S) | 661.05362 | unknown cluster 4d |
| 339.19984 | sulfonate 4 | 662.05309 | unknown cluster 4e |
| 340.20320 | sulfonate 4 (+ ¹³ C- ¹² C) | 662.05309 | unknown cluster 4f |
| 421.14220 | unknown cluster 1a | | |

Appendix 2: Full scan negative ion mass spectra

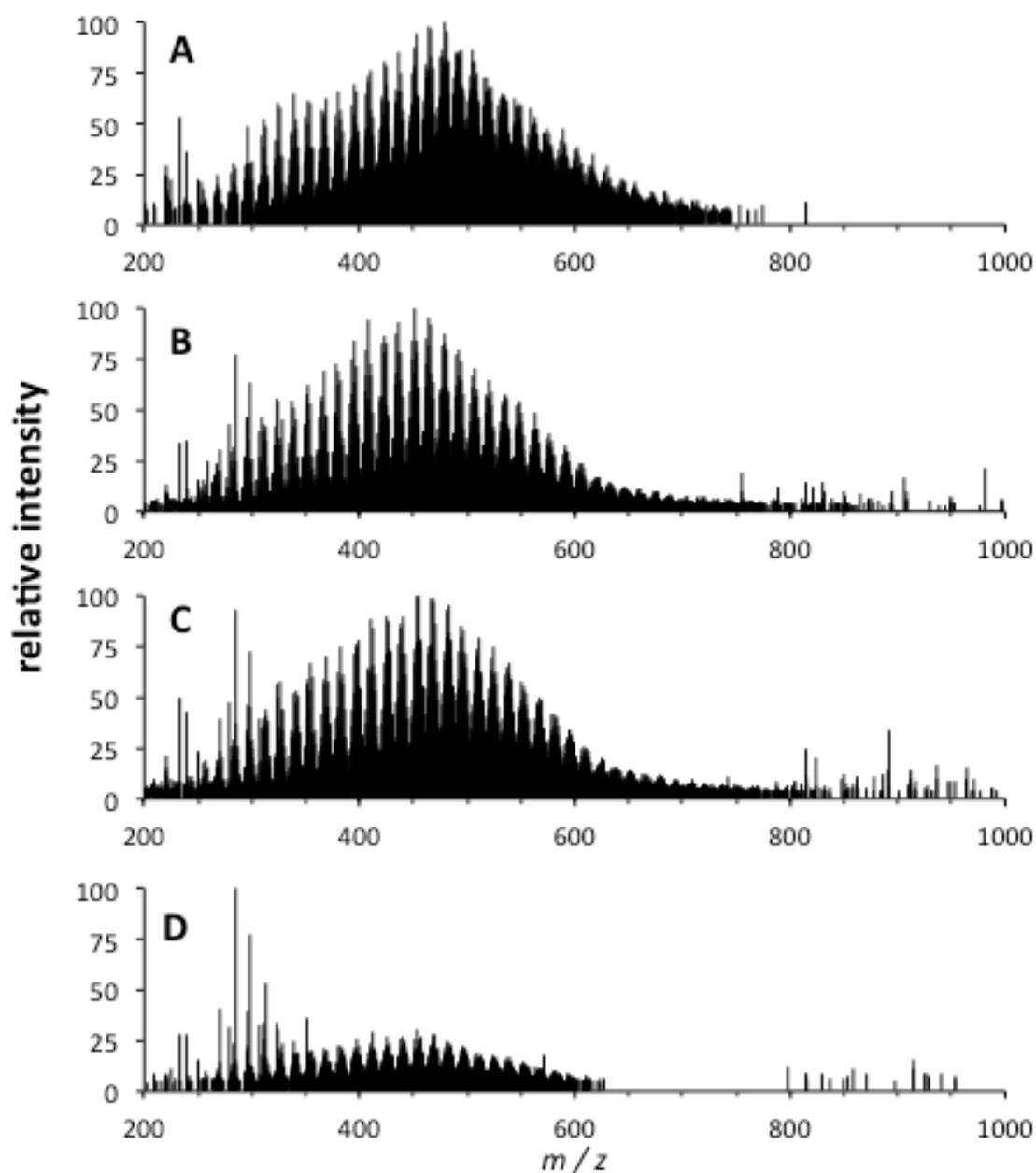


Figure A2.1. Negative ion 7 T ESI FT-ICR mass spectra of (A) SRFA UNT replicate 2, DEUB (B) UNT, (C) BHR, and (D) BDR.

Table A2.1. Number of peaks and total ion count (TIC) of the full scanned range (200 – 1000 m/z) and the analyzed range (200 – 600 m/z shaded rows) for all 7 T ESI FT-ICR mass spectra.

| sample | peaks | TIC (x 10 ⁵) |
|------------|-------|---------------------------|
| SRFA-UNT-1 | 3245 | 11.6 |
| | 2707 | 10.3 |
| SRFA-UNT-2 | 2872 | 17.6 |
| | 2361 | 15.6 |
| SRFA-BHR | 2622 | 14.1 |
| | 2164 | 12.4 |
| SRFA-BDR | 2036 | 6.84 |
| | 1994 | 6.73 |
| DEUB-UNT | 4378 | 25.0 |
| | 3407 | 21.9 |
| DEUB-BHR | 4437 | 18.0 |
| | 3421 | 15.6 |
| DEUB-BDR | 2242 | 9.45 |
| | 2194 | 9.30 |

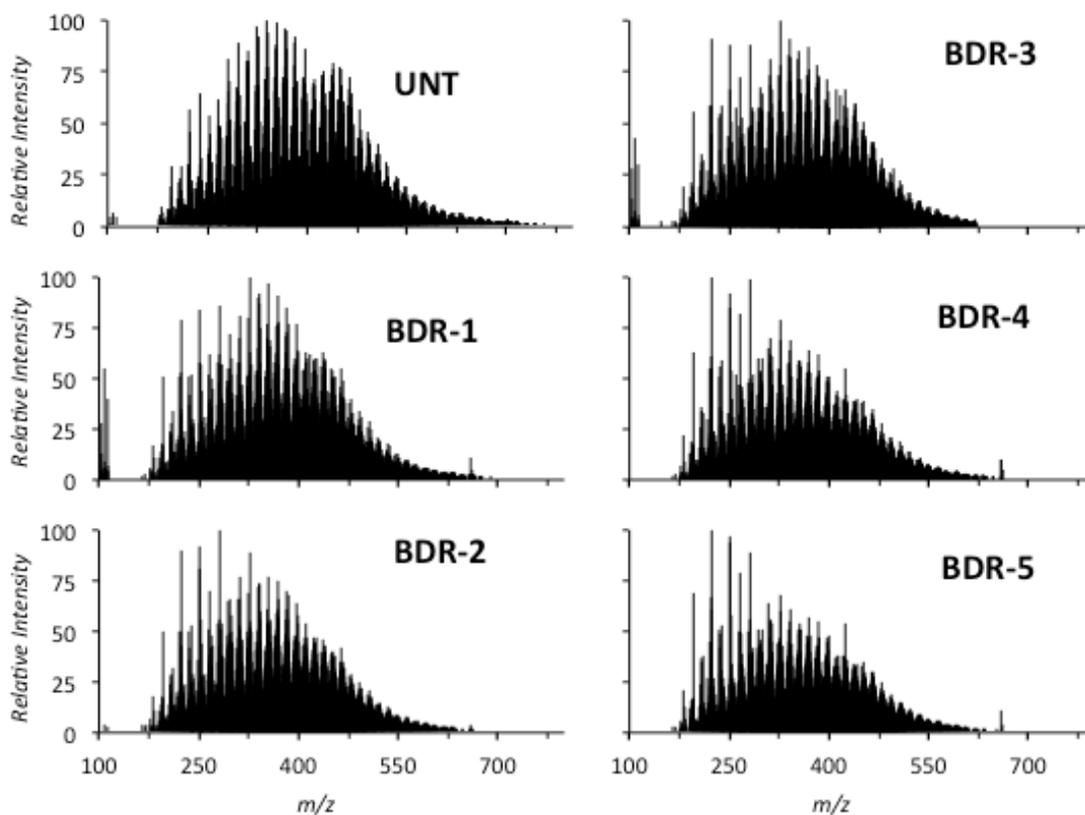


Figure A2.2. 12 T ESI FT-ICR mass spectra of SRFA UNT and BDR-1 through BDR-5. All spectra depicted exclude peaks which were doubly or triply charged, were present in the blank spectrum at high intensity, and/or had $S/N < 10$. No peaks with $m/z > 800$ were observed in any sample.

Appendix 3: Peak and molecular formula lists of UNT SRFA at
selected nominal masses (for Chapter 3)

Table A3.1. List of peaks, relative intensities, and molecular formulae at 467 m/z in the 7 T ESI FT-ICR mass spectra of SRFA (highlighted: Peaks A – E in Figure 3.4).

| m/z | relative intensity | | | | molecular formula | | | | error (ppm) |
|-----------|--------------------|------|------|-------|-------------------|-----------------|----|----|-------------|
| | UNT-1 | BHR | BDR | UNT-2 | ¹² C | ¹³ C | H | O | |
| 467.01035 | 11.4 | 0.0 | 0.0 | 10.3 | 18 | 0 | 12 | 15 | 0.004 |
| 467.01791 | 7.6 | 0.0 | 0.0 | 7.3 | - | - | - | - | - |
| 467.02557 | 10.6 | 0.0 | 0.0 | 11.9 | 22 | 0 | 12 | 12 | 0.065 |
| 467.03619 | 13.7 | 10.1 | 0.0 | 14.1 | - | - | - | - | - |
| 467.04675 | 52.0 | 48.7 | 36.1 | 59.3 | 19 | 0 | 16 | 14 | 0.026 |
| 467.05436 | 14.1 | 0.0 | 0.0 | 11.1 | - | - | - | - | - |
| 467.06196 | 17.2 | 0.0 | 0.0 | 13.7 | 23 | 0 | 16 | 11 | 0.055 |
| 467.08004 | 0.0 | 0.0 | 40.5 | 0.0 | - | - | - | - | - |
| 467.08067 | 6.0 | 0.0 | 0.0 | 0.0 | - | - | - | - | - |
| 467.08311 | 57.1 | 80.9 | 41.0 | 64.7 | 20 | 0 | 20 | 13 | 0.015 |
| 467.09073 | 8.9 | 0.0 | 0.0 | 0.0 | - | - | - | - | - |
| 467.09831 | 8.9 | 0.0 | 0.0 | 11.1 | 24 | 0 | 20 | 10 | 0.137 |
| 467.11646 | 0.0 | 0.0 | 38.8 | 0.0 | - | - | - | - | - |
| 467.11949 | 80.4 | 82.1 | 44.9 | 82.4 | 21 | 0 | 24 | 12 | 0.029 |
| 467.13474 | 6.9 | 0.0 | 0.0 | 8.0 | 25 | 0 | 24 | 9 | 0.035 |
| 467.14984 | 0.0 | 0.0 | 19.1 | 0.0 | 29 | 0 | 24 | 6 | 0.378 |
| 467.15275 | 0.0 | 0.0 | 30.7 | 0.0 | - | - | - | - | - |
| 467.15588 | 98.3 | 97.9 | 57.5 | 96.4 | 22 | 0 | 28 | 11 | 0.030 |
| 467.17115 | 0.0 | 0.0 | 0.0 | 7.7 | 26 | 0 | 28 | 8 | 0.006 |
| 467.18630 | 0.0 | 0.0 | 18.1 | 0.0 | 30 | 0 | 28 | 5 | 0.212 |
| 467.18919 | 0.0 | 0.0 | 38.3 | 0.0 | - | - | - | - | - |
| 467.19227 | 45.2 | 79.1 | 28.0 | 41.9 | 23 | 0 | 32 | 10 | 0.014 |
| 467.22557 | 0.0 | 0.0 | 28.2 | 0.0 | - | - | - | - | - |
| 467.22865 | 10.2 | 39.5 | 0.0 | 9.2 | 24 | 0 | 36 | 9 | 0.027 |
| 467.26501 | 0.0 | 9.6 | 0.0 | 0.0 | 25 | 0 | 40 | 8 | 0.067 |

Table A3.2. List of peaks, relative intensities, and molecular formulae at 469 *m/z* in the 7 T ESI FT-ICR mass spectra of SRFA (highlighted: Peaks A' – E' in Figure 3.4).

| <i>m/z</i> | relative intensity | | | | ¹² C | ¹³ C | molecular formula | | error (ppm) |
|------------|--------------------|------|------|-------|-----------------|-----------------|-------------------|----|-------------|
| | UNT-1 | BHR | BDR | UNT-2 | | | H | O | |
| 469.00484 | 6.5 | 0.0 | 0.0 | 10.1 | 21 | 0 | 10 | 13 | 0.054 |
| 469.02602 | 20.2 | 15.7 | 0.0 | 22.3 | 18 | 0 | 14 | 15 | 0.044 |
| 469.03366 | 8.7 | 0.0 | 0.0 | 10.5 | - | - | - | - | - |
| 469.04126 | 23.3 | 12.3 | 0.0 | 28.7 | 22 | 0 | 14 | 12 | 0.005 |
| 469.05176 | 15.3 | 11.4 | 0.0 | 8.9 | - | - | - | - | - |
| 469.06239 | 49.5 | 52.8 | 29.9 | 54.2 | 19 | 0 | 18 | 14 | 0.021 |
| 469.07006 | 7.9 | 11.9 | 0.0 | 7.4 | - | - | - | - | - |
| 469.07765 | 21.0 | 11.6 | 0.0 | 22.3 | 23 | 0 | 18 | 11 | 0.015 |
| 469.08820 | 0.0 | 10.3 | 0.0 | 0.0 | - | - | - | - | - |
| 469.09572 | 0.0 | 0.0 | 42.6 | 0.0 | - | - | - | - | - |
| 469.09877 | 59.8 | 83.2 | 33.2 | 55.9 | 20 | 0 | 22 | 13 | 0.007 |
| 469.11404 | 15.8 | 9.4 | 0.0 | 18.3 | 24 | 0 | 22 | 10 | 0.028 |
| 469.12914 | 0.0 | 0.0 | 24.3 | 0.0 | 28 | 0 | 22 | 7 | 0.303 |
| 469.13207 | 0.0 | 0.0 | 37.1 | 0.0 | - | - | - | - | - |
| 469.13517 | 73.7 | 90.8 | 51.8 | 76.9 | 21 | 0 | 26 | 12 | 0.034 |
| 469.15042 | 15.0 | 0.0 | 0.0 | 13.2 | 25 | 0 | 26 | 9 | 0.030 |
| 469.16551 | 0.0 | 0.0 | 19.0 | 0.0 | 29 | 0 | 26 | 6 | 0.317 |
| 469.16846 | 0.0 | 0.0 | 45.3 | 0.0 | - | - | - | - | - |
| 469.17156 | 62.0 | 99.9 | 33.6 | 54.1 | 22 | 0 | 30 | 11 | 0.040 |
| 469.18680 | 12.1 | 0.0 | 0.0 | 8.9 | 26 | 0 | 30 | 8 | 0.006 |
| 469.20188 | 0.0 | 0.0 | 20.5 | 0.0 | 30 | 0 | 30 | 5 | 0.365 |
| 469.20489 | 0.0 | 0.0 | 49.8 | 0.0 | - | - | - | - | - |
| 469.20794 | 15.4 | 71.7 | 0.0 | 13.9 | 23 | 0 | 34 | 10 | 0.042 |
| 469.24123 | 0.0 | 0.0 | 27.2 | 0.0 | - | - | - | - | - |
| 469.24430 | 0.0 | 21.7 | 0.0 | 0.0 | 24 | 0 | 38 | 9 | 0.009 |
| 469.55345 | 9.4 | 0.0 | 0.0 | 0.0 | - | - | - | - | - |

Table A3.3. List of peaks, relative intensities, and molecular formulae at 470 *m/z* in the 7 T ESI FT-ICR mass spectra of SRFA (highlighted: Peaks A” – E” in Figure 3.4).

| <i>m/z</i> | relative intensity | | | | ¹² C | ¹³ C | molecular formula | | error (ppm) |
|------------|--------------------|------|------|-------|-----------------|-----------------|-------------------|----|-------------|
| | UNT-1 | BHR | BDR | UNT-2 | | | H | O | |
| 470.02327 | 7.7 | 0.0 | 0.0 | 0.0 | - | - | - | - | - |
| 470.03083 | 12.4 | 0.0 | 0.0 | 7.1 | - | - | - | - | - |
| 470.04141 | 13.2 | 0.0 | 0.0 | 10.0 | - | - | - | - | - |
| 470.04907 | 7.9 | 0.0 | 0.0 | 0.0 | - | - | - | - | - |
| 470.05963 | 13.3 | 14.8 | 0.0 | 11.6 | - | - | - | - | - |
| 470.06572 | 11.9 | 12.3 | 0.0 | 13.3 | 18 | 1 | 18 | 14 | 0.063 |
| 470.06875 | 0.0 | 0.0 | 45.5 | 0.0 | - | - | - | - | - |
| 470.07781 | 6.8 | 15.3 | 0.0 | 8.8 | - | - | - | - | - |
| 470.09601 | 9.2 | 0.0 | 0.0 | 7.7 | - | - | - | - | - |
| 470.10208 | 14.6 | 15.0 | 25.6 | 13.4 | 19 | 1 | 22 | 13 | 0.098 |
| 470.10507 | 0.0 | 0.0 | 41.3 | 0.0 | - | - | - | - | - |
| 470.13849 | 19.8 | 18.1 | 30.8 | 17.0 | 20 | 1 | 26 | 12 | 0.070 |
| 470.14145 | 0.0 | 0.0 | 62.9 | 0.0 | - | - | - | - | - |
| 470.17488 | 17.4 | 20.2 | 28.3 | 12.8 | 21 | 1 | 30 | 11 | 0.076 |
| 470.17781 | 0.0 | 0.0 | 71.1 | 0.0 | - | - | - | - | - |
| 470.21123 | 0.0 | 13.5 | 27.8 | 0.0 | 22 | 1 | 34 | 10 | 0.144 |
| 470.21420 | 0.0 | 0.0 | 35.6 | 0.0 | - | - | - | - | - |
| 470.24759 | 0.0 | 0.0 | 18.2 | 0.0 | 23 | 1 | 38 | 9 | 0.142 |
| 470.57952 | 0.0 | 10.4 | 0.0 | 0.0 | - | - | - | - | - |

Appendix 4: Reduction and mass spectrometric analysis of acetovanillone

Two 5.0 mL aliquots of 84 μM acetovanillone (adjusted to pH 10 w/ NaOH) were reduced by adding 50 μL of 10 mg/mL NaBH_4 or NaBD_4 solution (in 1.25 mM NaOH). Solutions shaken and stored in loosely capped vials over 3 days. Absorbance spectra were stable after this period. Aqueous solutions were filtered using a 0.2 μm filter and diluted with equal volumes of LC-grade methanol (for a final solvent composition of 50:50 water/methanol) immediately prior to analysis by a Thermo ESI LTQ Orbitrap XL. Mass spectra were recorded from 100 – 250 m/z in negative ion mode with 100 co-added scans to achieve an average resolution of 52,000 (FWHM), with all other experimental parameters similar to those used for the analysis of DOM samples. These spectra (Figure A4.1) revealed that the total ion counts of the two reduced samples were only slightly less than that of the untreated sample, and confirms the expected appearance of $M+2.0157$ and $M+3.0219$ mass shifts arising from borohydride and borodeuteride reduction, respectively. The $[\text{M-H}]^-$ peak of the original species remained in both reduced samples, which is indicative of incomplete reduction, and the only other significant peaks were at 150 m/z in all three spectra (possible organic impurity and/or fragment ion), and 135 m/z (possibly $(\text{CH}_3)_4\text{B}^-$) in the two reduced samples, all of which were substantially less than the $[\text{M-H}]^-$ ions. Therefore, contamination and large changes in ionization efficiency resulting from reduction are not expected to interfere significantly with the qualitative identification of ketones/aldehydes.

