LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 7, 2009, 81–95

Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: a qualitative assessment

William C. Hockaday,[1] Jeremiah M. Purcell,[2] Alan G. Marshall,[2] Jeffery A. Baldock,[3] and Patrick G. Hatcher[4]
[1]Rice University, Earth Science, P.O. Box 1892 MS 126, Houston, TX, USA
[2]Florida State University, National High Magnetic Field Laboratory, 1800 E. Paul Dirac Dr., Tallahassee, FL, USA
[3]CSIRO Land and Water, PMB #2, Glen Osmond, SA, Australia
[4]Old Dominion University, Department of Chemistry and Biochemistry, 4541 Hampton Blvd., Norfolk, VA, USA

Abstract

Fourier-transform ion cyclotron resonance mass spectrometry (MS) has demonstrated potential to revolutionize the fields of limnology and chemical oceanography by identifying the individual molecular components of organic matter in natural waters. The use of MS for this purpose is made possible by the electrospray technique which successfully ionizes polar, nonvolatile organic molecules. Another recently developed ion source, atmospheric pressure photoionization (APPI), extends MS capabilities to less polar molecules. This article presents early results on the application of APPI MS to natural organic matter. We compare APPI MS and electrospray MS data for dissolved organic matter from Lake Drummond (Virginia, USA). Collectively, electrospray and APPI MS identify more than 6000 molecular species to which we assign unique molecular formulas. Fewer than 1000 molecular species are common to both electrospray and APPI mass spectra, indicating that the techniques are highly complementary in the types of molecules they ionize. Access to a broad range of molecules provided by combining APPI and electrospray has prompted a qualitative analysis. The goal is to assess the extent to which molecular MS data correspond with elemental (CHNOS) and structural characteristics determined by combustion elemental analyses and ^{13}C nuclear magnetic resonance (NMR). Because the data obtained by these different methods are not directly comparable, we propose a novel data analysis procedure that facilitates their comparison. The bulk elemental composition calculated from electrospray MS data are in close agreement (± 15%) with values determined by combustion elemental analysis. APPI and electrospray MS detect protein contributions in agreement with ^{13}C NMR (6 wt %) but underestimate carbohydrates relative to ^{13}C NMR. Nevertheless, MS results agree with NMR on the relative proportions of noncarbohydrate compounds in the organic matter: lignins > lipids > peptides. Finally, we use a molecular mixing model to simulate a ^{13}C NMR spectrum from the MS datasets. The correspondence of the simulated and measured ^{13}C NMR signals (74%) suggests that, collectively, the molecular species identified by APPI and electrospray MS comprise a large portion of the organic matter in Lake Drummond. These results add credibility to electrospray and APPI MS in limnology and oceanography applications, but further characterization of ion source behavior is fundamental to the accurate interpretation of MS data.

Introduction

Determining the molecular identity of organic molecules in natural waters is fundamental to understanding the origin, reactivity, and environmental fate of natural and anthropogenic compounds. The composition of dissolved organic matter is relevant to issues of drinking water quality, aquatic ecology, biogeochemical cycling, and many others. The answers to fundamental questions, such as, "What happens to terrestrial organic matter in the Ocean?"(Hedges et al. 1997), require molecular information about terrestrial and marine end members. Nevertheless, more than half of the organic matter in natural waters is uncharacterized at the molecular level (Hedges et al. 2000). Recent advances in nuclear magnetic resonance (NMR) (see review by Cardoza et al. 2004) and the use of modeling for interpretation of complex NMR spectra have

*Corresponding author: E-mail: wch2@rice.edu

Acknowledgments

W.C.H. thanks the National Science Foundation–supported visiting scientist program (DMR-06-54118) at the National High Field Magnet Lab for making available the FT-ICR instrumentation and Rice University's Keith-Weiss postdoctoral fellowship program. W.C.H. and J.M.P. are grateful to J. Quinn for his technical assistance with modifications to the APPI source.

improved our understanding of organic matter structure and composition (Nelson et al. 1999; Hedges et al. 2002; Baldock et al. 2004; Nelson and Baldock 2005; Dickens et al. 2006; Perdue et al. 2007). Likewise, the coupling of electrospray ionization (ESI) to high-field Fourier-transform ion cyclotron resonance (FT-ICR) MS has provided unprecedented resolution of the molecular components of natural organic matter (NOM) as well as molecular details of biogeochemical processes (Kujawinski et al. 2002; Kim et al. 2003a, 2003b, 2004, 2006; Kramer et al. 2004; Kujawinski et al. 2004; Koch et al. 2005; Dittmar and Koch 2006; Grannas et al. 2006; Hertkorn et al. 2006; Kujawinski and Behn 2006; Mopper et al. 2007; Reemtsma et al. 2008; Sleighter and Hatcher 2008).