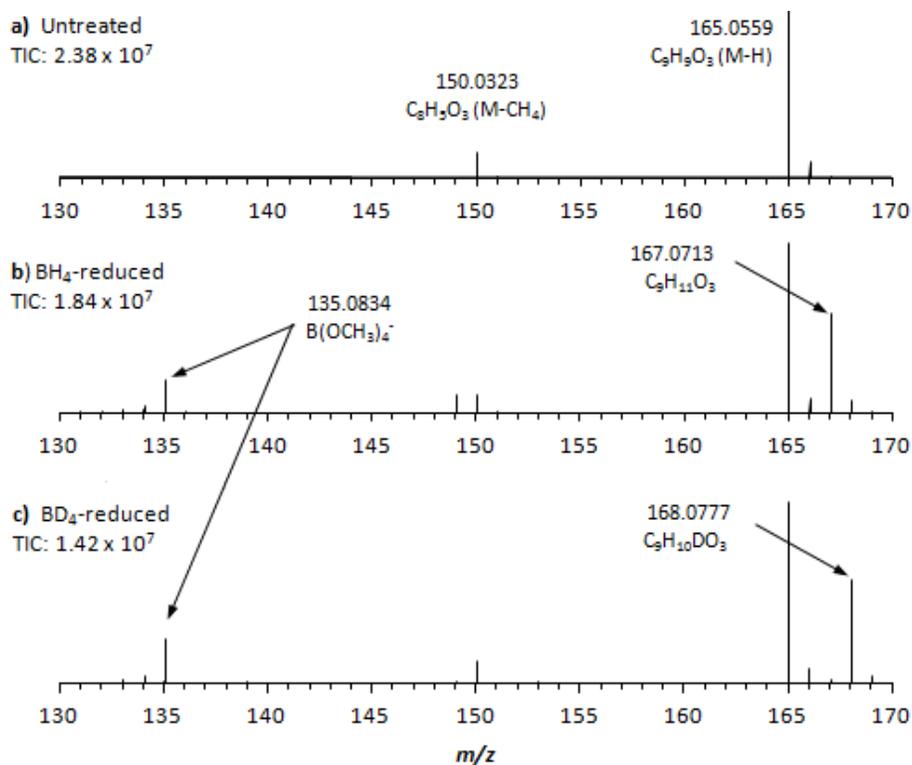


Figure A4.1. Negative ion ESI-orbitrap mass spectra of 42 μ M acetovanillone in 50:50 methanol/water: (a) untreated, (b) borohydride-reduced, and (c) borodeuteride-reduced aliquots. TIC = total ion count. Masses of molecular formula assignments are within 4 ppm of the measured masses.

Appendix 5: Molecular formula assignment details for Chapter 4

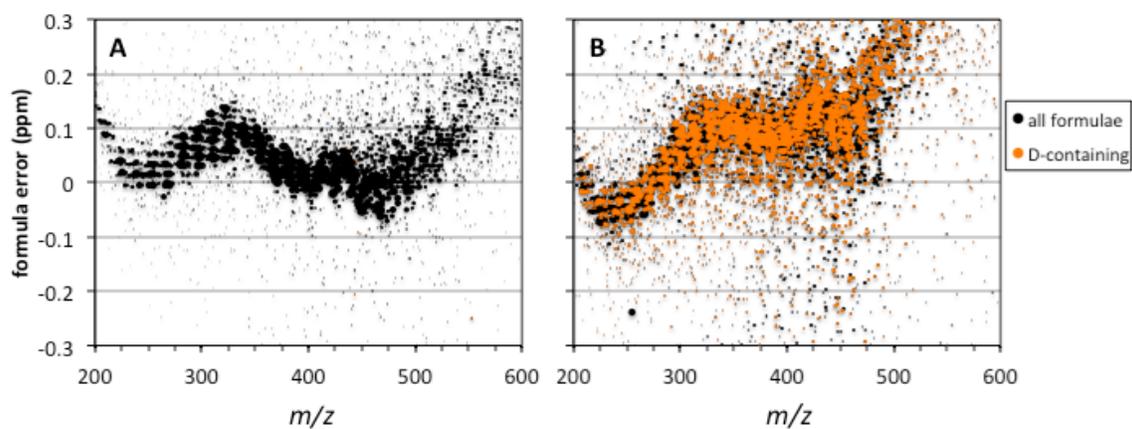


Figure A5.1. Molecular formula error vs. m/z for (A) UNT and (B) BDR-3 calculated using a 0.3 ppm error limit using function **F2** (mass shift-based assignment of D-containing compounds).

Table A5.1. Molecular formula assignment details for UNT in the 200 to 400 m/z and 400 to 500 m/z ranges using 0.2 and 0.3 ppm error limits, respectively and various maximum allowances of N and S.

| FE_{max} | 0.2 ppm | | | | | 0.3 ppm | | | | |
|-----------------------------------------|----------------|---------------|---------------|---------------|--------------|----------------|---------------|---------------|---------------|--------------|
| N_{max}, S_{max} | (0, 0) | (1, 1) | (2, 2) | (3, 3) | (4,2) | (0, 0) | (1, 1) | (2, 2) | (3, 3) | (4,2) |
| # peaks | 2761 | 3674 | 3733 | 4000 | 3775 | 3029 | 4059 | 4157 | 4323 | 4218 |
| % peaks | 40.9 | 54.4 | 55.3 | 59.2 | 55.9 | 44.9 | 60.1 | 61.6 | 64.0 | 62.5 |
| % TIC | 81.9 | 86.2 | 86.4 | 88.3 | 86.6 | 83.1 | 88.5 | 88.9 | 89.6 | 89.2 |
| % C _{mass} ^a | 55.4 | 55.3 | 55.3 | 54.9 | 55.3 | 55.4 | 55.3 | 55.4 | 55.3 | 55.3 |
| % H _{mass} ^a | 5.05 | 5.01 | 5.01 | 5.03 | 5.01 | 5.03 | 4.99 | 4.98 | 4.99 | 4.98 |
| % O _{mass} ^a | 39.6 | 39.4 | 39.4 | 39.1 | 39.3 | 39.6 | 39.4 | 39.3 | 39.1 | 39.3 |
| % N _{mass} ^a | 0 | 0.17 | 0.19 | 0.34 | 0.21 | 0 | 0.18 | 0.21 | 0.27 | 0.24 |
| % S _{mass} ^a | 0 | 0.048 | 0.11 | 0.62 | 0.13 | 0 | 0.10 | 0.16 | 0.33 | 0.18 |
| O/C _{avg} ^a | 0.543 | 0.542 | 0.542 | 0.542 | 0.542 | 0.544 | 0.543 | 0.541 | 0.539 | 0.542 |
| H/C _{avg} ^a | 1.086 | 1.081 | 1.081 | 1.092 | 1.081 | 1.083 | 1.078 | 1.077 | 1.078 | 1.078 |
| AI _{avg} ^a | 0.074 | 0.079 | 0.082 | 0.125 | 0.084 | 0.074 | 0.080 | 0.086 | 0.088 | 0.089 |
| DBE _{avg} ^a | 9.42 | 9.48 | 9.51 | 9.43 | 9.503 | 9.52 | 9.59 | 9.64 | 9.64 | 9.64 |
| FE ^{b,c} | 0.082 | 0.085 | 0.086 | 0.089 | 0.086 | 0.105 | 0.110 | 0.112 | 0.114 | 0.113 |

^a relative intensity weighted; ^b root mean square error; ^c error in ppm.

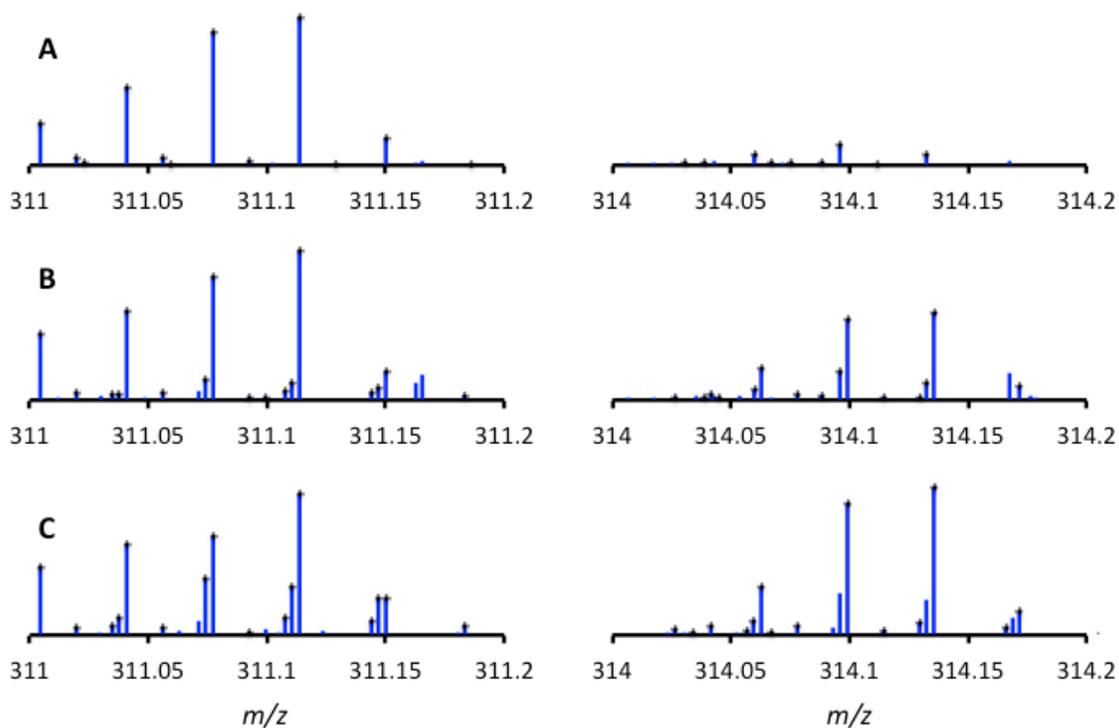


Figure A5.2. Expanded regions of the negative ion 12T ESI FT-ICR mass spectra of (A) UNT, (B) BDR-1, and (C) BDR-4 SRFA at 311 (left) and 314 (right) m/z . Black markers on peaks indicate those which have been assigned a molecular formula.

**Table A5.2. Molecular formula for peaks at 311 and 314 m/z in SRFA UNT.
 Intensities are normalized to the most intense peak the 295 to 314 m/z range.**

| m/z | intensity | ^{12}C | ^{13}C | H | D | O | N | ^{32}S | FE (ppm) |
|-----------|-----------|-----------------|-----------------|----|---|----|---|-----------------|----------|
| 311.00449 | 27.0 | 12 | 0 | 8 | 0 | 10 | 0 | 0 | 0.065 |
| 311.01975 | 4.4 | 16 | 0 | 8 | 0 | 7 | 0 | 0 | 0.078 |
| 311.02315 | 1.2 | 13 | 0 | 12 | 0 | 7 | 0 | 1 | 0.170 |
| 311.04089 | 50.3 | 13 | 0 | 12 | 0 | 9 | 0 | 0 | 0.112 |
| 311.05614 | 4.7 | 17 | 0 | 12 | 0 | 6 | 0 | 0 | 0.092 |
| 311.05952 | 0.9 | 14 | 0 | 16 | 0 | 6 | 0 | 1 | 0.121 |
| 311.07727 | 86.3 | 14 | 0 | 16 | 0 | 8 | 0 | 0 | 0.095 |
| 311.09255 | 2.4 | 18 | 0 | 16 | 0 | 5 | 0 | 0 | 0.171 |
| 311.11364 | 96.1 | 15 | 0 | 20 | 0 | 7 | 0 | 0 | 0.045 |
| 311.12887 | 1.0 | 19 | 0 | 20 | 0 | 4 | 0 | 0 | -0.038 |
| 311.15004 | 17.4 | 16 | 0 | 24 | 0 | 6 | 0 | 0 | 0.092 |
| 311.18636 | 0.9 | 17 | 0 | 28 | 0 | 5 | 0 | 0 | -0.117 |
| 314.03062 | 1.3 | 15 | 0 | 9 | 0 | 7 | 1 | 0 | -0.015 |
| 314.03874 | 1.3 | 15 | 1 | 10 | 0 | 7 | 0 | 0 | 0.030 |
| 314.05990 | 7.5 | 12 | 1 | 14 | 0 | 9 | 0 | 0 | 0.128 |
| 314.06706 | 1.8 | 16 | 0 | 13 | 0 | 6 | 1 | 0 | 0.158 |
| 314.07516 | 1.3 | 16 | 1 | 14 | 0 | 6 | 0 | 0 | 0.140 |
| 314.08812 | 1.8 | 13 | 0 | 17 | 0 | 8 | 1 | 0 | -0.062 |
| 314.09626 | 13.6 | 13 | 1 | 18 | 0 | 8 | 0 | 0 | 0.047 |
| 314.11154 | 0.8 | 17 | 1 | 18 | 0 | 5 | 0 | 0 | 0.123 |
| 314.13266 | 7.5 | 14 | 1 | 22 | 0 | 7 | 0 | 0 | 0.094 |

Table A5.3. Molecular formula for peaks at 311 and 314 m/z in SRFA BDR-1. Intensities are normalized to the most intense peak the 295 to 314 m/z range.

| m/z | intensity | ^{12}C | ^{13}C | H | D | O | N | ^{32}S | FE (ppm) |
|-----------|-----------|-----------------|-----------------|----|---|----|---|-----------------|----------|
| 311.00449 | 38.3 | 12 | 0 | 8 | 0 | 10 | 0 | 0 | 0.065 |
| 311.01976 | 4.4 | 16 | 0 | 8 | 0 | 7 | 0 | 0 | 0.110 |
| 311.03486 | 3.3 | 12 | 1 | 9 | 1 | 9 | 0 | 0 | 0.075 |
| 311.03782 | 3.1 | 13 | 0 | 8 | 2 | 9 | 0 | 0 | 0.197 |
| 311.04088 | 51.8 | 13 | 0 | 12 | 0 | 9 | 0 | 0 | 0.080 |
| 311.05612 | 4.6 | 17 | 0 | 12 | 0 | 6 | 0 | 0 | 0.028 |
| 311.07419 | 12.1 | 14 | 0 | 12 | 2 | 8 | 0 | 0 | 0.148 |
| 311.07726 | 72.2 | 14 | 0 | 16 | 0 | 8 | 0 | 0 | 0.063 |
| 311.09248 | 1.9 | 18 | 0 | 16 | 0 | 5 | 0 | 0 | -0.053 |
| 311.09952 | 1.8 | 14 | 0 | 16 | 1 | 7 | 1 | 0 | 0.059 |
| 311.10762 | 5.5 | 14 | 1 | 17 | 1 | 7 | 0 | 0 | 0.041 |
| 311.11056 | 10.0 | 15 | 0 | 16 | 2 | 7 | 0 | 0 | 0.098 |
| 311.11364 | 86.7 | 15 | 0 | 20 | 0 | 7 | 0 | 0 | 0.045 |
| 311.14399 | 4.9 | 15 | 1 | 21 | 1 | 6 | 0 | 0 | -0.009 |
| 311.14694 | 7.7 | 16 | 0 | 20 | 2 | 6 | 0 | 0 | 0.081 |
| 311.15001 | 17.1 | 16 | 0 | 24 | 0 | 6 | 0 | 0 | -0.004 |
| 314.02640 | 1.9 | 12 | 0 | 9 | 1 | 10 | 0 | 0 | 0.011 |
| 314.03874 | 1.5 | 15 | 1 | 10 | 0 | 7 | 0 | 0 | 0.030 |
| 314.04165 | 3.5 | 16 | 0 | 9 | 1 | 7 | 0 | 0 | -0.008 |
| 314.04501 | 1.6 | 13 | 0 | 13 | 1 | 7 | 0 | 1 | -0.043 |
| 314.05989 | 6.7 | 12 | 1 | 14 | 0 | 9 | 0 | 0 | 0.096 |
| 314.06281 | 18.5 | 13 | 0 | 13 | 1 | 9 | 0 | 0 | 0.090 |
| 314.07804 | 3.5 | 17 | 0 | 13 | 1 | 6 | 0 | 0 | 0.006 |
| 314.08810 | 2.1 | 13 | 0 | 17 | 0 | 8 | 1 | 0 | -0.125 |
| 314.09624 | 16.7 | 13 | 1 | 18 | 0 | 8 | 0 | 0 | 0.017 |
| 314.09918 | 47.7 | 14 | 0 | 17 | 1 | 8 | 0 | 0 | 0.041 |
| 314.11438 | 1.6 | 18 | 0 | 17 | 1 | 5 | 0 | 0 | -0.138 |
| 314.12956 | 1.7 | 14 | 1 | 18 | 2 | 7 | 0 | 0 | 0.082 |
| 314.13264 | 10.1 | 14 | 1 | 22 | 0 | 7 | 0 | 0 | 0.030 |
| 314.13557 | 51.5 | 15 | 0 | 21 | 1 | 7 | 0 | 0 | 0.055 |
| 314.17196 | 7.8 | 16 | 0 | 25 | 1 | 6 | 0 | 0 | 0.070 |

Table A5.4. Molecular formula for peaks at 311 and 314 m/z in SRFA BDR-4. Intensities are normalized to the most intense peak the 295 to 314 m/z range.

| m/z | intensity | ^{12}C | ^{13}C | H | D | O | N | ^{32}S | FE (ppm) |
|-----------|-----------|-----------------|-----------------|----|---|----|---|-----------------|----------|
| 311.00448 | 39.6 | 12 | 0 | 8 | 0 | 10 | 0 | 0 | 0.033 |
| 311.01975 | 4.5 | 16 | 0 | 8 | 0 | 7 | 0 | 0 | 0.078 |
| 311.03485 | 5.7 | 12 | 1 | 9 | 1 | 9 | 0 | 0 | 0.043 |
| 311.03779 | 9.9 | 13 | 0 | 8 | 2 | 9 | 0 | 0 | 0.101 |
| 311.04088 | 53.1 | 13 | 0 | 12 | 0 | 9 | 0 | 0 | 0.080 |
| 311.05611 | 4.3 | 17 | 0 | 12 | 0 | 6 | 0 | 0 | -0.004 |
| 311.07417 | 32.6 | 14 | 0 | 12 | 2 | 8 | 0 | 0 | 0.084 |
| 311.07726 | 58.1 | 14 | 0 | 16 | 0 | 8 | 0 | 0 | 0.063 |
| 311.09252 | 1.8 | 18 | 0 | 16 | 0 | 5 | 0 | 0 | 0.075 |
| 311.10758 | 10.1 | 14 | 1 | 17 | 1 | 7 | 0 | 0 | -0.088 |
| 311.11055 | 27.8 | 15 | 0 | 16 | 2 | 7 | 0 | 0 | 0.066 |
| 311.11364 | 82.0 | 15 | 0 | 20 | 0 | 7 | 0 | 0 | 0.045 |
| 311.14399 | 8.0 | 15 | 1 | 21 | 1 | 6 | 0 | 0 | -0.009 |
| 311.14696 | 21.3 | 16 | 0 | 20 | 2 | 6 | 0 | 0 | 0.145 |
| 311.15004 | 21.2 | 16 | 0 | 24 | 0 | 6 | 0 | 0 | 0.092 |
| 314.02637 | 3.5 | 12 | 0 | 9 | 1 | 10 | 0 | 0 | -0.084 |
| 314.03405 | 1.8 | 12 | 0 | 13 | 0 | 7 | 1 | 1 | 0.172 |
| 314.04170 | 5.7 | 16 | 0 | 9 | 1 | 7 | 0 | 0 | 0.151 |
| 314.05676 | 2.1 | 12 | 1 | 10 | 2 | 9 | 0 | 0 | -0.011 |
| 314.05982 | 8.5 | 12 | 1 | 14 | 0 | 9 | 0 | 0 | 0.127 |
| 314.06280 | 28.4 | 13 | 0 | 13 | 1 | 9 | 0 | 0 | 0.058 |
| 314.06705 | 1.7 | 16 | 0 | 13 | 0 | 6 | 1 | 0 | 0.126 |
| 314.07804 | 5.2 | 17 | 0 | 13 | 1 | 6 | 0 | 0 | 0.006 |
| 314.09918 | 77.2 | 14 | 0 | 17 | 1 | 8 | 0 | 0 | 0.041 |
| 314.11443 | 2.4 | 18 | 0 | 17 | 1 | 5 | 0 | 0 | 0.021 |
| 314.12954 | 7.6 | 14 | 1 | 18 | 2 | 7 | 0 | 0 | 0.018 |
| 314.13557 | 86.4 | 15 | 0 | 21 | 1 | 7 | 0 | 0 | 0.055 |
| 314.16592 | 4.6 | 15 | 1 | 22 | 2 | 6 | 0 | 0 | 0.001 |
| 314.17197 | 13.8 | 16 | 0 | 25 | 1 | 6 | 0 | 0 | 0.102 |

Appendix 6: Supporting optical data for Chapter 4

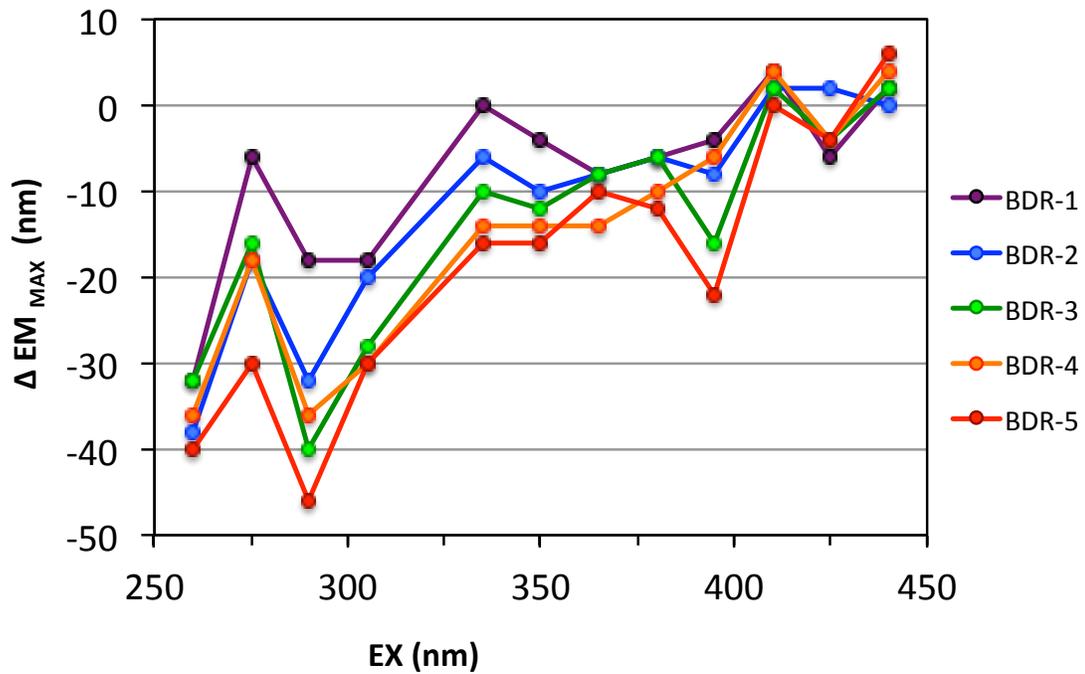


Figure A6.1. Shifts in the wavelength of maximum emission (EM_{MAX}) for various excitation wavelengths (EX) in BDR-1 through BDR-5 relative to those in UNT. $\Delta EM_{MAX} = EM_{MAX} (UNT) - EM_{MAX} (BDR)$

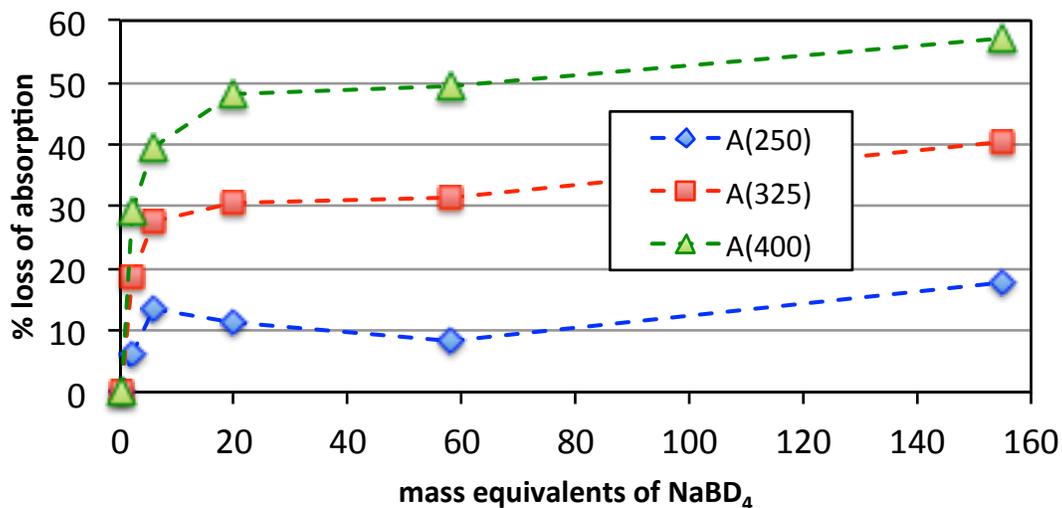


Figure A6.2. Percent loss of absorption (relative to UNT) of all samples at selected wavelengths versus mass equivalents of NaBD₄.

Comparison of PPL extracted and whole untreated and reduced SRFA via UV-visible spectrophotometry. The UV-visible absorption spectra of solid phase extracted and “whole” (pre-extraction) solutions of Suwannee River fulvic acid (SRFA) were compared to one another. This was done to test for any wavelength-dependent biases in the comparison of absorption spectra for untreated and reduced SRFA that may result from extraction via the functionalized polystyrene divinyl benzene (PPL) cartridges that were employed for purifying reduced samples.

A 0.10 mg/mL SRFA stock of was prepared by dissolving 4 mg SRFA in 40 mL Milli-Q water and adjusting the pH to 10.3 with 10 μ L 4 M sodium hydroxide. 20 mL of this stock was transferred to a 30 mL glass vial and 34 mg of solid sodium borohydride (NaBH₄) was added to this vial to a final concentration of 1.7 mg/mL. The vial was loosely sealed with a Teflon-lined cap, gently shaken, and placed in the dark at room for

2 h. After the reduction period, 4 M hydrochloric acid (HCl) was added incrementally to the untreated and reduced aliquots (70 μ L total) to adjust the pH to 2.0 to decompose any residual NaBH_4 .

9.0 mL of each aliquot was loaded onto a new PPL cartridge (preconditioned with 2 mL methanol (MeOH) and 2 mL of 10 mM HCl), and allowed to elute by gravity. The cartridges were rinsed with an additional 2 mL of 10 mM HCl and dried using a steady flow of nitrogen (~10 min). The SRFA was eluted into a glass vial using 2.0 mL MeOH, which was then evaporated under nitrogen (~10 mins). The dried solid was then re-dissolved in 36.0 mL 24 mM potassium phosphate buffer. The remaining untreated and reduced aliquots were then diluted 1:4 to in 32 mM phosphate buffer (final phosphate buffer concentration was 24 mM). All four aliquots – untreated whole (UNT), untreated and extracted (UNT PPL), reduced whole (RED), and reduced and extracted (RED PPL) – were filtered using 0.2 μ m polyethersulfone membrane filters. Absorption spectra were recorded in a 1 cm cuvette versus air from 200 – 700 nm. The spectrum of a 24 mM phosphate buffer blank was manually subtracted from all spectra for analysis.

The absorbance of the untreated PPL extract was slightly lower than that of the whole SRFA sample across all wavelengths, with a gradually increasing relative loss at longer wavelengths. Conversely, the absorbance of the borohydride-reduced PPL extract was slightly higher than the corresponding whole sample. This reverse trend is surprising, since a slight loss of absorption would be expected following the extraction, since some (aliphatic) compounds within SRFA may not adsorb efficiently to the PPL solid phase during extraction, resulting in a loss of absorption. However, the apparent differences between the two extracted and whole aliquots are much smaller than the differences

between the reduced and untreated sample, thus suggesting that these optical differences may be negligible. Furthermore, the lack of a drastic wavelength dependence in the loss or gains of absorption following extraction, along with the fact that the absorption of the reduced sample was slightly *higher* than that of the whole sample suggests that the differences observed here may be due to dilution error, rather than compositional differences which would affect the observed optical properties.

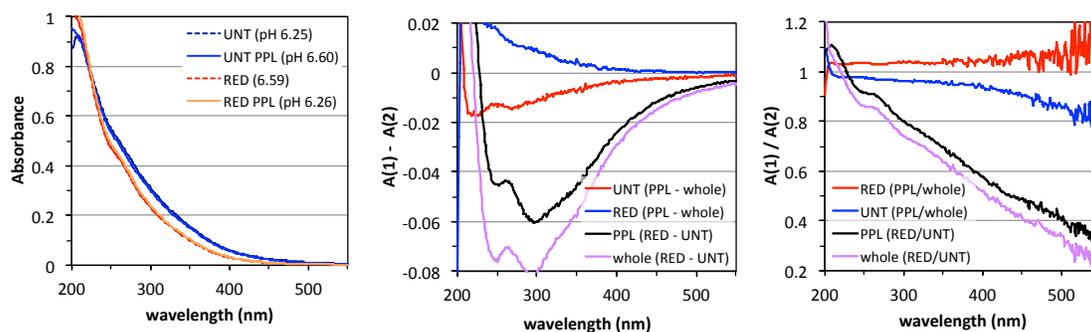


Figure A6.3. Absorption spectra (left) of whole and PPL extracted SRFA (untreated and reduced) and difference (middle) and fractional (right) spectra comparing the ratios of the four samples.

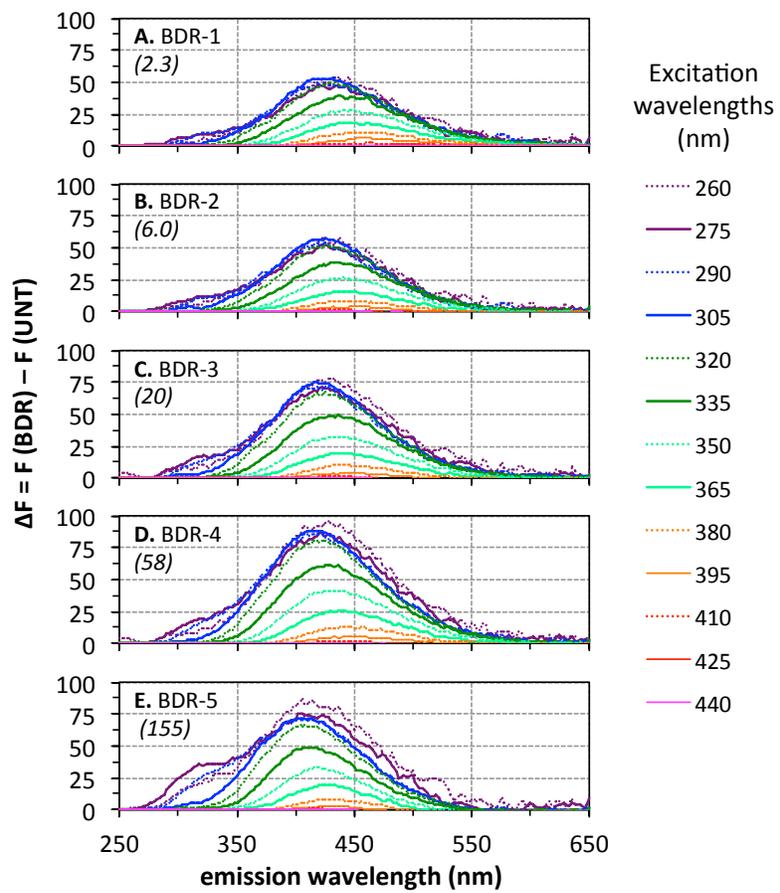


Figure A6.4. Difference emission spectra of BDR SRFA samples relative to UNT SRFA.

Appendix 7: Sample collection details for C₁₈ extracts

Extraction of Delaware River/Bay samples were collected and performed by Boyle et al.³² onboard the research vessel R/V Cape Henlopen. The individual water samples, collected at locations given in Table A5.1 were filtered through 0.2 µm Gellman filters, acidified to pH 2, and pumped through a C₁₈ solid phase extraction (SPE) cartridge with a flow rate of 50 mL/min, which were pretreated with 100 mL high purity methanol and 50 mL of acidified water. Cartridges were rinsed with 1 L of acidified (pH 2) water and stored in a refrigerator until further processing. Aqueous solutions of the extracted DOM were prepared by eluting the organic material from the cartridge with high purity methanol, evaporating to dryness with a rotary evaporator, re-dissolved in a few milliliters of deionized water (from a Mill-Q purification system), and adjusted to pH 7 with aqueous sodium hydroxide (NaOH) solutions. A similar procedure was used for the collection of Equatorial Atlantic samples, as described and performed by Andrew et al.³ onboard the research vessel R/V Endeavor.

Table A7.1. Sample collection date and location for all C₁₈ extracts.^a

| Sample | Date^a | Latitude | Longitude | Depth (m) |
|---------------|-------------------------|-----------------|------------------|------------------|
| DELB | 10/13/2006 | 38.992 | -75.132 | 2 |
| DERV | 10/15/2006 | 40.139 | -74.735 | 2 |
| DEUB | 12/--/2006 | 39.633 | -74.574 | 2 |
| EACR | 5/--/2009 | -6.0038 | 4.9928 | 4 |
| EAUW | 5/--/2009 | -0.7 | 0.0 | 40 |
| EADO | 5/--/2009 | 3.0020 | -22.9992 | 1000 |
| EASO | 5/--/2009 | 3.0020 | -22.9992 | 4 |

^a "--" indicates unknown value; ^b date given in month/day/year format.

Appendix 8: SRFA and PLFA at 311 m/z (for Chapter 5)

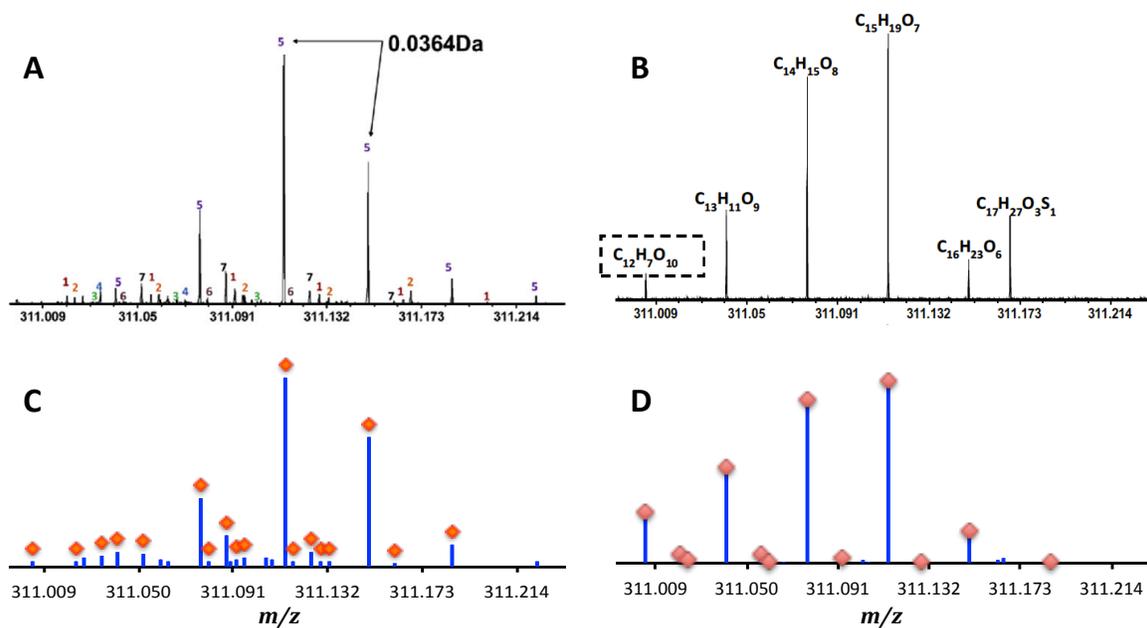


Figure A8.1. Expanded region of the negative ion ESI FT-ICR mass spectra of PLFA (A & C) and SRFA (B & D) as obtained by D'Andrilli et al. (A & B) and in this study (C & D). Red diamonds indicate peaks to which a molecular formula was assigned.

Appendix 9: Custom MATLAB functions and description of calculations

A9.1. Chapter 3 analysis code

```
=====
function [ PData, PData_T, Spec, Form, V_U2, V_HR, V_DR, Shifts, Shift_T,
ID_kets ] = D_ch3( U1, HR, DR, U2 )
%performs initial borodeuteride mass labeling test calculations

% Input Mass spectrum
% 12 SRFA U1 (full)
% 13 SRFA U2 (full)
% 18 SRFA U3 (full)
% 46 SRFA UpC (full)
% 6 SRFA H (full)
% 1 SRFA D (full)
% 24 LB U (full)
% 19 LB H (full)
% 47 LB D (full)
% 39 UB U (full)
% 34 UB H (full)
% 29 UB D (full)

% Input numbers correspond to column numbers in the variable 'IntensityNeg'
% in workspace 'BloughData_neg.08.31.2011.mat'

%Creates matrix 'PData' for the specified spectra in
%BloughData_neg.08.31.2011.mat, with the following columns:
%
% mz RI(U1) RI(HR) RI(DR) RI(U2)
% 1 2 3 4 5
%
load BloughData_neg.08.31.2011.mat
PData = zeros(1,5);
IntensityNeg(isnan(IntensityNeg))=0;
pz = 0;
for i = 1 : size(IntensityNeg,1)
    if (IntensityNeg(i, U1)>0 || IntensityNeg(i, HR)>0 || IntensityNeg(i, DR)>0
|| IntensityNeg(i, U2)>0 ) && PeaksNeg(i,1)<601;
        pz = pz + 1;
        PData(pz, 1) = PeaksNeg(i,1);
        PData(pz, 2) = IntensityNeg (i, U1);
        PData(pz, 3) = IntensityNeg (i, HR);
        PData(pz, 4) = IntensityNeg (i, DR);
        PData(pz, 5) = IntensityNeg (i, U2);
    end
end

%Creates table 'PData_I', which stores number of peaks and TIC of each
%spectrum
PData_I = zeros(2,4);
```

```

for i = 1 : pz
    for n = 1 : 4
        if PData (i, n+1) > 0
            PData_I(1,n) = PData_I(1,n) + 1;
            PData_I(2,n) = PData_I(2,n) + PData(i,n+1);
        end
    end
end
Names = {'peaks', 'TIC '};
PData_T = table(PData_I, 'RowNames', Names);

%Normalizes intensities in 'PData'
for n = 1 : 4
    PData(:,n+1) = ( 100 / max(PData(:,n+1)) ) .* PData(:,n+1);
end

%Creates spectral plot data in 'Spectra', with the following columns:
%
% U1(all) HR(+U1) HR(new) DR(+U1) DR(new)
% mz RI mz RI mz RI mz RI mz RI
% 1 2 3 4 5 6 7 8 9 10
%
Spec = zeros(1,9);
SZ = zeros(1,5);
for i = 1 : pz
    if PData(i,2) > 0
        SZ(1) = SZ(1) + 1;
        Spec(SZ(1), 1) = PData(i,1);
        Spec(SZ(1), 2) = PData(i,2);
    end
    if PData(i,3) > 0
        if PData(i,2) > 0
            SZ(2) = SZ(2) + 1;
            Spec(SZ(2), 3) = PData(i,1);
            Spec(SZ(2), 4) = PData(i,3);
        else
            SZ(3) = SZ(3) + 1;
            Spec(SZ(3), 5) = PData(i,1);
            Spec(SZ(3), 6) = PData(i,3);
        end
    end
    if PData(i,4) > 0
        if PData(i,2) > 0
            SZ(4) = SZ(4) + 1;
            Spec(SZ(4), 7) = PData(i,1);
            Spec(SZ(4), 8) = PData(i,4);
        else
            SZ(5) = SZ(5) + 1;
            Spec(SZ(5), 9) = PData(i,1);
            Spec(SZ(5),10) = PData(i,4);
        end
    end
end
end

%Creates matrix 'formulae' from 'PData', with the following columns :
%
% mz RI(U1) RI(HR) RI(DR) RI(U2) 12C 13C H O Err
% 1 2 3 4 5 6 7 8 9 10
%
```

```

MWs=[12.000000;13.003355;1.007825;15.994915];
Form = PData;
Form(1:pz, 6:10) = zeros(pz,5);
for i = 1 : pz
    if mod(floor(Form (i,1)),2)==1 && Form(i,6)==0 %odd nominal mass
        meas = Form (i,1)+1.007276; %converted neutral mass
        for c = 1 : floor(Form (i,1)/12)
            for h = 2*floor(0.1*c) : 2 : (2*c)+2;
                for o = 0 : c
                    calc = [c,0,h,o]*MWs;
                    err = abs((calc-meas)/meas)*1000000);
                    if err < 1
                        %assignes the 'i'th m/z a formula containing 12C,
                        %H, and/or O if err < 1 ppm
                        Form(i,6) = c;
                        Form(i,8) = h;
                        Form(i,9) = o;
                        Form(i,10) = err;
                        if c > 0
                            calc_C = Form(i,1) + 1.003355;
                            [err_CI, cz] = min( abs( 1000000 * (Form(:,1) -
                                calc_C) / calc_C ) );
                            if err_CI < 1
                                %identifies the 'cz'th m/z (i.e., the m/z
                                %with the smallest deviation) as the
                                %+13C-12C isotopologue of the 'i'th m/z
                                Form(cz,6) = c-1;
                                Form(cz,7) = 1;
                                Form(cz,8) = h;
                                Form(cz,9) = o;
                                Form(cz,10) = err_CI;
                            end
                        end
                    end
                end
            end
        end
    end
end

%VENN DIAGRAMS:
%matrix 'V_XX' (where XX = U2, HR, or DR):
%
%      U1-only   U1 & XX   XX-only
% # peaks      (1)       (2)       (3)
% # formulae   (2)
% % w/ form.   (3)
V_U2 = zeros(3,3);
for i = 1 : pz
    if Form(i,2) > 0 && Form(i,5) == 0
        V_U2(1,1) = V_U2(1,1) + 1;
        if Form(i,6) > 0
            V_U2(2,1) = V_U2(2,1) + 1;
        end
    elseif Form(i,2) > 0 && Form(i,5) > 0
        V_U2(1,2) = V_U2(1,2) + 1;
        if Form(i,6) > 0
            V_U2(2,2) = V_U2(2,2) + 1;
        end
    elseif Form(i,2) == 0 && Form(i,5) > 0
        V_U2(1,3) = V_U2(1,3) + 1;
        if Form(i,6) > 0
            V_U2(2,3) = V_U2(2,3) + 1;
        end
    end
end