Electrospray ionization (ESI) is selective for specific types of molecules. Hence, an important problem in the interpretation of ESI MS spectra of natural waters is that the proportion of NOM observed in mass spectra remains unknown, making it difficult to assess the qualitative or quantitative importance of MS data. For instance, the limnology and oceanography implications would differ significantly if the molecules identified by FT-ICR MS represent only 1% of dissolved NOM, as opposed to 70% of dissolved NOM.

The complexity of the electrospray ionization mechanism in addition to the complexity of organic matter in natural waters makes it difficult to assess how much of the sample is being ionized. Electrospray ionization is selective for species that exist as ions in solution or those with acidic or basic functionality capable of donating or stabilizing excess electrolytes (e.g., H^+, Na^+, Cl^-). Ion abundances are determined by a competition among solutes for charge at the surface of the electrospray droplet (Kebarle and Pesche 2000 and references therein). Therefore, ion abundance (peak magnitude) is not simply a function of analyte concentration. Minor constituents with high charge affinity can dominate the ion current. These so-called matrix effects make it impossible to calibrate peak magnitude as a function of dissolved analyte concentration in ESI MS. Therefore, a quantitative evaluation of ESI data for unknown mixtures such as dissolved organic matter in natural waters is difficult.

A newer ionization technique, atmospheric pressure photoionization, is less susceptible to the problems mentioned above. Several attributes of APPI make it well-suited to environmental samples: (*a*) there is little or no ion suppression due to charge competition among solutes (matrix effects); (*b*) there is minimal chemical noise because common solvents and salts are not ionized, and therefore, do not enter the mass analyzer (Hanold et al. 2004) (this feature is especially important in FT-ICR MS, in which space-charge effects influence sensitivity and resolution [Marshall et al. 1998]); (*c*) the APPI source is nearly field-free, which improves the efficiency of ion transport to the MS inlet, resulting in improved sensitivity (Robb et al. 2000; Hanold et al. 2004); and (*d*) photoionization is a soft ionization process, which does not cause extensive fragmentation of covalent bonds.

APPI is complementary to ESI, extending the analytical window to nonpolar and nonvolatile species. APPI has been successfully applied to a range of environmentally relevant molecules that ionize poorly via electrospray. These include aliphatic lipids (Cai and Syage 2006; Roy et al. 2006), isoprenoid lipids (Cai et al. 2005), condensed tannins (Gomez-Ariza et al. 2006), hydrophobic peptides, polycyclic aromatic hydrocarbons (Hanold et al. 2004; Hiroshi et al. 2004; Straube et al. 2004; Purcell et al. 2006; Zheng and Shamsi 2006), pesticides (Takino et al. 2003; Yoshioka et al. 2004), mycotoxins (Takino et al. 2003), and crude oil fractions (Purcell et al. 2006, 2007). Here, we report early results on the application of APPI FT-ICR MS to dissolved organic matter from natural waters.

The first part of this method assessment compares the NOM molecular species ionized by APPI and ESI. The second part of our assessment addresses the following question: Are the ions detected by ESI MS and APPI MS an accurate representation of NOM? Given the collective capacity of ESI and APPI sources to ionize polar and nonpolar, acidic and basic, aromatic and aliphatic molecules, it seems reasonable to begin asking whether the molecules observed are representative of the NOM mixture as a whole. On the basis of molecular weight (<1000 Da), ESI MS and APPI MS disagree with the traditional description of NOM as macromolecular or supramolecular products of polycondensation reactions (humification) (Swift 1999; Piccolo 2001). A new view has emerged, however, which reconciles observations of macromolecular size to MS-based observations of low molecular weight. The new view describes NOM and humic substances as a dynamic, noncovalent association of small, recognizable biological molecules in various stages of decomposition (Hedges et al. 2000; Burdon 2001; Wrobel et al. 2003; Sutton and Sposito 2005; Smejkalova and Piccolo 2008). A similar aggregation behavior has been observed in noncovalent oligomers in fossil fuels (Piccolo et al. 2002; Mullins et al. 2008). For instance, the large apparent size of humic substances could be attributed to the high concentration and/or choice of solvents in size exclusion chromatography experiments in which high-molecular-weight species are observed. ESI and APPI MS are typically conducted at much lower concentration (0.5 mg/mL or less), such that noncovalent adducts dissociate before ionization takes place. Therefore, we assess whether MS data are representative of NOM based on three characteristics that are independent of molecular size. These include a comparison of MS results to (1) elemental composition measured by combustion elemental analysis, (2) the distribution of plant biomolecules (lipids, lignins, proteins, and carbohydrates) determined by ^{13}C NMR combined with a molecular mixing model used previously in other NMR applications (Baldock et al. 2004), and (3) bulk structural properties measured by ^{13}C NMR.