```

```

    end
end
V_U2(3,:) = 100 * (V_U2(2,:) ./ V_U2(1,:));
V_HR = zeros(3,3);
for i = 1 : pz
    if Form(i,2) > 0 && Form(i,3) == 0
        V_HR(1,1) = V_HR(1,1) + 1;
        if Form(i,6) > 0
            V_HR(2,1) = V_HR(2,1) + 1;
        end
    elseif Form(i,2) > 0 && Form(i,3) > 0
        V_HR(1,2) = V_HR(1,2) + 1;
        if Form(i,6) > 0
            V_HR(2,2) = V_HR(2,2) + 1;
        end
    elseif Form(i,2) == 0 && Form(i,3) > 0
        V_HR(1,3) = V_HR(1,3) + 1;
        if Form(i,6) > 0
            V_HR(2,3) = V_HR(2,3) + 1;
        end
    end
end
end
V_HR(3,:) = 100 * (V_HR(2,:) ./ V_HR(1,:));
V_DR = zeros(3,3);
for i = 1 : pz
    if Form(i,2) > 0 && Form(i,4) == 0
        V_DR(1,1) = V_DR(1,1) + 1;
        if Form(i,6) > 0
            V_DR(2,1) = V_DR(2,1) + 1;
        end
    elseif Form(i,2) > 0 && Form(i,4) > 0
        V_DR(1,2) = V_DR(1,2) + 1;
        if Form(i,6) > 0
            V_DR(2,2) = V_DR(2,2) + 1;
        end
    elseif Form(i,2) == 0 && Form(i,4) > 0
        V_DR(1,3) = V_DR(1,3) + 1;
        if Form(i,6) > 0
            V_DR(2,3) = V_DR(2,3) + 1;
        end
    end
end
end
V_DR(3,:) = 100 * (V_DR(2,:) ./ V_DR(1,:));

%Identifies peaks as M+3, M+6, or M+9 masses, where M is an (odd nominal) m/z
%of a peak in U1 to which a molecular formula has been assigned. Stores
%data in 'Shifts' (derived from 'Form'), with the following columns:
%
% mz RI(U1) RI(HR) RI(DR) RI(U2) 12C 13C H O Err M+3? M+6? M+9?
% 1 2 3 4 5 6 7 8 9 10 11 12 13
%
%columns 10, 11, 12, 13 = 0 or 1
%
Shifts = Form;
Shifts(1:pz, 11:13) = zeros(pz,3);
for R = 1 : 3
    for i = 1 : pz
        if Shifts(i,2) > 0 && mod(floor(Shifts(i,1)),2) == 1 && Shifts(i,6) > 0
            M3 = Shifts(i,1) + 3.021927*R;
            [err_s, idx] = min( abs( 1000000 * (Shifts(:,1) - M3) / M3 ) );
            if err_s < 1
                Shifts(idx, 10 + R) = 1;
            end
        end
    end
end

```

```

        end
    end
end

%Creates Shift_T
%
%           %M+3R
%           Ntot R=1 R=2 R=3
%   I(all)  1   2   3   4
%   II(only) 2
%
Shift_T = zeros(2,4);
for i = 1 : pz
    if Shifts(i,2) > 0
        Shift_T(1,1) = Shift_T(1,1) + 1;
        if Shifts(i,11) == 1
            Shift_T(1,2) = Shift_T(1,2) + 1;
        end
        if Shifts(i,11) == 1 || Shifts(i,12) == 1
            Shift_T(1,3) = Shift_T(1,3) + 1;
        end
        if Shifts(i,11) == 1 || Shifts(i,12) == 1 || Shifts(i,13) == 1
            Shift_T(1,4) = Shift_T(1,4) + 1;
        end
    end
    if Shifts(i,2) == 0 && Shifts(i,4) > 0
        Shift_T(2,1) = Shift_T(2,1) + 1;
        if Shifts(i,11) == 1
            Shift_T(2,2) = Shift_T(2,2) + 1;
        end
        if Shifts(i,11) == 1 || Shifts(i,12) == 1
            Shift_T(2,3) = Shift_T(2,3) + 1;
        end
        if Shifts(i,11) == 1 || Shifts(i,12) == 1 || Shifts(i,13) == 1
            Shift_T(2,4) = Shift_T(2,4) + 1;
        end
    end
end
end

%IDENTIFICATION OF KETONE/ALDEHYDE-CONTAINING SPECIES
%
% matrices 'ID_kets' containing the following columns:
%   mz  DBE  O/C  H/C  ID
%   1   2   3   4   5
%
ID_kets = zeros(1,5);
Zn = 0;
for i = 1 : pz
    if mod(floor(Shifts(i,1)),2) == 1 && Shifts(i,2) > 0 && Shifts(i,6) > 0
        Zn = Zn + 1;
        ID_kets(Zn, 1) = Shifts(i,1);
        ID_kets(Zn, 2) = Shifts(i,6) + Shifts(i,7) + 1 - Shifts(i,8)/2;
        ID_kets(Zn, 3) = Shifts(i,9) / (Shifts(i,6) + Shifts(i,7));
        ID_kets(Zn, 4) = Shifts(i,8) / (Shifts(i,6) + Shifts(i,7));
        M3 = Shifts(i,1) + 3.021927;
        M6 = Shifts(i,1) + 6.043854;
        [err_m3, idx_m3] = min( abs( 1000000 * (Shifts(:,1) - M3) / M3 ) );
        [err_m6, idx_m6] = min( abs( 1000000 * (Shifts(:,1) - M6) / M6 ) );
        if err_m3 < 1 && Shifts(idx_m3,4) > 0
            ID_kets(Zn,5) = ID_kets(Zn,5) + 1;
        end
        if err_m6 < 1 && Shifts(idx_m6,4) > 0
            ID_kets(Zn,5) = ID_kets(Zn,5) + 2;
        end
    end
end

```

```

        end
    end
end

%Counts number of non, singly, and/or doubly reduced species at odd nominal
%m/z to which a molecular formula was assigned.
Z0 = 0;
Z1 = 1;
Z2 = 2;
Z3 = 3;
for i = 1 : Zn
    if ID_kets(i,5) == 0
        Z0 = Z0 + 1;
    elseif ID_kets(i,5) == 1
        Z1 = Z1 + 1;
    elseif ID_kets(i,5) == 2
        Z2 = Z2 + 1;
    elseif ID_kets(i,5) == 3
        Z3 = Z3 + 1;
    end
end
ID_kets = sortrows(ID_kets,[5 1]);
disp([' non-reduced:      ',num2str(Z0),' (',num2str(100*Z0/Zn),'%')]);
disp([' singly-only:     ',num2str(Z1),' (',num2str(100*Z1/Zn),'%')]);
disp([' doubly-only:     ',num2str(Z2),' (',num2str(100*Z2/Zn),'%')]);
disp([' singly & doubly: ',num2str(Z3),' (',num2str(100*Z3/Zn),'%')]);

end

```

49.2. Blank subtraction and excision of ions with multiple charges

```

function [ PEAKS, PEAKS_T ] = ms_list( mz_min, mz_max, raw, tol, blank, z_lim )

%Pre-processes a raw mass spectral peak list:
% * includes peaks in a specified m/z range
% * blank subtracts peaks
% * removes multiply-charged peaks
% * displays mass distribution information (stored in 'PEAKS_T')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%REQUIRED INPUT ARGUMENTS
%
% 'mz_min' and 'mz_max': Positive numbers specifying lower and upper limits
% of the the m/z range to be included in the final peak list
% (mz_min must be less than mz_max).
%
% 'raw': Two-column matrix containing the raw peak list of the mass
% spectrum of the sample to be analyzed, with m/z in the first column
% and raw intensities in the second column.
%
%OPTIONAL INPUT ARGUMENTS
%
% 'tol': Error limit (in ppm; positive number) used to find and excise
% peaks that are present in the blank spectrum and to search for and excise
% multiply-charged peaks.
%
% 'blank': Two-column matrix containing the raw peak list of the mass
% spectrum of a corresponding solvent blank, with m/z in the first

```

```

% column and raw intensities in the second column.
%
% 'z_lim': Single integer or vector of integers specifying charge states
% that are to be excised based on 1.003355 / z mass spacings. If
% z_lim = 0, no species will be excised in this manner.
% Default value = [2 3].
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%OUTPUT ARGUMENTS
%
% 'PEAKS': Two-column matrix containing m/z (first column) and
% normalized intensity data (second column) of the sample after the
% following processes are performed on the raw data:
% (1) Excision of peaks outside the specified m/z range
% (2) If 'tol' (and 'z_lim') is/are given:
%     Excision of peaks which differ by 1.003355/z within 'tol' ppm,
%     where z is the charge state(s) of the ion specified by 'z_lim'
%     (i.e., the monoisotopic peak and the corresponding 13C
%     isotopologue). If more than one peak is found within 'tol' ppm
%     of a theoretical isotopologue's m/z, the peak with the smallest
%     deviation from the theoretical value will be excised.
% (3) If 'tol' & 'blank' are given:
%     Excision of peaks which have m/z values within 'tol' ppm of a
%     peak present in 'blank'.
% (4) Normalization of intensities to base peak
%
%
% 'PEAKS_T': Table with the following rows (displayed when function ends)
%
% (1) # peaks excised (blank)
% (2) # peaks excised (z)
% (3) # peaks remaining
% (4) total ion count (TIC)
% (5) number-average m/z (AMN) = sum(mz*intensity)/sum(intensity)
% (6) weight-average m/z (AMW) = sum(mz^2*intensity)/sum(mz*intensity)
% (7) polydispersity = AMN / AMW
% (8) intensity range = maximum intensity / minimum intensity
% (9) m/z of base peak
% (10) raw intensity of base peak
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

PEAKS1 = raw;

T_values = zeros(10,1);

%converts 'isnan' values to 0 (if needed)
PEAKS1(isnan(PEAKS1)) = 0;

%default value for 'z_lim' if not specified
if nargin < 6
    z_lim = [2 3];
end

for i = 1 : size(PEAKS1,1)

    %Excises peaks (replaces row with [0 0]) with intensity equal to 0
    if PEAKS1(i,2) == 0;
        PEAKS1(i,1:2)=zeros(1,2);
    end

    %Excises peaks outside the specified m/z range
    if PEAKS1(i,1) < mz_min || PEAKS1(i,1) > mz_max

```

```

        PEAKS1(i,1:2) = zeros(1,2);

    else
        %Excises multi-z peaks if 'tol' is given
        if nargin > 3 && sum(z_lim) ~= 0
            thr_mass = PEAKS1(i,1) + (1.003355 ./ z_lim);
            for k = 1 : size(thr_mass)
                %Finds peak with smallest deviation from theoretical
                %isotopologue mass
                [err, z] = min( abs( 1000000 * (PEAKS1(:,1) - thr_mass(k)) /
thr_mass(k) ) );
                if err <= tol
                    PEAKS1(i,1:2) = zeros(1,2);
                    PEAKS1(z,1:2) = zeros(1,2);
                    T_values(2) = T_values(2) + 1;
                end
            end
        end

        %Excises peaks shared in 'blank' (if specified)
        if nargin > 4;
            for k = 1 : size(blank,1)
                if abs( 1000000 * ( PEAKS1(i,1) - blank(k,1) ) / blank(k,1) )
<= tol
                    PEAKS1(i,1:2) = zeros(1,2);
                    T_values(1) = T_values(1) + 1;
                end
            end
        end
    end

end

end

[T_values(10), BP_index] = max(PEAKS1(:,2)); %m/z of base peak
T_values(9) = PEAKS1(BP_index,1); %raw intensity of base peak
T_values(4) = sum( PEAKS1(:,2) ); %TIC

%stores non-excised values in 'PEAKS' (removes rows of zeros)
x = 0;
PEAKS = zeros(1,2);
for i = 1 : size(PEAKS1,1)
    if PEAKS1(i,1) > 0
        x = x + 1;
        PEAKS(x,1) = PEAKS1(i,1);
        PEAKS(x,2) = 100* PEAKS1(i,2) / T_values(10);
    end
end

T_values(3) = size( PEAKS,1 ); %number of peaks
T_values(5) = sum( PEAKS(:,1).*PEAKS(:,2) ) / sum(PEAKS(:,2)); %AMW(n)
T_values(6) = sum( PEAKS(:,1).*PEAKS(:,1).*PEAKS(:,2) ) / sum(
PEAKS(:,1).*PEAKS(:,2) ); %AMW(w)
T_values(7) = T_values(6) / T_values(5); %Polydispersity
T_values(8) = max( PEAKS(:,2) ) / min( PEAKS(:,2) ); %intensity range

T_rows = {'peaks excised (blank)'
'peaks excised (z) '
'peaks remaining '
'TIC (raw) '
'avg mz (number) '
'avg mz (weight) '
'polydispersity '
'intensity range '

```

```

        'base peak mz      '
        'base peak raw int.  '};

PEAKS_T = table(T_values, 'RowNames', T_rows);
disp(PEAKS_T);

```

```
end
```

```

=====
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
=====

```

A9.3. Molecular formula assignment (“brute-force”)

```

function [FORM, FORM_T] = ms_form1(peaks, tol, n_max, s_max, d_max)
%Calculates molecular formulae for m/z values in 'peaks'
%using a "brute force" algorithm

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%REQUIRED INPUT ARGUMENTS
%
% 'peaks': Matrix containing at least two columns with m/z and
% intensity data in the first and second columns, respectively.
%
% 'tol': Error threshold (in ppm) allowed for molecular formula assignment
%
%OPTIONAL INPUT ARGUMENTS
%
% 'n_max': maximum number of 14N atoms allowed. Default value = 0.
%
% 's_max': maximum number of 32S atoms allowed. Default value = 0.
%
% 'd_max': maximum number of deuterium (2H) atoms allowed.
% Default value = 0.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%OUTPUT ARGUMENT
%
% 'FORM': Matrix of peaks to which a molecular formula has been
% assigned with the following columns:
%
%      ...Number of atoms ...
% [mz int 12C 13C H D O N 32S 34S FE O/C H/C N/C AI  DBE DBE-O]
% 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
%
% Column 11: formula error (in ppm)
% Column 12: O/C molar ratio (12C & 13C included)
% Column 13: H/C molar ratio (H & D included)
% Column 14: N/C molar ratio
% Column 15: aromaticity index (AI; described by Koch and Dittmar in
% Rapid. Comm. Mass Spectrom. 2005, 20, 926-932.
% AI = (1 + C - O - S - 0.5H) / (C - O - S - N)
% Column 16: Double bond equivalent (DBE)
% DBE = 1 + C - H/2 - N/2

```

```

%      Column 17: Double bond equivalent minus # O atoms
%
%
%
% 'FORM_T': 18 x 3 table containing the following data of the calculated
% molecular formulae:
%
%           number or   percentage or
%           average     minimum       maximum
%           (1)         (2)           (3)
% (1) Intensity
% (2) Peaks
% (3) #CHO
% (4) #CHON
% (5) #CHONS
% (6) #CHOS
% (7) % mass C
% (8) % mass H
% (9) % mass O
% (10) % mass N
% (11) % mass S
% (12) O/C
% (13) H/C
% (14) N/C
% (15) AI
% (16) DBE
% (17) DBE-O
% (18) error (ppm)
%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%EXPLANATION OF MOLECULAR FORMULA ASSIGNMENT
%
%PART 1. For each mass in 'peaks', all chemically reasonable molecular
%formulae with the atoms 12C, 1H, 2H, 16O, 14N, 32S are calculated using
%the following criteria:
%
% 33 < %C < 100
% 2 <= #H <= 2(#C) + 2
% 0 <= C
% N <= (user specified)
% S <= (user specified)
% DBE > 1 (double-bond equivalent, where DBE = C + 1 - H/2 + N/2)
% DBE is an integer
% O+N+S > 0
%
%The difference in ppm between the experimentally measured m/z (MM)
%and the theoretical m/z of each formula (CM) is calculated using the
%following equation:
%
% FE = 1000000*(MM-CM)/CM
%
% where CM = 12*C + 1.0078250*H + 2.0141018*D + 15.9949146*O +
% 14.0030740*N + 33.9678669*S
%
% If the absolute value of FE is less than or equal to the error
% threshold specified by 'tol', the formula is assumed to be correct. If
% more than one formula is possible, the one with the lowest number of
% D, N, and S is used.
%
%PART 3. +13C-12C and +34S-32S isotopologue peaks are searched for each
%peak to which a molecular formula was assigned. If a peak whose measured
%mass is within 'tol' ppm of the calculated mass of the isotopologue
%peaks is found, the formula of this isotopologue is used.
%
%

```



```

end
end
end
end
end
TotalTime = toc;

%calculates molar ratios, etc for each formula
for i = 1 : size(FORM1,1)
    if FORM1(i,3) > 0
        c = FORM1(i,3) + FORM1(i,4);
        h = FORM1(i,5) + FORM1(i,6);
        o = FORM1(i,7);
        n = FORM1(i,8);
        s = FORM1(i,9) + FORM1(i,10);
        FORM1(i,12) = o/c;
        FORM1(i,13) = h/c;
        FORM1(i,14) = n/c;
        FORM1(i,15) = (1+c-o-s-(0.5*h)) / (c-o-s-n); %AI
        if FORM1(i,15) < 0
            FORM1(i,15) = 0;
        end
        FORM1(i,16) = c + 1 - (h/2) + (n/2); %DBE
        FORM1(i,17) = c + 1 - (h/2) + (n/2) - o; %DBE-O
    end
end

%Deletes rows with no formula
FORM = zeros(1,17);
fx = 0;
for i=1:size(FORM1,1)
    if FORM1(i,3) > 0
        fx = fx + 1;
        FORM(fx,1:17) = FORM1(i, 1:17);
    end
end

%Builds formula distribution table
numb_avg = zeros (19,1);
perc_min = zeros (19,1);
zero_max = zeros (19,1);

%TIC assigned (number and percent)
numb_avg(1) = sum(FORM(:,2));
perc_min(1) = 100 * numb_avg(1) / sum(peaks(:,2));

%peaks assigned (number and percent)
numb_avg(2) = size(FORM,1);
perc_min(2) = 100 * numb_avg(2) / size(peaks,1);

for i=1:size(FORM,1)
    if FORM(i,8)==0 && sum(FORM(i,9:10))==0
        numb_avg(3) = numb_avg(3) + 1; % # CHO
    elseif FORM(i,8)>0 && sum(FORM(i,9:10))==0
        numb_avg(4) = numb_avg(4) + 1; % # CHON
    elseif FORM(i,8)>0 && sum(FORM(i,9:10))>0
        numb_avg(5) = numb_avg(5) + 1; % # CHONS
    elseif FORM(i,8)==0 && sum(FORM(i,9:10))>0
        numb_avg(6) = numb_avg(6) + 1; % # CHOS
    end
end
end

```

```

perc_min(3:6) = numb_avg(3:6) .* ( 100/sum(numb_avg(3:6)) );

%intensity-weighted mass sums
mass_tot = sum( FORM(:,2) .* ( FORM(:,1)+1.01) );
mass_c = sum( FORM(:,2) .* ( 12.000*FORM(:,3) + 13.003*FORM(:,4) ) );
mass_h = sum( FORM(:,2) .* ( 1.008*FORM(:,5) + 2.041*FORM(:,6) ) );
mass_o = sum( FORM(:,2) .* ( 15.995*FORM(:,7) ) );
mass_n = sum( FORM(:,2) .* ( 14.003*FORM(:,8) ) );
mass_s = sum( FORM(:,2) .* ( 31.972*FORM(:,9) + 33.968*FORM(:,10) ) );
%percent (mass) C, H, O, N, S
perc_min(7) = 100 * mass_c / mass_tot;
perc_min(8) = 100 * mass_h / mass_tot;
perc_min(9) = 100 * mass_o / mass_tot;
perc_min(10) = 100 * mass_n / mass_tot;
perc_min(11) = 100 * mass_s / mass_tot;

for r = 12 : 17
    numb_avg(r) = sum(FORM(:,r).*FORM(:,2)) / sum(FORM(:,2));
    perc_min(r) = min(FORM(:,r));
    zero_max(r) = max(FORM(:,r));
end

%error (RMSE, min, max)
numb_avg(18) = rms(FORM(:,11));
perc_min(18) = min(abs(FORM(:,11)));
zero_max(18) = max(abs(FORM(:,11)));

%time (min)
numb_avg(19) = TotalTime / 60;

Categories = { 'TIC'           '
               'peaks        '
               'CHO          '
               'CHON         '
               'CHONS        '
               'CHOS         '
               '% mass C     '
               '% mass H     '
               '% mass O     '
               '% mass N     '
               '% mass S     '
               'O/C          '
               'H/C          '
               'N/C          '
               'AI           '
               'DBE          '
               'DBE-O        '
               'err (ppm)    '
               'time (min)  ' };

FORM_T = table(numb_avg, perc_min, zero_max, 'RowNames',Categories);

disp(' formula distribution ')
disp(FORM_T)

end

%=====
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====

```

A9.4. Molecular formula assignment ("low-mass moiety")

```
function [FORM, FORM_T] = ms_form2(peaks, tol, n_max, s_max, d_max)
%Calculates molecular formulae for m/z values in 'peaks'
%using a "low-mass moiety" algorithm as described by Green and Perdue in
%Anal. Chem. 2015, 87, 5079-5085 and
%Anal. Chem. 2015, 87, 5086-5094

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%REQUIRED INPUT ARGUMENTS
%
% 'peaks': Matrix containing at least two columns with m/z and
% intensity data in the first and second columns, respectively.
%
% 'tol': Error threshold (in ppm) allowed for molecular formula assignment
%
%OPTIONAL INPUT ARGUMENTS
%
% 'n_max': maximum number of 14N atoms allowed. Default value = 0.
%
% 's_max': maximum number of 32S atoms allowed. Default value = 0.
%
% 'd_max': maximum number of deuterium (2H) atoms allowed.
% Default value = 0.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%OUTPUT ARGUMENT
%
% 'FORM': Matrix of peaks to which a molecular formula has been
% assigned with the following columns:
%
%           ...Number of atoms ...
% [mz int 12C 13C H D O N 32S 34S FE O/C H/C N/C AI  DBE DBE-O]
% 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
%
%   Column 11: formula error (in ppm)
%   Column 12: O/C molar ratio (12C & 13C included)
%   Column 13: H/C molar ratio (H & D included)
%   Column 14: N/C molar ratio
%   Column 15: aromaticity index (AI; described by Koch and Dittmar in
%   Rapid. Comm. Mass Spectrom. 2005, 20, 926-932.
%   AI = (1 + C - O - S - 0.5H) / (C - O - S - N)
%   Column 16: Double bond equivalent (DBE)
%   DBE = 1 + C - H/2 - N/2
%   Column 17: Double bond equivalent minus # O atoms
%
%
%
% 'FORM_T': 18 x 3 table containing the following data of the calculated
% molecular formulae:
%
%           number or  percentage or
%           average   minimum     maximum
%           (1)        (2)         (3)
% (1) Intensity
% (2) Peaks
% (3) #CHO
% (4) #CHON
% (5) #CHONS
```

```

% (6) #CHOS
% (7) % mass C
% (8) % mass H
% (9) % mass O
% (10) % mass N
% (11) % mass S
% (12) O/C
% (13) H/C
% (14) N/C
% (15) AI
% (16) DBE
% (17) DBE-O
% (18) error (ppm)
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%EXPLANATION OF MOLECULAR FORMULA ASSIGNMENT
%
%PART 1. Each measured m/z is converted to its neutral mass (by addition of
%the mass of H+), and a base molecular formula of a hydrocarbon with the
%lowest possible H/C molar ratio is calculated for that nominal mass.
%This base formula is altered by adding or subtracting C4O-3 and/or CH4O-1
%subunits until the calculated mass of the formula is within 'tol' ppm of
%the measured mass. Heteroatoms are included by varying the number of 32S
%and/or NH subunits to the base hydrocarbon. The following constraints must
%be satisfied:
%
%   2 <= #H <= 2(#C) + 2
%   O <= C
%   N <= (user specified)
%   S <= (user specified)
%   O+N+S > 0
%
%The difference in ppm between the experimentally measured m/z (MM)
%and the theoretical m/z of each formula (CM) is calculated using the
%following equation:
%
%   FE = 1000000*(MM-CM)/CM
%
%   where CM = 12*C + 1.0078250*H + 2.0141018*D + 15.9949146*O +
%   14.0030740*N + 33.9678669*S
%
%   If the absolute value of FE is less than or equal to the error
%   threshold specified by 'tol', the formula is assumed to be correct.
%
%PART 2. +13C-12C and +34S-32S isotopologue peaks are searched for each
%peak to which a molecular formula was assigned. If a peak whose measured
%mass is within 'tol' ppm of the calculated mass of the isotopologue
%peaks is found, the formula of this isotopologue is used.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
tic

if nargin < 5
    d_max = 0;
    if nargin < 4
        s_max = 0;
        if nargin < 3
            n_max = 0;
        end
    end
end
end

```

```

%Calculates molecular formulae
FORM1 = zeros( size(peaks,1),18 );
FORM1(1:size(peaks,1),1:size(peaks,2)) = peaks;
Masses = [12; 13.0033548; 1.0078250; 2.0141018; 15.9949146; 14.0030740;
31.9720710; 33.9678669];
for x = 1 : size(FORM1,1)
    if sum(FORM1(x,3:10)) == 0
        MM = FORM1(x,1) + 1.0072764;
        for d = 0 : d_max;
            for s = 0 : s_max;
                for n = 0 : n_max;
                    form = zeros(1,8);
                    sum_HA = d*3.0219268 + s*31.9720710 + n*15.0108990;
                    MMC = MM - sum_HA;
                    if mod(round(MMC),2) == 0
                        i = 0;
                        %calculate initial molecular formula
                        form(1) = floor(MMC/12);
                        form(3) = round(MMC - 12*form(1)) + n + d;
                        form(4) = d;
                        form(6) = n;
                        form(7) = s;
                        while i < 100
                            %calculate current formula mass
                            CM = form * Masses;
                            %calculate formula error
                            FE = 1000000*(MM - CM)/CM;
                            if abs(FE) <= tol
                                %end 'while' loop, check formula
                                i = 100;
                            else
                                %calculate number of CH4-O replacements needed
                                (M)
                                M = round((MM - CM)/0.0363854);
                                %calculate formula error after M CH4-O
                                replacements
                                FE = 1000000*(MM - CM - M*0.0363854) / CM;
                                if abs(FE) <= tol
                                    %perform M +CH4-O replacements
                                    form(1) = form(1) + M;
                                    form(3) = form(3) + 4*M;
                                    form(5) = form(5) - M;
                                    %end 'while' loop, check formula
                                    i = 100;
                                else
                                    %perform +O3-C4 replacement
                                    form(1) = form(1) - 4;
                                    form(5) = form(5) + 3;
                                    i = i + 1;
                                end
                            end
                        end
                        %check formulae
                        CM = form * Masses;
                        %calculate formula error
                        FE = 1000000*(MM - CM)/CM;
                        if abs(FE) <= tol
                            if sum(FORM1(x,3:10)) == 0 || ( sum(FORM1(x,3:10))
> 0 && sum(form(6:8)) < sum(FORM1(x,8:10)) );
                                if form(1) > 0 %C > 0
                                    if form(3) >= 2 && form(3) <= 2*form(1)+2
                                        %2 <= H <= 2C+2
                                        if sum(form(5:8))>0 % O+N+S > 0
                                            if form(5) <= form(1) && form(5)>=0
                                                ; % 0 <= O<= C

```



```

        FORM1(i,17) = c + 1 - (h/2) + (n/2) - o; %DBE-O
    end
end

%Deletes rows with no formula
FORM = zeros(1,17);
fx = 0;
for i=1:size(FORM1,1)
    if sum(FORM1(i,3:10)) > 0
        fx = fx + 1;
        FORM(fx,1:17) = FORM1(i, 1:17);
    end
end

%Builds formula distribution table
numb_avg = zeros (19,1);
perc_min = zeros (19,1);
zero_max = zeros (19,1);

%TIC assigned (number and percent)
numb_avg(1) = sum(FORM(:,2));
perc_min(1) = 100 * numb_avg(1) / sum(peaks(:,2));

%peaks assigned (number and percent)
numb_avg(2) = size(FORM,1);
perc_min(2) = 100 * numb_avg(2) / size(peaks,1);

for i=1:size(FORM,1)
    if FORM(i,8)==0 && sum(FORM(i,9:10))==0
        numb_avg(3) = numb_avg(3) + 1; % # CHO
    elseif FORM(i,8)>0 && sum(FORM(i,9:10))==0
        numb_avg(4) = numb_avg(4) + 1; % # CHON
    elseif FORM(i,8)>0 && sum(FORM(i,9:10))>0
        numb_avg(5) = numb_avg(5) + 1; % # CHONS
    elseif FORM(i,8)==0 && sum(FORM(i,9:10))>0
        numb_avg(6) = numb_avg(6) + 1; % # CHOS
    end
end
perc_min(3:6) = numb_avg(3:6) .* ( 100/sum(numb_avg(3:6)) );

%intensity-weighted mass sums
mass_tot = sum( FORM(:,2).* ( FORM(:,1)+1.01) );
mass_c = sum( FORM(:,2).* ( 12.000*FORM(:,3) + 13.003*FORM(:,4) ) );
mass_h = sum( FORM(:,2).* ( 1.008*FORM(:,5) + 2.041*FORM(:,6) ) );
mass_o = sum( FORM(:,2).* ( 15.995*FORM(:,7) ) );
mass_n = sum( FORM(:,2).* ( 14.003*FORM(:,8) ) );
mass_s = sum( FORM(:,2).* ( 31.972*FORM(:,9) + 33.968*FORM(:,10) ) );
%percent (mass) C, H, O, N, S
perc_min(7) = 100 * mass_c / mass_tot;
perc_min(8) = 100 * mass_h / mass_tot;
perc_min(9) = 100 * mass_o / mass_tot;
perc_min(10) = 100 * mass_n / mass_tot;
perc_min(11) = 100 * mass_s / mass_tot;

for r = 12 : 17
    numb_avg(r) = sum(FORM(:,r).*FORM(:,2)) / sum(FORM(:,2));
    perc_min(r) = min(FORM(:,r));
    zero_max(r) = max(FORM(:,r));
end

%error (RMSE, min, max)
numb_avg(18) = rms(FORM(:,11));
perc_min(18) = min(abs(FORM(:,11)));

```

```

zero_max(18) = max(abs(FORM(:,11)));

%time (min)
numb_avg(19) = TotalTime / 60;

Categories = {'TIC          '
             'peaks        '
             'CHO          '
             'CHON         '
             'CHONS        '
             'CHOS         '
             '% mass C      '
             '% mass H      '
             '% mass O      '
             '% mass N      '
             '% mass S      '
             'O/C          '
             'H/C          '
             'N/C          '
             'AI           '
             'DBE          '
             'DBE-O        '
             'err (ppm)    '
             'time (min)  '};

FORM_T = table(numb_avg, perc_min, zero_max, 'RowNames',Categories);

disp(' formula distribution ')
disp(FORM_T)

end

%=====
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====

=====


```

49.5. Ketone/aldehyde identification

```

function [LAB, LAB_T, RED, RED_T] = ms_ketid(form, peaks, tol, n_max, s_max,
d_max)

%Identifies species in 'form' which comprise species with borodeuteride
%reducible moieties (i.e. ketones, aldehydes) by searching M+3.021927 &
%M+6.043854 m/z in 'peaks'
%
%Calculates molecular formulae for unassigned masses in 'peaks' by calling
%function 'ms_form.m'

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ABBREVIATIONS:
%   "M+3" = M + 3.021 926 8
%   "M+6" = M + 6.043 853 6
%

```

```

% where M refers to an m/z in 'form'
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%INPUT ARGUMENTS
%
% 'form': Matrix of peaks to which a molecular formula has been
% assigned with the following columns (see function 'f_mf_calc.m' for more
% details):
%
%           ...Number of atoms ...
% [mz int 12C 13C H D O N 32S 34S FE O/C H/C N/C AI  DBE DBE-O]
%  1  2   3   4   5 6 7 8   9 10 11 12 13 14 15 16 17
%
%
% 'peaks': Matrix containing at least two columns, with the first
% column containing m/z values and the second column containing intensities
% (should be a borodeuteride-reduced mass spectral peak list).
%
%
% 'tol': Error threshold (in ppm) allowed for molecular formula assignment
% and M+3/6 searches
%
%OPTIONAL INPUT ARGUMENTS
%
% 'n_max': maximum number of 14N atoms allowed. Default value = 0.
%
% 's_max': maximum number of 32S atoms allowed. Default value = 0.
%
% 'd_max': maximum number of deuterium (2H) atoms allowed.
% Default value = 2.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%OUTPUT ARGUMENTS
%
% 'LAB': Same format as 'form' with column 18 data added, which identifies
% the peak composition based on presence/absence of M+3 or M+6 masses in
% the searched mass spectral peak list. Values and meanings are as follows:
%   0 : non-reduced species only (default)
%   1 : singly-reduced species only
%   2 : doubly-reduced species only
%   3 : singly and doubly reduced species
% theoretical M+3 and M+6 masses are determined from the calculated
% (theoretical) mass of the given molecular formula. An M+3 or M+6 mass
% marker is considered to be present if there is a peak with m/z within
% 'tol' ppm of the theoretical M+3 or M+6 mass.
%
%
% 'LAB_T': Table with rows and columns defined as follows:
%
%   Columns: Peaks comprising species which are ___ reduced:
%           (1) non
%           (2) singly-reduced only
%           (3) doubly-reduced only
%           (4) singly- & doubly-reduced only
%
%
% 'RED': Same format as 'FORM' from function 'f_mf_calc.m'
%
% 'RED_T': Same format as 'FORM_T' from function 'f_mf_calc.m'
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
tic

if nargin < 6
    d_max = 2;

```

```

    if nargin < 5
        s_max = 0;
        if nargin < 4
            n_max = 0;
        end
    end
end
end

LAB = form;
LAB(:,18) = zeros(size(form,1),1);
RED_temp(1:size(peaks,1),1:2) = peaks(:,1:2);

if nargin > 2
    RED_temp(1:size(peaks,1),3:17) = zeros(size(peaks,1),15);
end

Masses = [12; 13.0033548; 1.0078250; 2.0141018; 15.9949146; 14.0030740;
31.9720710; 33.9678669];

for i = 1:size(LAB,1);
    %calculated M minus the mass of H+
    M = LAB(i,3:10)*Masses - 1.0072764;
    %M = LAB(i,1);
    M3 = M + 3.0219268; %theoretical M+3 mass
    M6 = M + 6.0438536; %theoretical M+6 mass
    if d_max >= 1
        %Search for m/z in 'RED_temp' closest to theoretical M+3
        [M3_diff, M3_in] = min( abs( M3 - RED_temp(:,1) ) );
        %if M+3 found & formula contains at least 1 O or N AND at least 1
DBE...
        if tol >= 1000000 * M3_diff / M3 && sum(LAB(i,7:8)) > 0 && LAB(i,16) >
0
            %increase ketone/aldehyde ID by 1
            LAB(i,18) = LAB(i,18) + 1;
            %if formulae are to be calculated for 'peaks'...
            if nargin > 2
                %mass shift-based molecular formula stored in 'RED_temp'
                RED_temp(M3_in,3:17) = LAB(i,3:17);
                RED_temp(M3_in,5) = RED_temp(M3_in,5) + 1;
                RED_temp(M3_in,6) = 1;
                calc = (RED_temp(M3_in,3:10)*Masses)-1.0072764;
                meas = RED_temp(M3_in,1);
                RED_temp(M3_in,11) = 1000000 * (meas - calc)/calc;
                RED_temp(M3_in,13) = sum(RED_temp(M3_in,5:6)) /
sum(RED_temp(M3_in,3:4));
                AI_d = 0.5 + sum(RED_temp(M3_in,3:4)) - RED_temp(M3_in,7) -
sum(RED_temp(M3_in,9:10)) - RED_temp(M3_in,5)/2;
                AI_c = sum(RED_temp(M3_in,3:4)) - RED_temp(M3_in,7) -
sum(RED_temp(M3_in,9:10)) - RED_temp(M3_in,8);
                RED_temp(M3_in,15) = AI_d / AI_c;
                RED_temp(M3_in,16) = RED_temp(M3_in,16) - 1;
                RED_temp(M3_in,17) = RED_temp(M3_in,17) - 1;
            end
        end
    end

    if d_max >= 2
        %Search for m/z in 'RED_temp' closest to theoretical M+6
        [M6_diff, M6_in] = min( abs( M6 - RED_temp(:,1) ) );
        %if M+6 found & formula contains at least 2 O or N AND at least 2
DBE...
        if tol >= 1000000 * M6_diff / M6 && sum(LAB(i,7:8)) > 1 && LAB(i,16) >
1
            %increase ketone/aldehyde ID by 2

```

```

LAB(i,18) = LAB(i,18) + 2;
%if formulae are to be calculated for 'peaks'...
if nargout > 2
    %mass shift-based molecular formula stored in 'RED_temp'
    RED_temp(M6_in,3:17) = LAB(i,3:17);
    RED_temp(M6_in,5) = RED_temp(M6_in,5) + 2;
    RED_temp(M6_in,6) = 2;
    calc = (RED_temp(M6_in,3:10)*Masses)-1.0072764;
    meas = RED_temp(M6_in,1);
    RED_temp(M6_in,11) = 1000000 * (meas - calc)/calc;
    RED_temp(M6_in,13) = sum(RED_temp(M6_in,5:6)) /
sum(RED_temp(M6_in,3:4));
    AI_d = 1 + sum(RED_temp(M6_in,3:4)) - RED_temp(M6_in,7) -
sum(RED_temp(M6_in,9:10)) - RED_temp(M6_in,5)/2 - 1;
    AI_c = sum(RED_temp(M6_in,3:4)) - RED_temp(M6_in,7) -
sum(RED_temp(M6_in,9:10)) - RED_temp(M6_in,8);
    RED_temp(M6_in,15) = AI_d / AI_c;
    RED_temp(M6_in,16) = RED_temp(M6_in,16) - 2;
    RED_temp(M6_in,17) = RED_temp(M6_in,17) - 2;
end
end
end
end

%Creates 'LAB_T'
num = zeros(4,1);
for i = 1:size(LAB,1);
    if LAB(i,18) == 0
        num(1) = num(1) + 1;
    elseif LAB(i,18) == 1
        num(2) = num(2) + 1;
    elseif LAB(i,18) == 2
        num(3) = num(3) + 1;
    elseif LAB(i,18) == 3
        num(4) = num(4) + 1;
    end
end
perc = 100 * num / sum(num);
Categories = {'non'      '
'singly  '
'doubly  '
'both    '};
LAB_T = table(num, perc, 'RowNames',Categories);
disp(LAB_T);

%Calculates molecular formulae for remaining peaks in 'RED_temp'
if nargout > 2
    [RED, RED_T] = ms_form2(RED_temp, tol, n_max, s_max, 0);
end
end

%=====
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====
=====

```

A9.6. Peak/formula list subset selection

```
function [ Selected ] = ms_sub( List, ColVar, VarMin, VarMax )
%Returns a truncated list of peaks or molecular formula based on a minimum
%and maximum value of a given variable
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%INPUT ARGUMENTS
%
%'List': Complete list of molecular formulae (or m/z values)
%
%'ColVar': Column number of 'List' containing the variable on which to base
%the truncation
%
%'VarMin': Minimum value of column specified by 'ColVar' allowed in the
%output variable 'Selected'.
%
%'VarMax': Maximum value of column specified by 'ColVar' allowed in the
%output variable 'Selected'. If 'VarMax' is not given, then 'VarMin'
%specifies the only value allowed in the column specified by 'ColVar' is
%'VarMin'.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%OUTPUT ARGUMENTS
%
%'Selected': List of molecular formula (or m/z values; derived from 'List')
%consisting only of rows in which the values of the column specified by
%'ColVar' is greater than or equal to 'VarMin' and less than 'VarMax' (if
%'VarMax' is specified)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

x = 0;
cols = size(List,2);
Selected = zeros(1,cols);
for i = 1 : size(List,1)
    if nargin == 3
        if List(i,ColVar) == VarMin
            x = x + 1;
            Selected(x,1:cols) = List(i,1:cols);
        end
    elseif nargin == 4
        if List(i,ColVar) >= VarMin && List(i,ColVar) < VarMax
            x = x + 1;
            Selected(x,1:cols) = List(i,1:cols);
        end
    end
end
end

%=====
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====
```