Materials and procedures

Sample preparation—The dissolved organic matter used in this study was collected from the Great Dismal Swamp

National Forest located near Suffolk, VA, USA. Surface water was collected in December (2005) from a pier in Lake Drummond, a shallow, blackwater swamp lake centrally located within the Great Dismal Swamp. The waters are known for their antibacterial properties attributed to high tannin levels from the swamp's Atlantic white cedar [*Chamaecyparis thyoides (L.)*] and bald cypress [*Taxodium distichum (L.)*] stand. Dissolved organic carbon concentrations were 105 mg/L at the time of sampling, and dissolved organic matter was isolated by filtering the water through previously combusted (450 °C) Whatman GF/F (0.7 μm pore size) filters to remove particulates, and then by C_{18} solid-phase extraction (3M; Empore, St. Paul, MN). The natural pH of the Lake Drummond water was 3.4 and was acidified to pH 2 using 6N HCl before solid-phase extraction. The organic matter adsorbed to the C_{18} was immediately eluted with HPLC-grade methanol (Fisher Scientific). This extraction procedure typically recovers 60% to 70% of DOM from freshwater samples (Kim et al. 2003b), though the extraction is more quantitative (100% recovery) for phenolic lignin-type compounds (Louchouarn et al. 2000). This solution was directly analyzed by mass spectrometry. For combustion elemental analysis and NMR spectroscopy, the methanol was evaporated, first under a stream of nitrogen gas, and then at 50 °C in a drying oven for 24 h.

Elemental analyses—The elemental composition of the dry solid DOM extract was performed by combustion analysis. Carbon, hydrogen, and nitrogen content were determined by catalytic combustion with a model ESC4010 elemental analyzer (Costech Analytical Technologies Inc., Valencia, CA). The sulfur content was determined by combustion method ASTM D4239 B at Galbraith Laboratories (Knoxville, TN). Ash content was determined after combustion at 600 °C for 24 h. Oxygen content was calculated as the residual mass after summation of C + N + H + S + ash. All measurements were performed in duplicate.

^{13}C NMR—The ^{13}C NMR spectrum of the dry solid C_{18} extract was collected using a 300-MHz Bruker DSX spectrometer (Billerica, MA) with a 4-mm magic angle spinning (MAS) probe operated at 13 kHz. Cross-polarization was achieved by a 90-degree ^{1}H pulse and a ramped amplitude (0.5–1.0 db) ^{1}H-^{13}C contact pulse (2 ms pulse length) (Metz et al. 1994). A recycle delay of 1 s was applied between scans, and ^{13}C signal was collected with two-pulse phase-modulated proton decoupling. Acquisition parameters and chemical shift were calibrated with respect to polycrystalline glycine. Exponential multiplication, one zero fill, and 30-Hz line broadening were applied during data processing. Peaks were integrated according to Baldock et al. (2004) for compatibility with the molecular mixing model. Peak areas were corrected for spinning side bands.

ESI source—The external ESI source used in this study comprised mainly a fused silica capillary secured by an electrical contact at the opening of a heated steel capillary entering the skimmer region of the mass spectrometer. The voltage bias across the silica and steel capillaries was approximately 3 kV. The appropriate polarity was selected for positive or negative ion detection. Sample flow rate through the fused silica capillary was maintained by a syringe pump at 0.5 μL/min.

APPI source—The APPI ion source was supplied by Thermo Fisher and outfitted with a krypton vacuum UV lamp (10 eV). Its operating principles and parameters are described in detail by Purcell et al. (2006). Toluene was used a dopant, to increase the ionization efficiency of the NOM. Direct addition of toluene to the NOM sample resulted in precipitation of the NOM. Therefore, a slight modification was made to this ion source to allow the introduction of toluene vapor through the auxiliary gas inlet of the nebulizer. The toluene was introduced as a vapor by injection into a segment of heated (90 °C) copper tubing purged with carbon dioxide gas at a flow rate of 10 mL/min. The nebulizer heater was operated at 300 °C, with carbon dioxide as the drying gas. The syringe pump used to introduce the sample to the nebulizer was operated at 50 μL/min.

9.4-T FT-ICR MS—All MS experiments were performed at the National High Field Magnet Laboratory (Tallahassee, FL) using a custom-built FT-ICR mass spectrometer equipped with a 9.4-T Oxford magnet. Ions were produced at atmospheric pressure by external ESI and APPI ion sources and entered the evacuated skimmer region of the spectrometer through a common interface consisting of a heated metal capillary. A detailed description of the spectrometer hardware and operational parameters during signal acquisition is given by Purcell et al. (2006). Spectrometer parameters (e.g., ion accumulation period and multipole rf amplitude and frequency) were not held constant for each ionization technique. Rather, these parameters were optimized to accommodate differences in the mass-to-charge ratio and the abundance of ions. The goal of parameter optimization was to collect the broadest possible range of ions generated by each ion source as opposed to focusing on ions of the same mass-to-charge ratio (m/z) for each of the three ionization techniques. In other words, we systematically adjusted the ion optics so as to select a representative sample of the ions being formed (Brown and Rice 2000). Ions were accumulated in the first octapole for a period of 20 s (APPI) or 25 s (ESI) before transfer to a second octapole, where they were collisionally cooled with helium gas for 10 ms and finally transferred to a Penning ion trap by a 10-V offset (1 ms in duration). Ions were excited to a cyclotron orbital radius by applying a broadband frequency (~90–600 kHz) sweeping pulse. The induced current was detected on two opposing electrodes on the ion trap. Approximately 100 time–domain transient signals were summed, Hanning-apodized, and Fourier-transformed in the Modular ICR Data Acquisition and Analysis System (MIDAS) (Senko et al. 1996). Mass calibration was referred to an external standard purchased from Agilent Technologies (Santa Clara, CA) (Ledford et al. 1984; Shi et al. 2000). The accurate mass (m/z) values were determined to 6 decimal places. Mass error for all spectra is <1 ppm.

Algorithms for data reduction and structure interpretation

The information contained in ultra-high-resolution FT-ICR mass spectra of natural organic matter is expansive. Typical spectra contain 10,000 signals, each having an m/z value and a corresponding measure of relative abundance. On the other hand, one-dimensional ^{13}C NMR spectra contain information on carbon functional groups and their relative abundance, with little information about how these bonds are arranged into molecular structures. Therefore, we employ various mathematical algorithms to extract relevant information from NMR and MS data: molecular formulas, structural information, and molecule type (e.g., carbohydrate, peptide). The algorithms described below are all previously published, and we refer the reader to the original publications. Several novel modifications to these algorithms are discussed in the Assessment section.

Molecular mixing model for ^{13}C NMR data—The molecular mixing model is described in detail by Baldock et al. (2004). Briefly, the model uses ^{13}C NMR peak areas in addition to C and N elemental data to estimate the relative proportions of major biomolecule classes. For samples of terrestrial origin, the model is parameterized with the NMR average peak area distributions of 4 representative biomolecule classes: cellulose, lignins, lipids, and peptides. Charcoal carbon is also included as a component in the mixing model because of its abundance in the soil organic matter and dissolved organic matter of ecosystems that have been exposed to fire (Hockaday et al. 2006). The mixing model calculates the linear combination of these five components that achieves the best fit to the measured NMR peak area distribution. We use the terms charcoal carbon and condensed aromatic structures interchangeably in this article.

Molecular formula calculator for MS peak assignment—Each m/z value determined by FT-ICR MS was assigned a unique molecular formula with the Formula Calculator Program version 1.1 (© 1998 NHMFL). In some cases, multiple formulas ($C_cH_hN_nO_oS_s$) agreed with the measured mass (±1 ppm) and also obeyed the rules of chemical bonding. The correct formula assignment was determined from reliable patterns which result from isotopic distributions, elemental substitution patterns, and homologous series of molecules. The procedure used for validating formula assignments is explained in detail by Koch et al. (2007).