A9.7. Comparison of peaks and/or molecular formulae

```
function [ TAB_p, ALIGNED, TAB_ni ] = ms_compare( tol, varargin )

%Compares two or more mass spectral data sets (molecular formulae or
%peak lists). Returns either a table of Venn diagram data or a dendrogram
%displaying Bray-Curtis linkages.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%INPUT ARGUMENTS
%
% 'tol' m/z error (in ppm) tolerance used for peak alignment
%
% 'varargin'. Matrices (at least 2) comprising peak lists (2 columns) or
% molecular formulae lists (at least 18 columns) to be compared.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%OUTPUT ARGUMENTS
%
% 'ALIGNED' Aligned peaks or molecular formulae containing the following
% columns:
%
%   mz int1 int2...intN
%   1  2     3     1+N
%
%
% 'TAB_p'
% If 2 or 3 peak or formula lists are compared:
%   Table of venn diagram data. Columns: (1) number and (2)percentage of
%   peaks or formulae in the subset specified by each row, and (3) number
%   and (4) percentages of D-containing formulae in the subset specified
%   by each row.
%
%           variables (columns):
%           A_num A_per D_num D_per
%   rows   (1)  (2)  (3)  (4)
%   (1) A-only
%   (2) A & B
%   (3) B-only
%   (4) B & C
%   (5) C-only
%   (6) C & A
%   (7) A,B,C
%
% If 4 or more peak or formula lists are compared:
%   Matrix containing Bray-Curtis linkage matrix, where the numbers of
%   the rows and column corresponds to a single sample in the order in
%   which they were entered in the function.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Aligns molecular formulae or mz
Masses = [12; 13.0033548; 1.0078250; 2.0141018; 15.9949146; 14.0030740;
31.9720710; 33.9678669];
TEMP = 0;
start = 1;
for i = 1 : nargin-1
    out = varargin{i};
    out_rows = size(out,1);
    out_cols = size(out,2);
```

```

stop = start + out_rows - 1;
if out_cols > 2 %formula & intensity data are inputted
    alignd_cols = 9; %formula
    TEMP(start:stop,2:9) = out(:,3:10);
    TEMP(start:stop,i+9) = out(:,2);
    start = stop+1;
elseif out_cols == 2 %only m/z & intensity data are inputted
    alignd_cols = 1; %m/z
    TEMP(start:stop,1) = out(:,1);
    TEMP(start:stop,i+1) = out(:,2);
    start = stop+1;
end
end
if alignd_cols == 9
    for i = 1 : size(TEMP,1)
        TEMP(i,1) = (TEMP(i,2:9)*Masses) - 1.0072764;
    end
end
TEMP = sortrows(TEMP, 1);
ALIGNED = zeros(1, size(TEMP,2));
n_aligned = 1;
curr_strt = 1;
test_row = 1;
while test_row < size(TEMP,1);
    test_row = test_row + 1;
    if alignd_cols == 1;
        diff = 1000000 * ( TEMP(test_row,1) - TEMP(curr_strt,1) ) /
TEMP(curr_strt,1);
        if diff > tol
            %adds aligned data to 'aligned'
            ALIGNED (n_aligned, 1) = mean( TEMP (curr_strt:test_row-1, 1) );
            for spec = 1 : nargin-1
                ALIGNED (n_aligned, spec+1 ) = sum ( TEMP (curr_strt:test_row-
1, spec+1) );
            end
            %start new loop
            n_aligned = n_aligned + 1;
            curr_strt = test_row;
            %in_rnge = 1;
        end
    elseif alignd_cols == 9;
        if ~isequal( TEMP (test_row,1:8), TEMP (curr_strt,1:8) )
            %adds aligned data to 'masses' and 'aligned'
            ALIGNED (n_aligned, 1:9) = TEMP (curr_strt, 1:9);
            for spec = 1 : nargin - 1
                ALIGNED (n_aligned, spec+9 ) = sum ( TEMP (curr_strt:test_row-
1, spec+9) );
            end
            %start new loop
            n_aligned = n_aligned + 1;
            curr_strt = test_row;
            %in_rnge = 1;
        end
    end
end
if alignd_cols == 9
    if ~isequal( ALIGNED (n_aligned-1,1:8), TEMP (size(TEMP,1),1:8) )
        ALIGNED (n_aligned,1:7+nargin) = TEMP (size(TEMP,1),1:7+nargin);
    end
end
%Calculates similarity or Venn diagram data
if nargin == 3
    A_num = zeros(3,1);

```

```

D_num = zeros(3,1);
for i = 1 : size(ALIGNED,1)
    if ALIGNED(i,alignd_cols+1) > 0 && ALIGNED(i,alignd_cols+2) == 0
        A_num(1) = A_num(1) + 1;
        if alignd_cols == 9
            if ALIGNED(i,5) > 0
                D_num(1) = D_num(1) + 1;
            end
        end
    elseif ALIGNED(i,alignd_cols+1) > 0 && ALIGNED(i,alignd_cols+2) > 0
        A_num(2) = A_num(2) + 1;
        if alignd_cols == 9
            if ALIGNED(i,5) > 0
                D_num(2) = D_num(2) + 1;
            end
        end
    elseif ALIGNED(i,alignd_cols+1) == 0 && ALIGNED(i,alignd_cols+2) > 0
        A_num(3) = A_num(3) + 1;
        if alignd_cols == 9
            if ALIGNED(i,5) > 0
                D_num(3) = D_num(3) + 1;
            end
        end
    end
end
end
A_per = 100 * A_num / sum(A_num);
D_per = 100 * D_num / sum(D_num);
Subsets = {'A ', 'AB ', 'B '};
TAB_p = table(A_num, A_per, D_num, D_per, 'RowNames', Subsets);

elseif nargin == 4
    A_num = zeros(7,1);
    D_num = zeros(7,1);
    for i = 1 : size(ALIGNED,1)
        if ALIGNED(i,alignd_cols+1) > 0 && ALIGNED(i,alignd_cols+2) == 0 &&
ALIGNED(i,alignd_cols+3) == 0
            A_num(1) = A_num(1) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(1) = D_num(1) + 1;
                end
            end
        elseif ALIGNED(i,alignd_cols+1) > 0 && ALIGNED(i,alignd_cols+2) > 0 &&
ALIGNED(i,alignd_cols+3) == 0
            A_num(2) = A_num(2) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(2) = D_num(2) + 1;
                end
            end
        elseif ALIGNED(i,alignd_cols+1) == 0 && ALIGNED(i,alignd_cols+2) > 0 &&
ALIGNED(i,alignd_cols+3) == 0
            A_num(3) = A_num(3) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(3) = D_num(3) + 1;
                end
            end
        elseif ALIGNED(i,alignd_cols+1) == 0 && ALIGNED(i,alignd_cols+2) > 0 &&
ALIGNED(i,alignd_cols+3) > 0
            A_num(4) = A_num(4) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(4) = D_num(4) + 1;
                end
            end
        end
    end
end

```

```

        end
        elseif ALIGNED(i,alignd_cols+1) == 0 && ALIGNED(i,alignd_cols+2) == 0
&& ALIGNED(i,alignd_cols+3) > 0
            A_num(5) = A_num(5) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(5) = D_num(5) + 1;
                end
            end
        end
        elseif ALIGNED(i,alignd_cols+1) > 0 && ALIGNED(i,alignd_cols+2) == 0 &&
ALIGNED(i,alignd_cols+3) > 0
            A_num(6) = A_num(6) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(6) = D_num(6) + 1;
                end
            end
        end
        elseif ALIGNED(i,alignd_cols+1) > 0 && ALIGNED(i,alignd_cols+2) > 0 &&
ALIGNED(i,alignd_cols+3) > 0
            A_num(7) = A_num(7) + 1;
            if alignd_cols == 9
                if ALIGNED(i,5) > 0
                    D_num(7) = D_num(7) + 1;
                end
            end
        end
    end
end
A_per = 100 * A_num / sum(A_num);
D_per = 100 * D_num / sum(D_num);
Subsets = {'A ', 'AB ', 'B ', 'BC ', 'C ', 'CA ', 'ABC'};
TAB_p = table(A_num, A_per, D_num, D_per, 'RowNames', Subsets);

elseif nargin > 4

    %Creates a presence/absence-based data matrix from which to calculate
    %Bray-Curtis distance matrix between samples.
    BC_data_PA=ALIGNED( :, alignd_cols+1:size(ALIGNED,2) );
    for i=1:size(BC_data_PA,1);
        for j=1:size(BC_data_PA,2);
            if BC_data_PA(i,j)>0
                BC_data_PA(i,j)=1;
            end
        end
    end

    %Calculates Bray-Curtis distance matrix; see note below)
    BC_data_PA = (BC_data_PA)'; % transpose so will operate on
columns
    BC_data_PA = (BC_data_PA + 0.00000001); % prevent divide by zero errors
    [n,p] = size(BC_data_PA);
    D = zeros(n);
    % BRAY-CURTIS SIMILARITY:
    if p>1
        for i = 1:n-1
            x1 = BC_data_PA(i+1:n,:);
            x2 = BC_data_PA(1:n-i,:);
            D(i+1:n+1:n*(n-i))=(1-(sum(abs(x1-x2)')'./sum(abs(x1+x2)'))));
        end
    end
    D
    D = D+D';
    D(1:(n+1):n*n) = ones(n,1);
    TAB_p = (1-D); % change output from similarity to distance (DLJ)

    if nargout == 3

```

```

%stores local (nominal mass) maxima of each column
comp_nm = (floor(ALIGNED(1,1)) : floor(ALIGNED(size(ALIGNED,1),1)))';
comp_nm(1 : size(comp_nm,1), 2 : nargin-1) = zeros(size(comp_nm,1),
nargin-2);
for z = 1 : nargin - 2
    for i = 1 : size(comp_nm)
        for j = 1 : size(ALIGNED,1)
            if ALIGNED(j,alignd_cols+z) > comp_nm(i,z+1) &&
floor(ALIGNED(j,1)) == comp_nm(i,1)
                comp_nm(i,z+1) = ALIGNED(j,alignd_cols+z);
            end
        end
    end
end
end
BC_data_NI=ALIGNED( :, alignd_cols+1:size(ALIGNED,2) );

%normalizes relative intensities to (nonzero) local maxima
for j=1:size(BC_data_NI,2);
    for i=1:size(BC_data_NI,1);
        if BC_data_NI(i,j) > 0
            for z = 1 : size(comp_nm,1)
                if comp_nm(z,1) == floor(ALIGNED(i,1)) &&
comp_nm(z,j+1) > 0
                    BC_data_NI(i,j) = 100 * BC_data_NI(i,j) /
comp_nm(z,j+1);
                end
            end
        end
    end
end
end

%Calculates Bray-Curtis distance matrix; see note below
BC_data_NI = (BC_data_NI)'; % transpose so will operate on
columns
BC_data_NI = (BC_data_NI + 0.00000001); % prevent divide by zero errors
[n,p] = size(BC_data_NI);
D = zeros(n);
% BRAY-CURTIS SIMILARITY:
if p>1
    for i = 1:n-1
        x1 = BC_data_NI(i+1:n,:);
        x2 = BC_data_NI(1:n-i,:);
        D(i+1:n+1:n*(n-i))=(1-(sum(abs(x1-x2))'./sum(abs(x1+x2))')));
    end
end;
D = D+D';
D(1:(n+1):n*n) = ones(n,1);
TAB_ni = (1-D); % change output from similarity to distance
(DLJ)
end

end

%note for lines 228 - 242 & 274 - 288:
%
% =====
% Copyright (c) 1997 B. Planque - Sir Alister Hardy Foundation for Ocean
Science
% bp@wpo.nerc.ac.uk
% Permission is granted to modify and re-distribute this code
% in any manner as long as this notice is preserved.
% All standard disclaimers apply.
% =====
%
% modified by David L. Jones (Feb-2001) after "distance.m" IN "EDAT Toolbox"

```

```

% to only calculate a Bray-Curtis distance matrix between columns
%
% This file is part of the FATHOM Toolbox for Matlab.
%
% modified by Daniel R. Baluha (June 2015) after "f_braycurtis.m" in
%"FATHOM toolbox"
%
%
%=====
% Daniel R. Baluha
% University of Maryland, College Park
% June 2015
% MATLAB version R2013b
% OS X version 10.9.5
%=====

```

A9.8. UV-visible absorption and emission analysis

```

function [ab_corr, em_corr, ex_data] = optics4( FileNames, Parameters)

%Applies the following data analyses/corrections to absorbance (and
%fluorescence) data for a single sample. Raw data are read from the raw
%text files that are exported directly from the UVPC-2401 (and the AB2).

% 1. ABSORBANCE, LOAD & FORMAT
%   % Raw absorbance data from .txt files exported directly from the
%   % UVPC-2401 with filenames specified in 'FileNames' are opened and
%   % converted to n x 2 matrices.

% 2. ABSORBANCE, BLANK SUBTRACTION
%   % Absorbance values in 'ab_corr' are equal to those in the
%   % absorbance spectrum of the sample minus those in the
%   % absorbance spectrum of the blank.

% 3. ABSORBANCE, BASELINE CORRECTION
%   % An average and standard deviation of absorbance is calculated
%   % from the wavelength range specified by 'Parameters(1:2,1)'. All
%   % new absorbance values in 'ab_corr' equal the initial absorbance
%   % values minus the average baseline absorbance. Only performed if
%   % Parameters(1:2,1) are all non-zero.

% 4. ABSORBANCE, SPECTRAL SLOPE CALCULATION
%   % CDOM absorbance spectral slope (S) calculated and displayed using
%   % a non-linear least squares fitting of the equation
%   %
%   %  $ab_{wl} = ab_{ref} * e^{-S (wl - ref)}$ 
%   %
%   %   ref: Wavelength specified in 'Parameters(3,2)'
%   %
%   %   ab_wl & ab_ref: absorbance at a wavelength within the range
%   %   specified by 'Parameters(1:2,2)' and Parameters(3,2)
%   %
%   % Only performed if Paramters(1:3,2) are all non-zero.

% 5. EMISSION, LOAD & FORMAT
%   % Raw emission data from .txt files exported directly from the
%   % AB2 with filenames specified in 'FileNames' are opened and
%   % converted to n x m matrices.

```

```

% 6. EMISSION, BLANK SUBTRACTION
% Emission intensity values in 'em_corr' are equal to those in the
% emission spectrum of the blank subtracted from those in the
% emission spectrum of the sample.

% 7. EMISSION, SMOOTHING
% Individual emission spectra in 'eem_corr' are smoothed using
% a 'smoothing'-point moving average, where 'smoothing' is an odd
% integer specified in the input variable 'Parameters(1,4)'.
% Performed only if 'smoothing' is an odd integer.

% 8. EMISSION, SCATTER PEAK EXCISION
% Replaces intensity data at wavelengths +/- 10 nm of the center of
% a scatter peak with interpolated intensities. Wavelengths of the
% center of the 1st order Raman, 1st order Rayleigh, and 2nd order
% Rayleigh peaks ('scatter_wl') from excitation at wavelength 'ex'
% are determined from the following equations, (respectively):
%
% scatter_wl = 10000000 / ( 10000000/ex - 3400 )
% scatter_wl = ex
% scatter_wl = 2*ex
%
% Peak excision performed only if Parameters(1:3,5) are equal to 1
% for the corresponding scatter peak.

% 9. EMISSION, WAVELENGTH OF MAXIMUM EMISSION
% Calculates wavelength of maximum emission for each excitation
% wavelength, given in 'ex_data'

%10. EMISSION, INNER FILTER EFFECT CORRECTION
% Calculates and applies an inner filter affect correction factor
% for all intensity values based on the corrected absorbance
% values.

%11. EMISSION, CONVERSION TO QUININE SULFATE UNITS
% Converts intensities in 'em_corr' to quinine sulfate units (QSU),
% where 1 QSU = intensity of 1 ppb quinine sulfate at
% excitation/emission wavelengths 350/450 nm.
%

%12. EMISSION, QUANTUM YIELD DETERMINATION

%INPUT ARGUMENTS

% 'FileNames' 2x1, 4x1, 6x1, or 8x1 cell with the following data:
% { ab_sample; ab_blank; ( ) }
% { em_sample; em_blank; ( ) }
% { em_qs; ab_qs; ( ) }
% { em_h2so4; ab_h2so4 }

% 'Parameters' 3x6 double (matrix) with the following data:
% [sc_min, bl_min, ex_min, smoothing, ram_1, c_smp_e/c_smp_a;
% sc_max, bl_max, ex_max, pife_corr, ray_1, c_qs_a;
% sc_ref, em_norm, ex_int, sife_corr, ray_2, c_qs_e ]

%OUTPUT ARGUMENTS

% 'ab_corr' Corrected absorbance data
%
% 'em_corr' Corrected emission data
%
% 'ex_data' n x 2 matrix containing the following:
%

```

```

%      column 1: excitation wavelengths
%      column 2: wavelength of maximum emission
%      column 3: fluorescence quantum yield

%OUTPUT TEXT

% baseline avg: average absorbance of the specified baseline region before
% correction

% baseline std: Standard deviation of the absorbance of the specified
% baseline region before correction.

% spectral slope (S): Spectral slope calculated by non-linear least squares
% fitting

% residual: defined by SUM(exp - thr)^2 for fitted data, where 'exp' and
% 'thr' are the experimental and theoretical absorbance values,
% respectively.

%ABSORBANCE DILUTION FACTOR
if Parameters(1,6) > 0
    df_smp = Parameters(1,6);
else
    df_smp = 1;
end

%ABSORBANCE, LOAD & FORMAT
fileID = fopen( char( FileNames(1,1) ),'r' );
opened_smp = textscan( fileID,'%s','delimiter','\r' );
for i = 1:length( opened_smp{1} )
    row = str2num( char( opened_smp{1}{i} ) );
    if length(row)==2
        ab_raw_smp(i,:) = row;
    end
end
fileID = fopen( char( FileNames(2,1) ),'r' );
opened_blk = textscan( fileID,'%s','delimiter','\r' );
for i = 1:length( opened_blk{1} )
    row = str2num( char( opened_blk{1}{i} ) );
    if length(row)==2
        ab_raw_blk(i,:) = row;
    end
end

%ABSORBANCE, BLANK SUBTRACTION
ab_corr (:,1) = ab_raw_smp(:,1);
ab_corr (:,2) = ab_raw_smp(:,2) - ab_raw_blk(:,2);

%ABSORBANCE, BASELINE CORRECTION
bl_min = Parameters(1,2);
bl_max = Parameters(2,2);
if bl_min > 0 && bl_max > 0
    n=0;
    for i=1:size(ab_corr)
        if ab_corr(i,1) >= bl_min && ab_corr(i,1) <= bl_max
            n=n+1;
            base(n,1)=ab_corr(i,2);
        end
    end
    disp(['baseline avg = ',num2str(mean(base),'%10.4f\n' )]);

```

```

disp(['baseline std = ',num2str(std(base),'%10.4f\n' )]);
for i = 1:size(ab_corr,1)
    ab_corr(i,2) = ab_corr(i,2) - mean(base);
end
end

%ABSORBANCE, SPECTRAL SLOPE CALCULATION
sc_min = Parameters(1,1);
sc_max = Parameters(2,1);
sc_ref = Parameters(3,1);
if sc_min > 0 && sc_max > 0 && sc_ref > 0
    n=0;
    for i=1:size(ab_corr,1)
        if ab_corr(i,1) >= sc_min && ab_corr(i,1) <= sc_max
            n=n+1;
            xdata(n,1) = ab_corr(i,1);
            ydata(n,1) = ab_corr(i,2);
        end
        if ab_corr(i,1)==sc_ref
            ab_ref=ab_corr(i,2);
        end
    end
    fun=@(s,xdata) ab_ref.*exp(-s(1).*(xdata(:,1)-sc_ref));
    [s_slope, s_residual]=lsqcurvefit(fun,0.0001,xdata,ydata);
    disp(['spectral slope (S) = ',num2str(s_slope)]);
    disp(['residual = ',num2str(s_residual)]);
end

ex_min = Parameters(1,3);
ex_max = Parameters(2,3);
ex_int = Parameters(3,3);
if length(FileNames) >= 4 && ex_min > 0 && ex_max > 0 && ex_int > 0

    %EMISSION, LOAD & FORMAT
    %Load sample
    fileID = fopen( char( FileNames(3,1) ),'r' );
    opened = textscan( fileID,'%s','delimiter','\r' );
    for i=1:length( opened{1} )
        row=str2num( char( opened{1}{i} ) );
        if length(row)==2
            em_raw_smp(i,:) = row;
        end
    end
    em_raw_smp( all(~em_raw_smp,2), : ) = [];
    %Load blank
    fileID = fopen( char( FileNames(4,1) ),'r' );
    opened = textscan( fileID,'%s','delimiter','\r' );
    for i=1:length( opened{1} )
        row=str2num( char( opened{1}{i} ) );
        if length(row)==2
            em_raw_blk(i,:) = row;
        end
    end
    em_raw_blk( all(~em_raw_blk,2), : ) = [];
    %Pre-allocate excitation/emission wavelengths
    em_min = min(em_raw_smp(:,1));
    em_max = max(em_raw_smp(:,1));
    em_int = em_raw_smp(2,1) - em_raw_smp(1,1);
    EMs = em_min : em_int : em_max;
    EXs = ex_min : ex_int : ex_max;
    em_smp(2:length(EMs)+1, 1) = EMs;
    em_smp(1, 2:length(EXs)+1) = EXs;
    em_blk = em_smp;