Kendrick mass defect analysis—Kendrick mass defect (KMD) analysis identifies homologous series of molecules by identifying ions that are separated by integer multiples of a given functional group (Kendrick 1963; Hsu et al. 1992; Stenson et al. 2002). For example, Equation 1 has the effect of converting to a mass scale based on CH_2 rather than ^{12}C (12.00000 Da).

$$\text{Kendrick mass} = \text{measured mass in SI units} \times 14.00000/14.01565 \quad (1)$$

Species separated by 14.01565 Da (CH_2) have the same KMD value as determined by Equation 2, where the nominal Kendrick mass is defined as the highest integer value of the Kendrick mass.

$$\text{KMD} = \text{nominal Kendrick mass} - \text{Kendrick mass} \quad (2)$$

Hydrogen deficiency (z) and double bond equivalents (DBE)—Hydrogen-deficiency and double bond equivalents, as defined in Equations 3 and 4, are common measures of unsaturation. DBE is defined as the number of rings plus double bonds in a neutral molecule.

$$z\,(C_cH_hN_nO_oS_s) = -2c + h \quad (3)$$

$$DBE\,(C_cH_hN_nO_oS_s) = -\frac{z}{2} + \frac{n}{2} + 1 \quad (4)$$

Van Krevelen analysis—The van Krevelen analysis (Van Krevelen 1950) is a graphical and statistical treatment of elemental data (C, H, N, and O). The most common van Krevelen diagram is a two-dimensional plot of O/C and H/C ratio for each molecular formula. Natural biomolecules have relatively uniform and discrete elemental compositions such that different types of molecules can be distinguished by some combination of their H/C, O/C, and N/C ratios. The van Krevelen diagram can be used to visualize the elemental composition-space occupied by molecules identified by FT-ICR MS, and this facilitates an approximation of the biomolecule distribution (e.g., lipid, lignin, carbohydrate, and peptide) (Kim et al. 2003a).

Assessment

Atmospheric pressure photoionization is relatively new and has not previously been used for a detailed molecular analysis of NOM. Therefore, the first part of our assessment compares and contrasts APPI MS to the more widely used ESI MS. We describe the types of ions generated by each and employ several (published and novel) algorithms that simplify the comparison of large MS datasets. The second part of the assessment addresses the degree to which molecular MS data can be reconciled to bulk chemical characteristics measured by elemental analysis and NMR. Because the raw data generated by MS (mass-to-charge ratio), combustion analyses (mass percent C, H, N, O, and S), and NMR (chemical bond type) cannot be directly compared in a meaningful way, we propose a data analysis protocol that facilitates method comparisons. The flow diagram in Figure 1 summarizes the sequence in which we apply data analysis algorithms (described above). The double arrows along the counterclockwise flow of the diagram represent method comparisons that address the following questions: *Does the elemental composition of NOM determined by molecular MS agree with combustion elemental analyses? Do MS and NMR observe the same types of molecules in natural organic matter, and in similar quantities? Are some chemical structures in NOM ionized preferentially in ESI and APPI?* The answers to these questions have important implications for how we interpret MS data and our understanding of organic matter in natural waters.

Electrospray versus APPI

Ion source behavior—Positive-ion APPI, along with positive- and negative-ion ESI FT-ICR mass spectra, are shown in Figure 2. Ions generated by all three ionization techniques fall within the range $250 < m/z < 800$. Relative ion abundance also shows a similar distribution across ionization techniques as indicated

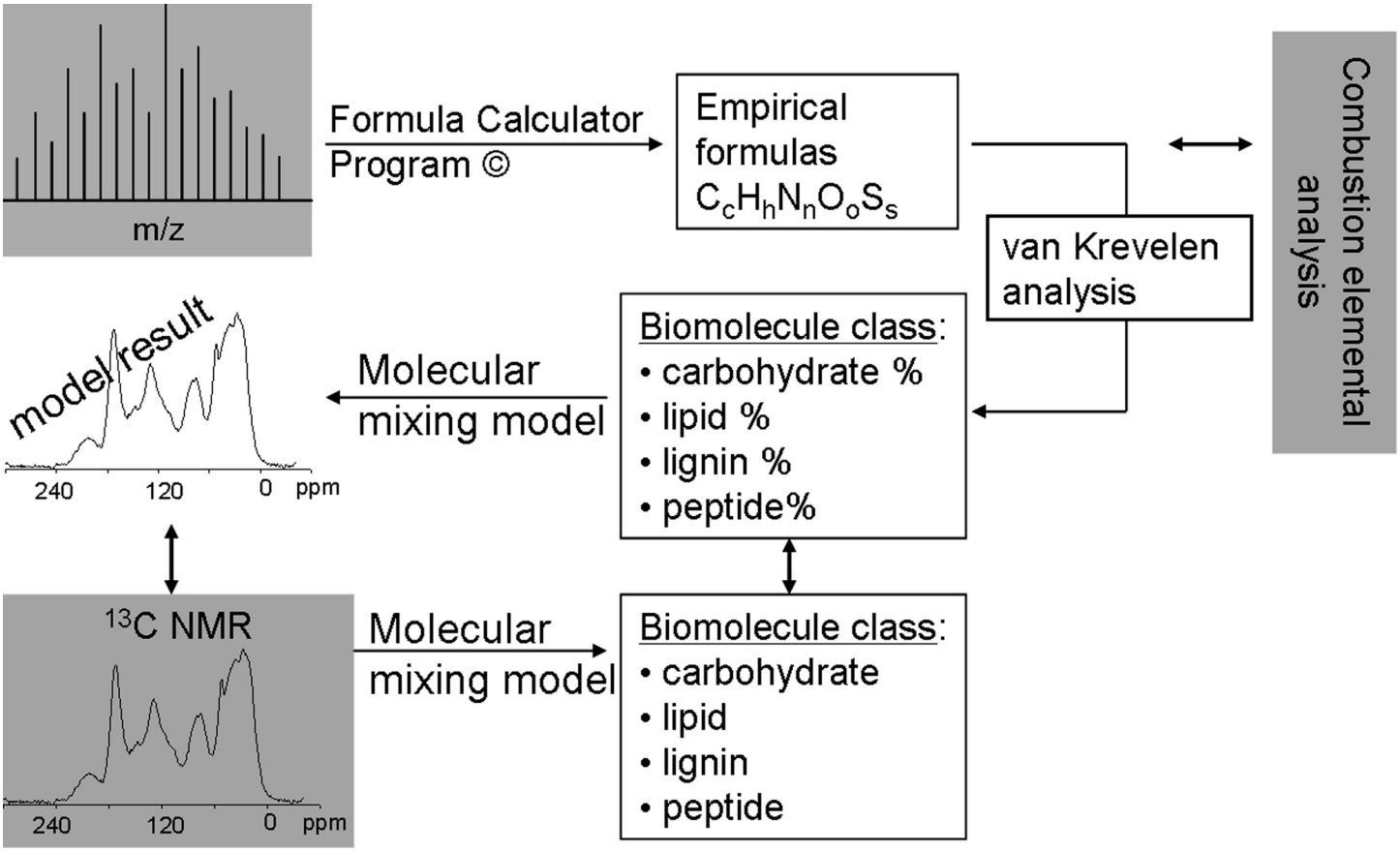

Fig. 1. Data assessment flow diagram. Gray shapes represent different analytical methods; double arrows indicate a method comparison; arrows represent published algorithms for data analysis; and boxes represent the processed outcome.

by close agreement in the average molecular weights (M_n) in Table 1 (calculated according to O'Loughlin and Chin 2001). The m/z spacing of 1.0034 between $^{12}C_c$ and $^{13}C_1{}^{12}C_{c-1}$ forms of the same ion indicates that ions are singly charged. Most ion species in positive-ion APPI, and ESI are H^+ and Na^+ charge adducts. Similarly, the negative-ion ESI mass spectrum exhibits singly-charged, deprotonated species, many of which are counterparts to species generated by positive-ion ESI.

It is also important to note that the negative-ion ESI mass spectrum is largely dominated by chlorinated species, including both single Cl^- charge adducts and molecular species with multiple covalently bound chlorine atoms. The latter give rise to characteristic isotopic distributions with 1.9970-Da mass differences between ^{35}Cl- and ^{37}Cl-bearing species. The source of chlorine in this sample is likely a residue from the HCl used to acidify the water before solid-phase (C_{18}) extraction of the dissolved organic matter, as dissolved chlorine concentrations in Lake Drummond are quite low (9 mg/L) relative to dissolved organic carbon concentrations (105 mg/L) (Johannesson et al. 2004). Nevertheless, natural organo-chlorine species are well documented, and we cannot rule out the possibility that they are preexisting components of Lake Drummond NOM (Myneni 2002; Repeta et al. 2004; Leri et al. 2007; Teuten and Reddy 2007).