```

```

%format sample
i = 1;
j = 1;
last = em_max;
for n = 1:size(em_raw_smp,1)
    if last > em_raw_smp(n,1)
        j = j + 1;
        i = 2;
    else
        i = i + 1;
    end
    if em_raw_smp(n,1)==em_smp(i,1);
        em_smp(i,j) = em_raw_smp(n,2);
    end
    last = em_raw_smp(n,1);
end
%format blank
i = 1;
j = 1;
last = em_max;
for n = 1:size(em_raw_blk,1)
    if last > em_raw_blk(n,1)
        j = j + 1;
        i = 2;
    else
        i = i + 1;
    end
    if em_raw_blk(n,1)==em_blk(i,1);
        em_blk(i,j) = em_raw_blk(n,2);
    end
    last = em_raw_blk(n,1);
end

%EMISSION, BLANK SUBTRACTION
em_corr = em_smp;
X = size(em_corr,1);
Y = size(em_corr,2);
em_corr(2:X,2:Y) = em_corr(2:X,2:Y) - em_blk(2:X,2:Y);

%EMISSION, SMOOTHING
if Parameters(1,4)>0 && mod(Parameters(1,4),2)==1
    smoothing = Parameters(1,4);
    for i = 1.5+(smoothing/2) : X - floor(smoothing/2)
        for j = 2 : Y
            i_start = i - floor(smoothing/2);
            i_end = i + floor(smoothing/2);
            em_corr(i,j) = sum( em_corr(i_start:i_end,j) ) / smoothing;
        end
    end
end

%EMISSION, SCATTER PEAK EXCISION (Raman peak)
if Parameters(1,5)==1
    for j = 2 : Y;
        curr_raw = em_corr(2:X,1);
        curr_raw(:,2) = em_corr(2:X,j);
        %center wavelength of the scatter peak
        scatter_wl = 10000000 / ( 10000000/em_corr(1,j) - 3400 );
        %indices of wavelengths +/- 10 nm of the center of the scatter peak
        [~, lower_n] = min( abs( (scatter_wl - 10) - curr_raw(:,1) ) );
        [~, upper_n] = min( abs( (scatter_wl + 10) - curr_raw(:,1) ) );
        %wavelength/intensity data outside of scatter peak range
    end
end

```

```

curr_ref = curr_raw;
curr_ref(lower_n : upper_n, :) = [];
%new emission wavelength/intensity data
curr_new = curr_raw(:,1);
curr_new(:,2) = interp1 ( curr_ref(:,1), curr_ref(:,2),
curr_new(:,1), 'spline' );
curr_new(isnan(curr_new))=0;
%replaces input em_corr with new (excised) em_corr data
em_corr(2:X,j)=curr_new(:,2);

end
end

%EMISSION, SCATTER PEAK EXCISION (1st-order Rayleigh peak)
if Parameters(2,5)==1
for j = 2 : Y;
curr_raw = em_corr(2:X,1);
curr_raw(:,2) = em_corr(2:X,j);
%center wavelength of the scatter peak
scatter_wl = em_corr(1,j) ;
%indices of wavelengths +/- 10 nm of the center of the scatter peak
[~, lower_n] = min( abs( (scatter_wl - 10) - curr_raw(:,1) ) );
[~, upper_n] = min( abs( (scatter_wl + 10) - curr_raw(:,1) ) );
%wavelength/intensity data outside of scatter peak range
curr_ref = curr_raw;
curr_ref(lower_n : upper_n, :) = [];
%new emission wavelength/intensity data
curr_new = curr_raw(:,1);
curr_new(:,2) = interp1 ( curr_ref(:,1), curr_ref(:,2),
curr_new(:,1), 'spline' );
curr_new(isnan(curr_new))=0;
%replaces input em_corr with new (excised) em_corr data
em_corr(2:size(em_corr,1),j)=curr_new(:,2);
end
end

%EMISSION, SCATTER PEAK EXCISION (2nd-order Rayleigh peak)
if Parameters(2,5)==1
for j = 2 : Y;
curr_raw = em_corr(2:X,1);
curr_raw(:,2) = em_corr(2:X,j);
%center wavelength of the scatter peak
scatter_wl = 2*em_corr(1,j) ;
%indices of wavelengths +/- 10 nm of the center of the scatter peak
[~, lower_n] = min( abs( (scatter_wl - 10) - curr_raw(:,1) ) );
[~, upper_n] = min( abs( (scatter_wl + 10) - curr_raw(:,1) ) );
%wavelength/intensity data outside of scatter peak range
curr_ref = curr_raw;
curr_ref(lower_n : upper_n, :) = [];
%new emission wavelength/intensity data
curr_new = curr_raw(:,1);
curr_new(:,2) = interp1 ( curr_ref(:,1), curr_ref(:,2),
curr_new(:,1), 'spline' );
curr_new(isnan(curr_new))=0;
%replaces input em_corr with new (excised) em_corr data
em_corr(2:X,j)=curr_new(:,2);
end
end

%EMISSION, WAVELENGTH OF MAXIMUM EMISSION
ex_data=zeros(Y-1,1);
for j = 2 : size(em_corr,2);

```

```

    ex_data(j-1,1)=em_corr(1,j);
    [~,idx] = max( em_corr( 2:X ,j ) );
    ex_data(j-1,2)=em_corr(idx+1,1);
end

%EMISSION, INNER FILTER EFFECT CORRECTION
ifec = ones(X,Y);
%Determines excitation wavelength (primary) correction factor
if Parameters(2,4)
    for z=1:size(ab_corr,1)
        for j=2:Y
            if ab_corr(z,1)==em_corr(1,j);
                ifec(2:X,j)=ifec(2:X,j)*3.1623^-(ab_corr(z,2)/df_smp);
            end
        end
    end
end
%Determines emission wavelength (secondary) correction factor
if Parameters(3,4)
    for z=1:size(ab_corr,1)
        for i=2:X;
            if ab_corr(z,1)==em_corr(i,1);
                ifec(i,2:Y)=ifec(i,2:Y)*3.1623^-(ab_corr(z,2)/df_smp);
            end
        end
    end
end
%Applies inner filter correction
for i=1:X
    for j=1:Y
        em_corr(i,j) = em_corr(i,j) / ifec (i,j);
    end
end

end

if length(FileNames) >= 5 && Parameters(2,6)>0 && Parameters(3,6)>0

    %Load em_qs
    fileID = fopen( char( FileNames(5,1) ),'r' );
    opened = textscan( fileID,'%s','delimiter','\r' );
    for i=1:length( opened{1} )
        row=str2num( char( opened{1}{i} ) );
        if length(row)==2
            em_qs(i,:) = row;
        end
    end
    em_qs( all(~em_qs,2), : ) = [];
    %concentration of quinine sulfate (in ppb) for emission measurements
    c_qs_e = Parameters(3,6);

    %processing and blank subtraction of QS emission spectrum
    if length(FileNames) >= 7
        %Loads em_h2so4 (if specified)
        fileID = fopen( char( FileNames(7,1) ),'r' );
        opened = textscan( fileID,'%s','delimiter','\r' );
        for i=1:length( opened{1} )
            row=str2num( char( opened{1}{i} ) );
            if length(row)==2
                em_qs_blk(i,:) = row;
            end
        end
    else

```

```

%Uses em_blk at ex=350 nm for blank subtraction
for j=2:Y
    if em_blk(1,j) == 350
        em_qs_blk_pre = [em_blk(2:X,1) em_blk(2:X,j)];
    end
end
n = 0;
for i = 1:size(em_qs_blk_pre,1)
    for z = 1:size(em_qs,1)
        if em_qs_blk_pre(i,1) == em_qs(z,1)
            n = n + 1;
            em_qs_blk(n,1:2)=em_qs_blk_pre(i,1:2);
        end
    end
end
end
em_qs(:,2) = em_qs(:,2) - em_qs_blk(:,2);

%smoothing of QS emission spectrum
if Parameters(1,4)>0 && mod(Parameters(1,4),2)==1
    smoothing = Parameters(1,4);
    for i = 1.5+(smoothing/2) : size(em_qs,1) - floor(smoothing/2)
        i_start = i - floor(smoothing/2);
        i_end = i + floor(smoothing/2);
        em_qs(i,2) = sum( em_qs(i_start:i_end,2) ) / smoothing;
    end
end

%EMISSION, CONVERSION TO QUININE SULFATE UNITS
if Parameters(3,2) == 1
    %gets emission of quinine sulfate at 450 nm
    [~, n_450] = min( abs( 450 - em_qs(:,1) ) );
    em_qs_450 = em_qs(n_450,2); %emission of 'c_qs_e' ppb QS
    em_corr(2:X,2:Y) = em_corr(2:X,2:Y) * (c_qs_e / em_qs_450);
    em_qs(:,2) = em_qs(:,2) * (c_qs_e / em_qs_450);
end

%Raman peak excision and interpolation (EM = 397 +/- 10 nm)
if Parameters(1,5)==1
    curr_raw = em_qs(:,1);
    curr_raw(:,2) = em_qs(:,2);
    %indices of wavelengths +/- 10 nm of the center of the scatter peak
    [~, lower_n] = min( abs( 387 - curr_raw(:,1) ) );
    [~, upper_n] = min( abs( 407 - curr_raw(:,1) ) );
    %wavelength/intensity data outside of scatter peak range
    curr_ref = curr_raw;
    curr_ref(lower_n : upper_n, :) = [];
    %new emission wavelength/intensity data
    curr_new = curr_raw(:,1);
    curr_new(:,2) = interp1 ( curr_ref(:,1), curr_ref(:,2),
curr_new(:,1), 'spline' );
    curr_new(isnan(curr_new))=0;
    %replaces input em_corr with new (excised) em_corr data
    em_qs(:,2)=curr_new(:,2);
end

%DETERMINATION OF QUANTUM YIELD
if length(FileNames) >= 6 && Parameters(3,6)~=0
    c_qs_a = Parameters(2,6);
    %Load ab_qs
    fileID = fopen( char( FileNames(6,1) ), 'r' );
    opened = textscan( fileID, '%s', 'delimiter', '\r' );
    for i=1:length( opened{1} )

```

```

row=str2num( char( opened{1}{i} ) );
if length(row)==2
    ab_qs(i,:) = row;
end
end
%blank subtract ab_qs
if length(FileNames) >= 7
    %Load and subtract ab_h2so4 from ab_qs
    fileID = fopen( char( FileNames(6,1) ), 'r' );
    opened = textscan( fileID, '%s', 'delimiter', '\r' );
    for i=1:length( opened{1} )
        row=str2num( char( opened{1}{i} ) );
        if length(row)==2
            ab_h2so4(i,:) = row;
        end
    end
    ab_qs(:,2) = ab_qs(:,2) - ab_h2so4(:,2);
else
    ab_qs(:,2) = ab_qs(:,2) - ab_raw_blk(:,2);
end
%sum_I_qs & A_qs
sum_I_qs = sum(em_qs(:,2));
[~, n_350] = min( abs( 350 - ab_qs(:,1) ) );
A_qs = ab_qs(n_350,2);
%for each excitation wavelength
for j = 2:Y
    sum_I_smp = sum(em_corr(2:X,j));
    [~, n_ex] = min( abs( em_corr(1,j) - ab_corr(:,1) ) );
    A_smp = ab_corr(n_ex,2);
    I_s_dv_q = sum_I_smp / sum_I_qs;
    A_q_dv_s = A_qs / A_smp;
    cs_a_dv_e = 1 / df_smp;
    cq_e_dv_a = c_qs_e / c_qs_a;
    ex_data(j-1,3) = 0.51 * I_s_dv_q * A_q_dv_s * cs_a_dv_e *
cq_e_dv_a;
end
end
end
end

```

A9.9. Calculations of average molecular weight, H/C and O/C molar ratios, and mass percentages.

Number-averaged molecular weight (AMW_N) and weight-averaged molecular weight (AMW_W), and polydispersity (PD) were calculated for a mass spectral peak list of x singly-charged species as follows:

$$(1) \quad AMW_N = \frac{\sum_{i=1}^x M_i I_i}{\sum_{i=1}^x I_i}$$

$$(2) \quad AMW_W = \frac{\sum_{i=1}^x M_i^2 I_i}{\sum_{i=1}^x M_i I_i}$$

$$(3) \quad PD = \frac{AMW_W}{AMW_N}$$

where M_i and I_i are the m/z and intensity of the i^{th} species in the mass spectral peak list.

Relative intensity-weighted average O/C and H/C molar ratios (O/C_{avg} and H/C_{avg}) were calculated for a list of x peaks with assigned molecular formulae as follows:

$$(4) \quad O/C_{avg} = \frac{1}{x} \sum_{i=1}^x RI_i \frac{O_i}{C_i}$$

$$(5) \quad H/C_{avg} = \frac{1}{x} \sum_{i=1}^x RI_i \frac{H_i}{C_i}$$

Where C_i , H_i , and O_i , are the numbers of carbon, hydrogen, and oxygen in the i^{th} molecular formula in the mass spectral peak list, and RI_i is the relative intensity of the corresponding peak. Numbers of carbon and hydrogen are calculated from all isotopes (i.e. ^{12}C , ^{13}C , ^1H , and ^2H).

The mass percent of each element X was calculated as follows:

$$(6) \quad \%X_{MP} = \sum_{i=1} \frac{100 n_{x,i} MW_x RI_i}{(m/z_i + 1.007) TIC}$$

Where MW_x is the molecular mass of atom x , TIC is the total ion intensity of all peaks with molecular formulae, and $n_{x,i}$, RI_i , and m/z_i are the numbers of x atoms, relative intensity, m/z of the i^{th} peak to which a molecular formula was assigned.

Bibliography

- (1) Hedges, J. I. Global Biogeochemical Cycles: Progress and Problems. *Mar. Chem.* **1992**, *39*, 67–93.
- (2) Falkowski, P.; Scholes, R. J.; Boyle, E.; Canadell, J.; Canfield, D.; Elser, J.; Gruber, N.; Hibbard, K.; Högberg, P.; Linder, S.; et al. The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science (80-.)*. **2000**, *290*, 291–296.
- (3) Andrew, A. a.; Del Vecchio, R.; Subramaniam, A.; Blough, N. V. Chromophoric Dissolved Organic Matter (CDOM) in the Equatorial Atlantic Ocean: Optical Properties and Their Relation to CDOM Structure and Source. *Mar. Chem.* **2013**, *148*, 33–43.
- (4) Moran, M. A.; Sheldon, W. M.; Zepp, R. G. Carbon Loss and Optical Property Changes during Long-Term Photochemical and Biological Degradation of Estuarine Dissolved Organic Matter. *Limnol. Ocean.* **2000**, *45*, 1254–1264.
- (5) Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. Illuminated Darkness: Molecular Signatures of Congo River Dissolved Organic Matter and Its Photochemical Alteration as Revealed by Ultrahigh Precision Mass Spectrometry. *Limnol. Ocean.* **2010**, *55*, 1467–1477.

- (6) Repeta, D. J.; Quan, T. M.; Aluwihare, L. I.; Accardi, A. Chemical Characterization of High Molecular Weight Dissolved Organic Matter in Fresh and Marine Waters. *Geochim. Cosmochim. Acta* **2002**, *66*, 955–962.
- (7) Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a Major Refractory Component of Marine Dissolved Organic Matter. *Geochim. Cosmochim. Acta* **2006**, *70*, 2990–3010.
- (8) Salonen, K.; Hammar, T. On the Importance of Dissolved Organic Matter in the Nutrition of Zooplankton in Some Lake Waters. *Oecologia* **1986**, *68*, 246–253.
- (9) Kujawinski, E. B.; Longnecker, K.; Blough, N. V.; Vecchio, R. Del; Finlay, L.; Kitner, J. B.; Giovannoni, S. J. Identification of Possible Source Markers in Marine Dissolved Organic Matter Using Ultrahigh Resolution Mass Spectrometry. *Geochim. Cosmochim. Acta* **2009**, *73*, 4384–4399.
- (10) Hedges, J. I. Why Dissolved Organics Matter. In *Biogeochemistry of Marine Dissolved Organic Matter*; Hansell, D. A.; Carlson, C. A., Eds.; Academic Press: San Diego, CA, 2002; pp. 1–27.
- (11) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A Simple and Efficient Method for the Solid-Phase Extraction of Dissolved Organic Matter (SPE-DOM) from Seawater. *Limnol. Ocean. Methods* **2008**, *6*, 230–235.

- (12) Sutton, R.; Sposito, G. Molecular Structure in Soil Humic Substances: The New View. *Environ. Sci. Technol.* **2005**, *39*, 9009–9015.
- (13) Nebbioso, A.; Piccolo, A. Basis of a Humeomics Science: Chemical Fractionation and Molecular Characterization of Humic Biosuprastructures. *Biomacromolecules* **2011**, *12*, 1187–1199.
- (14) Ma, X.; Green, S. a. Fractionation and Spectroscopic Properties of Fulvic Acid and Its Extract. *Chemosphere* **2008**, *72*, 1425–1434.
- (15) Conte, P.; Piccolo, A. Conformational Arrangement of Dissolved Humic Substances. Influence of Solution Composition on Association of Humic Molecules. *Environ. Sci. Technol.* **1999**, *33*, 1682–1690.
- (16) Bricaud, A.; Morel, A.; Prieur, L. Absorption by Dissolved Organic Matter of the Sea (yellow Substance) in the UV and Visible Domains. *Limnol. Ocean.* **1981**, *26*, 43–53.
- (17) Anderson, S.; Zepp, R.; Machula, J.; Santavy, D.; Hansen, L.; Mueller, E. Indicators of UV Exposure in Corals and Their Relevance to Global Climate Change and Coral Bleaching. *Hum. Ecol. Risk Assess.* **2001**, *7*, 1271–1282.
- (18) Chin, Y.-P.; Miller, P. L.; Zeng, L.; Cawley, K.; Weavers, L. K. Photosensitized Degradation of Bisphenol A by Dissolved Organic Matter. *Environ. Sci. Technol.* **2004**, *38*, 5888–5894.

- (19) Golanoski, K. S.; Fang, S.; Del Vecchio, R.; Blough, N. V. Investigating the Mechanism of Phenol Photooxidation by Humic Substances. *Environ. Sci. Technol.* **2012**, *46*, 3912–3920.
- (20) Pos, W. H.; Riemer, D. D.; Zika, R. G. Carbonyl Sulfide (OCS) and Carbon Monoxide (CO) in Natural Waters : Evidence of a Coupled Production Pathway. *Mar. Chem.* **1998**, *62*, 89–101.
- (21) Zhang, Y.; Del Vecchio, R.; Blough, N. V. Investigating the Mechanism of Hydrogen Peroxide Photoproduction by Humic Substances. *Environ. Sci. Technol.* **2012**, *46*, 11836–11843.
- (22) Del Vecchio, R.; Blough, N. V. On the Origin of the Optical Properties of Humic Substances. *Environ. Sci. Technol.* **2004**, *38*, 3885–3891.
- (23) Chin, Y.-P.; Aiken, G.; O’Loughlin, E. Molecular Weight, Polydispersity, and Spectroscopic Properties of Aquatic Humic Substances. *Environ. Sci. Technol.* **1994**, *28*, 1853–1858.
- (24) Helms, J. R.; Stubbins, A.; Ritchie, J. D.; Minor, E. C.; Kieber, D. J.; Mopper, K. Absorption Spectral Slopes and Slope Ratios as Indicators of Molecular Weight, Source, and Photobleaching of Chromophoric Dissolved Organic Matter. *Limnol. Ocean.* **2008**, *53*, 955–969.

- (25) Del Vecchio, R.; Blough, N. V. Photobleaching of Chromophoric Dissolved Organic Matter in Natural Waters: Kinetics and Modeling. *Mar. Chem.* **2002**, *78*, 231–253.
- (26) 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. *Limnol. Ocean.* **2001**, *46*, 38–48.
- (27) Zepp, R. G.; Sheldon, W. M.; Moran, M. A. Dissolved Organic Fluorophores in Southeastern US Coastal Waters: Correction Method for Eliminating Rayleigh and Raman Scattering Peaks in Excitation–emission Matrices. *Mar. Chem.* **2004**, *89*, 15–36.
- (28) Green, S. a; Blough, N. V. Optical Absorption and Fluorescence of Chromophoric Dissolved Organic Matter in Natural Waters. *Limnol. Ocean.* **1994**, *39*, 1903–1916.
- (29) Coble, P. G. Characterization of Marine and Terrestrial DOM in Seawater Using Excitation-Emission Matrix Spectroscopy. *Mar. Chem.* **1996**, *51*, 325–346.
- (30) D’Andrilli, J.; Foreman, C. M.; Marshall, A. G.; McKnight, D. M. Characterization of IHSS Pony Lake Fulvic Acid Dissolved Organic Matter by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Fluorescence Spectroscopy. *Org. Geochem.* **2013**, *65*, 19–28.

- (31) Coble, P. G.; Del Castillo, C. E.; Avril, B. Distribution and Optical Properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. *Deep. Res., Part II* **1998**, *45*, 2195–2223.
- (32) Boyle, E. S.; Guerriero, N.; Thiallet, A.; Vecchio, R. Del; Blough, N. V. Optical Properties of Humic Substances and CDOM: Relation to Structure. *Environ. Sci. Technol.* **2009**, *43*, 2262–2268.
- (33) Furman, G. S.; Lonsky, W. F. W. Charge-Transfer Complexes in Kraft Lignin. Part 2: Contribution to Color. *J. Wood Chem. Technol.* **1988**, 191–208.
- (34) Power, J. F.; Langford, C. H. Optical Absorbance of Dissolved Organic Matter in Natural Water Studies Using the Thermal Lens Effect. *Anal. Chem.* **1988**, *60*, 842–846.
- (35) Ziechmann, W. The Electron Donor- and Acceptor Properties of Humic Substances. *Geoderma* **1972**, *8*, 111–131.
- (36) Tossell, J. A. Quinone–hydroquinone Complexes as Model Components of Humic Acids: Theoretical Studies of Their Structure, Stability and Visible–UV Spectra. *Geochim. Cosmochim. Acta* **2009**, *73*, 2023–2033.
- (37) Ma, J.; Del Vecchio, R.; Golanoski, K. S.; Boyle, E. S.; Blough, N. V. Optical Properties of Humic Substances and CDOM: Effects of Borohydride Reduction. *Environ. Sci. Technol.* **2010**, *44*, 5395–5402.