Molecules containing chlorine(s) give rise to strong signals in negative-ion ESI MS, an order of magnitude greater than nonchlorinated species. These are seen as off-scale peaks in the bottom panel of Figure 2, illustrating a commonly encountered limitation of the ESI technique for investigations of organic molecules in environmental samples. Competition for charge among the solutes (analyte–analyte or analyte–contaminant) can result in low signal-to-noise ratio or the suppression of analytes with lower charge affinity. Even though ion suppression has diminished the signal from thousands of organic analytes, the negative-ion ESI mass spectrum still contains a great deal of information about the NOM, and therefore we include it in our analysis.

For positive-ion APPI with toluene gas as a dopant, two types of ions can be generated from the same precursor molecule, M: $M^{\cdot+}$ and $(M+H)^+$. Thus, for a given parent molecule M, positive-ion APPI can require resolution of the mass doublet corresponding to $M^{\cdot+}$ with one ^{13}C and $(M+H)^+$ composed of ^{12}C only. Resolution of the 4.5-mDa mass difference requires FT-ICR MS. Approximately 90% of the positive ions generated by APPI$^+$ are protonated species, $(M+H)^+$, formed by proton transfer from photo-excited toluene dopant (Purcell et al. 2007) or from methanol, which behaves as a protic solvent