- (38) Phillips, S. M.; Smith, G. D. Light Absorption by Charge Transfer Complexes in Brown Carbon Aerosols. *Environ. Sci. Technol. Lett* **2014**, *1*, 382–386.
- (39) Mopper, K.; Stubbins, A.; Ritchie, J. D.; Bialk, H. M.; Hatcher, P. G. Advanced Instrumental Approaches for Characterization of Marine Dissolved Organic Matter: Extraction Techniques, Mass Spectrometry, and Nuclear Magnetic Resonance Spectroscopy. *Chem. Rev.* **2007**, *107*, 419–442.
- (40) Abdulla, H. A. N.; Minor, E. C.; Dias, R. F.; Hatcher, P. G. Changes in the Compound Classes of Dissolved Organic Matter along an Estuarine Transect: A Study Using FTIR and ¹³C NMR. *Geochim. Cosmochim. Acta* **2010**, *74*, 3815–3838.
- (41) Zafiriou, O. C.; Jousot-Dubien, J.; Zepp, R. G.; Zika, R. G. Photochemistry of Natural Waters. *Environ. Sci. Technol.* **1984**, *18*, 358A – 371A.
- (42) Kostianen, R.; Kauppila, T. J. Effect of Eluent on the Ionization Process in Liquid Chromatography-Mass Spectrometry. *J. Chromatogr. A* **2009**, *1216*, 685–699.
- (43) Kido Soule, M. C.; Longnecker, K.; Giovannoni, S. J.; Kujawinski, E. B. Impact of Instrument and Experiment Parameters on Reproducibility of Ultrahigh Resolution ESI FT-ICR Mass Spectra of Natural Organic Matter. *Org. Geochem.* **2010**, *41*, 725–733.

- (44) Koch, B. P.; Ludwichowski, K.-U.; Kattner, G.; Dittmar, T.; Witt, M. Advanced Characterization of Marine Dissolved Organic Matter by Combining Reversed-Phase Liquid Chromatography and FT-ICR-MS. *Mar. Chem.* **2008**, *111*, 233–241.
- (45) 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 Res.* **2008**, *42*, 63–72.
- (46) Piccolo, A.; Spiteller, M. Electrospray Ionization Mass Spectrometry of Terrestrial Humic Substances and Their Size Fractions. *Anal. Bioanal. Chem.* **2003**, *377*, 1047–1059.
- (47) Koch, B. P.; Witt, M.; Engbrodt, R.; Dittmar, T.; Kattner, G. Molecular Formulae of Marine and Terrigenous Dissolved Organic Matter Detected by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Geochim. Cosmochim. Acta* **2005**, *69*, 3299–3308.
- (48) Rostad, C. E.; Leenheer, J. A. Factors That Affect Molecular Weight Distribution of Suwannee River Fulvic Acid as Determined by Electrospray Ionization/mass Spectrometry. *Anal. Chim. Acta* **2004**, *523*, 269–278.
- (49) Leenheer, J. A.; Rostad, C. E.; Gates, P. M.; Furlong, E. T.; Ferrer, I. Molecular Resolution and Fragmentation of Fulvic Acid by Electrospray Ionization/Multistage Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 1461–1471.

- (50) Aiken, G. R.; Mcknight, I. D. M.; Thorn, I. K. A.; Thurman, E. M. Isolation of Hydrophilic Organic Acids from Water Using Nonionic Macroporous Resins. *Org. Geochem.* **1992**, *18*, 567–573.
- (51) Leenheer, J. A.; Wershaw, R. L.; Reddy, M. M. Strong-Acid, Carboxyl-Group Structures in Fulvic Acid from the Suwannee River, Georgia. 1. Minor Structures. *Environ. Sci. Technol.* **1995**, *29*, 393–398.
- (52) Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E. M. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Anal. Chem* **2008**, *80*, 8908–8919.
- (53) Balogh, M. P. Debating Resolution and Mass Accuracy. *LC GC Eur.* **2004**, *17*, 152–159.
- (54) Stabenau, E. R.; Zika, R. G. Correlation of the Absorption Coefficient with a Reduction in Mean Mass for Dissolved Organic Matter in Southwest Florida River Plumes. *Mar. Chem.* **2004**, *89*, 55–67.
- (55) Reemtsma, T.; These, A. Comparative Investigation of Low-Molecular-Weight Fulvic Acids of Different Origin by SEC-Q-TOF-Ms: New Insights into Structure and Formation. *Environ. Sci. Technol.* **2005**, *39*, 3507–3512.
- (56) Plancque, G.; Amekraz, B.; Moulin, V.; Toulhoat, P.; Moulin, C. Molecular Structure of Fulvic Acids by Electrospray with Quadrupole Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 827–835.

- (57) Zhang, X.; Talley, J. W.; Boggess, B.; Ding, G.; Birdsell, D. Fast Selective Detection of Polar Brominated Disinfection Byproducts in Drinking Water Using Precursor Ion Scans. *Environ. Sci. Technol.* **2008**, *42*, 6598–6603.
- (58) Sleighter, R. L.; Chen, H.; Wozniak, A. S.; Willoughby, A. S.; Caricasole, P.; Hatcher, P. G. Establishing a Measure of Reproducibility of Ultrahigh-Resolution Mass Spectra for Complex Mixtures of Natural Organic Matter. *Anal. Chem.* **2012**, *84*, 9184–9191.
- (59) Kujawinski, E. B.; Freitas, M. a; Zang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. The Application of Electrospray Ionization Mass Spectrometry (ESI MS) to the Structural Characterization of Natural Organic Matter. *Org. Geochem.* **2002**, *33*, 171–180.
- (60) Comisarow, M. B.; Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974**, *25*, 282–283.
- (61) Marshall, A. G.; Hendrickson, C. L. Fourier Transform Ion Cyclotron Resonance Detection: Principles and Experimental Configurations. *Int. J. Mass Spectrom.* **2002**, *215*, 59–75.
- (62) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.

- (63) Hughey, C. a.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001**, *73*, 4676–4681.
- (64) Amster, I. J. Fourier Transform Mass Spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1325–1337.
- (65) Marshall, A. G.; Guan, S. Advantages of High Magnetic Field for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1819–1823.
- (66) Bresson, J. a; Anderson, G. a; Bruce, J. E.; Smith, R. D. For High Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectra via Time-Domain Data Extraction. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 799–804.
- (67) Sleighter, R. L.; Hatcher, P. G. Molecular Characterization of Dissolved Organic Matter (DOM) along a River to Ocean Transect of the Lower Chesapeake Bay by Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Mar. Chem.* **2008**, *110*, 140–152.
- (68) Peuravuori, J.; Pihlaja, K. Molecular Size Distribution and Spectroscopic Properties of Aquatic Humic Substances. *Anal. Chim. Acta* **1997**, *337*, 133–149.
- (69) Thurman, E. M.; Wershaw, R. L.; Malcolm, R. L.; Pinckney, D. J. Molecular Size of Aquatic Humic Substances. *Org. Geochem. Geochem* **1982**, *4*, 27–35.

- (70) Stenson, A. C.; Landing, W. M.; Marshall, A. G.; Cooper, W. T. Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 4397–4409.
- (71) Kujawinski, E. B.; Hatcher, P. G.; Freitas, M. A. High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Humic and Fulvic Acids: Improvements and Comparisons. *Anal. Chem.* **2002**, *74*, 413–419.
- (72) 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*, 554–560.
- (73) Sleighter, R. L.; McKee, G. a.; Liu, Z.; Hatcher, P. G. Naturally Present Fatty Acids as Internal Calibrants for Fourier Transform Mass Spectra of Dissolved Organic Matter. *Limnol. Ocean. Methods* **2008**, *6*, 246–253.
- (74) Cortés-Francisco, N.; Flores, C.; Moyano, E.; Caixach, J. Accurate Mass Measurements and Ultrahigh-Resolution: Evaluation of Different Mass Spectrometers for Daily Routine Analysis of Small Molecules in Negative Electrospray Ionization Mode. *Anal. Bioanal. Chem.* **2011**, *400*, 3595–3606.
- (75) Olsen, J. V.; de Godoy, L. M. F.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-Trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010–2021.

- (76) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method. *Anal. Chem.* **2007**, *79*, 4595–4602.
- (77) Koch, B. P.; Dittmar, T.; Witt, M.; Kattner, G. Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Anal. Chem.* **2007**, *79*, 1758–1763.
- (78) Gonsior, M.; Zwartjes, M.; Cooper, W. J.; Song, W.; Ishida, K. P.; Tseng, L. Y.; Jeung, M. K.; Rosso, D.; Hertkorn, N.; Schmitt-Kopplin, P. Molecular Characterization of Effluent Organic Matter Identified by Ultrahigh Resolution Mass Spectrometry. *Water Res.* **2011**, *45*, 2943–2953.
- (79) Yekta, S. S.; Gonsior, M.; Schmitt-Kopplin, P.; Svensson, B. H. Characterization of Dissolved Organic Matter in Full Scale Continuous Stirred Tank Biogas Reactors Using Ultrahigh Resolution Mass Spectrometry: A Qualitative Overview. *Environ. Sci. Technol.* **2012**, *46*, 12711–12719.
- (80) Herzsprung, P.; Hertkorn, N.; von Tümpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Understanding Molecular Formula Assignment of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Data of Natural Organic Matter from a Chemical Point of View. *Anal. Bioanal. Chem.* **2014**.

- (81) Kujawinski, E. B.; Behn, M. D. Automated Analysis of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra of Natural Organic Matter. *Anal. Chem.* **2006**, *78*, 4363–4373.
- (82) Minor, E. C.; Steinbring, C. J.; Longnecker, K.; Kujawinski, E. B. Characterization of Dissolved Organic Matter in Lake Superior and Its Watershed Using Ultrahigh Resolution Mass Spectrometry. *Org. Geochem.* **2012**, *43*, 1–11.
- (83) Longnecker, K.; Kujawinski, E. B. Composition of Dissolved Organic Matter in Groundwater. *Geochim. Cosmochim. Acta* **2011**, *75*, 2752–2761.
- (84) Van Krevelen, D. W. Organic Geochemistry—old and New. *Org. Geochem.* **1984**, *6*, 1–10.
- (85) Perminova, I. V; Dubinenkov, I. V; Kononikhin, A. S.; Konstantinov, A. I.; Zhrebker, A. Y.; Andzhushev, M. a; Lebedev, V. a; Bulygina, E.; Holmes, R. M.; Kostyukevich, Y. I.; et al. Molecular Mapping of Sorbent Selectivities with Respect to Isolation of Arctic Dissolved Organic Matter as Measured by Fourier Transform Mass Spectrometry. *Environ. Sci. Technol.* **2014**, *48*, 7461–7468.
- (86) Capley, E. N.; Tipton, J. D.; Marshall, A. G.; Stenson, A. C. Chromatographic Reduction of Isobaric and Isomeric Complexity of Fulvic Acids to Enable Multistage Tandem Mass Spectral Characterization. *Anal. Chem.* **2010**, *82*, 8194–8202.

- (87) Stenson, A. C. Reversed-Phase Chromatography Fractionation Tailored to Mass Spectral Characterization of Humic Substances. *Environ. Sci. Technol.* **2008**, *42*, 2060–2065.
- (88) Solouki, T.; Freitas, M. A.; Alomary, A. Gas-Phase Hydrogen/Deuterium Exchange Reactions of Fulvic Acids: An Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectral Study. *Anal. Chem.* **1999**, *71*, 4719–4726.
- (89) Kostyukevich, Y.; Kononikhin, A.; Popov, I.; Kharybin, O.; Perminova, I.; Konstantinov, A.; Nikolaev, E. Enumeration of Labile Hydrogens in Natural Organic Matter by Use of Hydrogen/Deuterium Exchange Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 11007–11013.
- (90) Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P. High-Precision Frequency Measurements: Indispensable Tools at the Core of the Molecular-Level Analysis of Complex Systems. *Anal. Bioanal. Chem.* **2007**, *389*, 1311–1327.
- (91) 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. *Environ. Sci. Technol.* **2013**, *47*, 9891–9897.
- (92) 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. *Mar. Chem.* **2004**, *92*, 23–37.
- (93) Gonsior, M.; Peake, B. M.; Cooper, W. T.; Podgorski, D.; D’Andrilli, J.; Cooper, W. J. Photochemically Induced Changes in Dissolved Organic Matter Identified by Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Environ. Sci. Technol.* **2009**, *43*, 698–703.
- (94) Thorn, K. A.; Younger, S. J.; Cox, L. G. Order of Functionality Loss during Photodegradation of Aquatic Humic Substances. *J. Environ. Qual.* **2010**, *39*, 1416–1428.
- (95) Leenheer, J. A.; Wilson, M. A.; Malcolm, R. L. Presence and Potential Significance of Aromatic-Ketone Groups in Aquatic Humic Substances. *Environ. Sci. Technol.* **1987**, *11*, 273–280.
- (96) Aeschbacher, M.; Graf, C.; Schwarzenbach, R. P.; Sander, M. Antioxidant Properties of Humic Substances. *Environ. Sci. Technol.* **2012**, *46*, 4916–4925.
- (97) Scott, D. T.; McKnight, D. M.; Blunt-Harris, E. L.; Kolesar, S. E.; Lovley, D. R. Quinone Moieties Act as Electron Acceptors in the Reduction of Humic Substances by Humics-Reducing Microorganisms. *Environ. Sci. Technol.* **1998**, *32*, 2984–2989.

- (98) Mantini, D.; Petrucci, F.; Pieragostino, D.; Del Boccio, P.; Di Nicola, M.; Di Ilio, C.; Federici, G.; Sacchetta, P.; Comani, S.; Urbani, A. LIMPIC: A Computational Method for the Separation of Protein MALDI-TOF-MS Signals from Noise. *BMC Bioinformatics* **2007**, *8*, 1–17.
- (99) Reemtsma, T. Determination of Molecular Formulas of Natural Organic Matter Molecules by (ultra-) High-Resolution Mass Spectrometry: Status and Needs. *J. Chromatogr. A* **2009**, *1216*, 3687–3701.
- (100) 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 Molecular Formulas Have Been Assigned for 4626 Indi- Mass Measurements Fr. *Anal. Chem.* **2003**, *75*, 1275–1284.
- (101) McIntyre, C.; McRae, C. Proposed Guidelines for Sample Preparation and ESI-MS Analysis of Humic Substances to Avoid Self-Esterification. *Org. Geochem.* **2005**, *36*, 543–553.
- (102) Canonica, S.; Hellrung, B.; Wirz, J. Oxidation of Phenols by Triplet Aromatic Ketones in Aqueous Solution. *J. Phys. Chem. A* **2000**, *104*, 1226–1232.
- (103) Perdue, E. M.; Green, N. W. Isobaric Molecular Formulae of C, H, and O – A View from the Negative Quadrants of van Krevelen Space. *Anal. Chem.* **2015**, *87*, 5079–5085.

- (104) Green, N. W.; Perdue, E. M. Fast Graphically Inspired Algorithm for Assignment of Molecular Formulae in Ultrahigh Resolution Mass Spectrometry. *Anal. Chem.* **2015**, *87*, 5086–5094.
- (105) Kim, S.; Kaplan, L. a.; Benner, R.; Hatcher, P. G. Hydrogen-Deficient Molecules in Natural Riverine Water Samples - Evidence for the Existence of Black Carbon in DOM. *Mar. Chem.* **2004**, *92*, 225–234.
- (106) Trompowsky, P. M.; De Melo Benites, V.; Madari, B. E.; Pimenta, A. S.; Hockaday, W. C.; Hatcher, P. G. Characterization of Humic like Substances Obtained by Chemical Oxidation of Eucalyptus Charcoal. *Org. Geochem.* **2005**, *36*, 1480–1489.
- (107) Mueller-Harvey, I. Analysis of Hydrolysable Tannins. *Anim. Feed Sci. Tech.* **2001**, *91*, 3–20.
- (108) Schofield, P.; Mbugua, D. M.; Pell, A. N. Analysis of Condensed Tannins: A Review. *Anim. Feed Sci. Tech.* **2001**, *91*, 21–40.
- (109) Schendorf, T. Effect of Borohydride Reduction and pH on the Optical Properties of Humic Substances. M.S. Thesis., University of Maryland College Park, College Park, MD, 2013.
- (110) Coble, P. G. Marine Optical Biogeochemistry: The Chemistry of Ocean Color. *Chem. Rev.* **2007**, *107*, 402–418.

- (111) Guo, R.; Ma, J. Reduction-Induced Molecular Signature of Humic Substances: Structural Evidence for Optical Changes. *RSC Adv.* **2014**, *4*, 25880.
- (112) Bhatia, M. P.; Das, S. B.; Longnecker, K.; Charette, M. a.; Kujawinski, E. B. Molecular Characterization of Dissolved Organic Matter Associated with the Greenland Ice Sheet. *Geochim. Cosmochim. Acta* **2010**, *74*, 3768–3784.
- (113) Antony, R.; Grannas, A. M.; Willoughby, A. S.; Sleighter, R. L.; Thamban, M.; Hatcher, P. G. Origin and Sources of Dissolved Organic Matter in Snow on the East Antarctic Ice Sheet. *Environ. Sci. Technol.* **2014**, *48*, 6151–6159.
- (114) 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. *Mar. Chem.* **2004**, *89*, 327–337.
- (115) McKnight, D. M.; Andrews, E. D.; Spaulding, S. A.; Aiken, G. R. Aquatic Fulvic Acids in Algal-Rich Antarctic Ponds. *Limnol. Ocean.* **1994**, *39*, 1972–1979.
- (116) Koch, B. P.; Dittmar, T. From Mass to Structure: An Aromaticity Index for High-Resolution Mass Data of Natural Organic Matter. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 926–932.
- (117) Wagner, S.; Dittmar, T.; Jaffé, R. Molecular Characterization of Dissolved Black Nitrogen via Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Org. Geochem.* **2015**, *79*, 21–30.

- (118) Ohno, T.; He, Z.; Sleighter, R. L.; Honeycutt, C. W.; Hatcher, P. G. Ultrahigh Resolution Mass Spectrometry and Indicator Species Analysis to Identify Marker Components of Soil- and Plant Biomass- Derived Organic Matter Fractions. *Environ. Sci. Technol.* **2010**, *44*, 8594–8600.
- (119) Ohno, T.; Ohno, P. E. Influence of Heteroatom Pre-Selection on the Molecular Formula Assignment of Soil Organic Matter Components Determined by Ultrahigh Resolution Mass Spectrometry. *Anal. Bioanal. Chem.* **2013**, *405*, 3299–3306.
- (120) Lechtenfeld, O. J.; Kattner, G.; Flerus, R.; McCallister, S. L.; Schmitt-Kopplin, P.; Koch, B. P. Molecular Transformation and Degradation of Refractory Dissolved Organic Matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **2014**, *126*, 321–337.
- (121) Koch, B. P.; Kattner, G.; Witt, M.; Passow, U. Molecular Insights into the Microbial Formation of Marine Dissolved Organic Matter: Recalcitrant or Labile? *Biogeosciences* **2014**, *11*, 4173–4190.
- (122) Leito, I.; Herodes, K.; Huopolainen, M.; Virro, K.; Ku, A.; Krueve, A.; Tanner, R. Towards the Electrospray Ionization Mass Spectrometry Ionization Efficiency Scale of Organic Compounds. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 379–384.
- (123) Oss, M.; Krueve, A.; Herodes, K.; Leito, I. Electrospray Ionization Efficiency Scale of Organic Compounds. *Anal. Chem.* **2010**, *82*, 2865–2872.

- (124) Dalzell, B. J.; Minor, E. C.; Mopper, K. M. Photodegradation of Estuarine Dissolved Organic Matter: A Multi-Method Assessment of DOM Transformation. *Org. Geochem.* **2009**, *40*, 243–257.
- (125) Minor, E. C.; Swenson, M. M.; Mattson, B. M.; Oyler, A. R. Structural Characterization of Dissolved Organic Matter: A Review of Current Techniques for Isolation and Analysis. *Environ. Sci. Process. Impacts* **2014**, *16*, 2064–2079.
- (126) Abbt-Braun, G.; Lankes, U.; Frimmel, F. H. Structural Characterization of Aquatic Humic Substances-The Need for a Multiple Method Approach. *Aquat. Sci.* **2004**, *66*, 151–170.