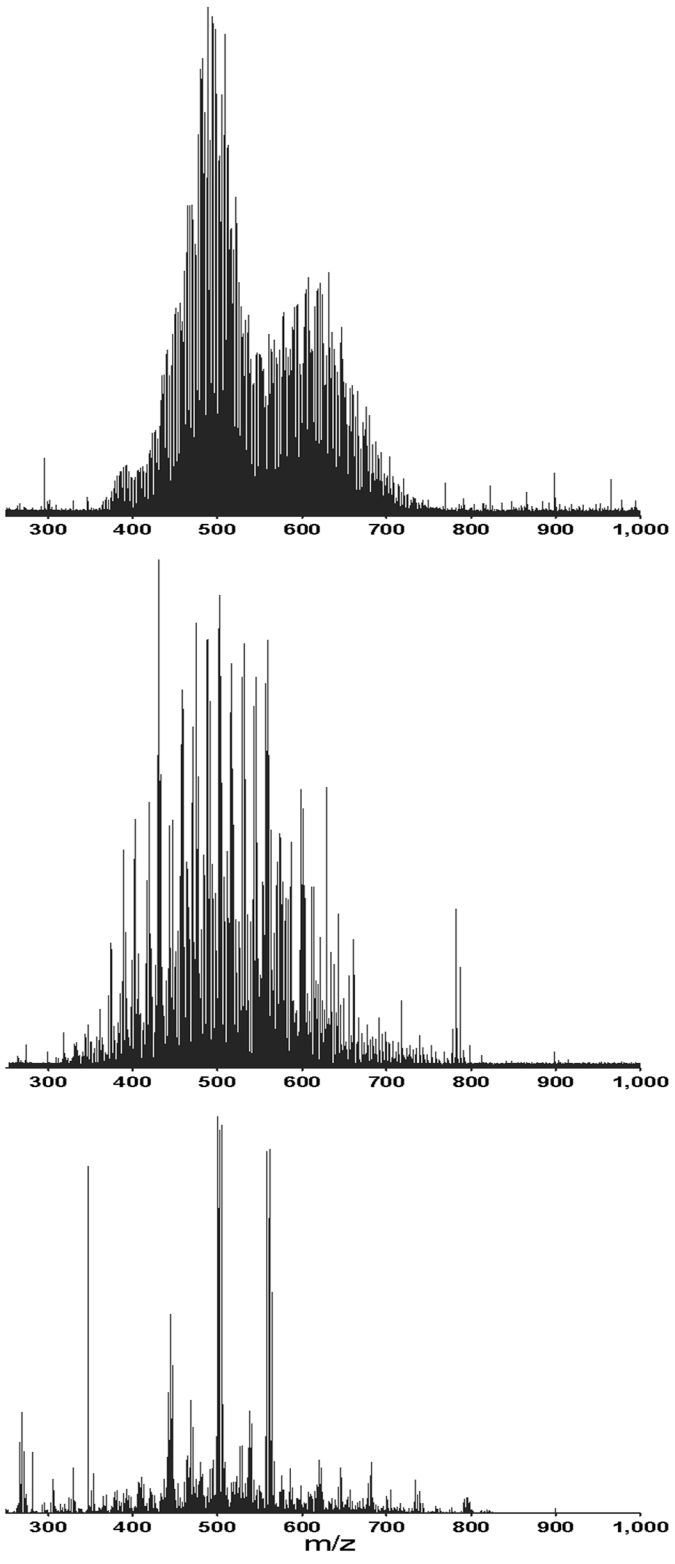

Fig. 2. FT-ICR mass spectra of C_{18} extractable dissolved NOM from Lake Drummond (Great Dismal Swamp National Forest) obtained with atmospheric pressure ion sources: (+)APPI (upper), (+)ESI (middle), (–)ESI (lower). High-magnitude peaks in the ESI^- spectrum are chlorinated species.

Table 1. APPI versus electrospray: molecular weight (MW) and elemental composition of dissolved organic matter.

	(+) APPI MS	(+) ESI MS	(–) ESI MS
MW range, Da	365–750	300–800	250–825
MW_n, Da	543	554	555
H/C_n	1.01	1.21	1.28
O/C_n	0.25	0.32	0.36
N/C_n	0.011	0.013	0.035
S/C_n	0.00016	0.0061	0.0088

MW_n, number average molecular weight (calculated according to O'Loughlin and Chin 2001). All averages exclude compositionally redundant ions (e.g., ^{13}C versus ^{12}C, H^+ and Na^+ adducts versus radical cations) for the same neutral precursor.

(Robb et al. 2000). Radical cations, $M^{\cdot+}$, account for the remaining 10% of the peaks, and are all highly conjugated species, with DBE ≥10 (see Equation 3). On average, the NOM ionized by APPI is more aromatic and less polar than NOM ionized by electrospray. Evidence is seen Table 1, in which the APPI ion suite has lower average H/C, O/C, N/C, and S/C ratios than the NOM ionized by positive- or negative-ion ESI.

Collectively, ESI and APPI FT-ICR MS identified more than 6000 unique molecular species in Lake Drummond NOM (species differing by ^{13}C vs. ^{12}C or H^+ vs. Na^+ are counted only once). A side-by-side comparison of the molecular formulas shows that fewer than 1000 of the 6000 species are common to ESI and APPI. Therefore, ESI and APPI are highly complementary for the analysis of NOM.

Structurally related molecules: data reduction with the zmo *table*—The suite of molecules ionized by all three techniques exhibits prevalent CH_2 homologous series. Upon grouping the molecular formulas by hydrogen deficiency (*z*) and Kendrick mass defect (KMD), it becomes apparent that most CH_2 series intersect with a homologous oxygen series. For example, the most intense peak in positive-ion ESI and APPI mass spectra has the (charge-neutral) molecular formula $C_{29}H_{28}O_7$ and belongs to a CH_2 homologous series $C_{29}H_{28}(CH_2)_mO_7$, with *m* values ranging from –7 to +10. Moreover, for each value *m*, there exists a homologous oxygen series $C_{29}H_{28}(CH_2)_mO_o$, with *o* values ranging from 2 to 13. Furthermore, the functionality of the oxygen can be deduced because in order to change the oxygen number, *o*, without altering *m* or *z*, the oxygen-containing functional group must not introduce a double-bond equivalent to the molecule. This limitation narrows the possible oxygen-containing functionality to hydroxyl and ether groups because all others (ester, carboxyl, amide, ketone, aldehyde, or cyclic ether) alter the value of *z* or *m*. Thus, the oxygen series listed in Table 2 are likely comprised of structurally related species.

The *zmo* table (Table 2) contains only 22 entries, but it represents approximately 3000 (50%) of the positive ions in the APPI and ESI mass spectra. For each entry, the *zmo* table conveys the molecular formula, the hydrogen deficiency (*z*),

Table 2. Data-reducing *zmo* table for MS data.

	(+) APPI			(+) ESI		
	z^a	m^b	o^c	z^a	m^b	o^c
$C_{23}H_{40}(CH_2)_mO_o$				–8	–3–1	8–12
$C_{23}H_{38}(CH_2)_mO_o$				–10	–3–5	9–13
$C_{23}H_{36}(CH_2)_mO_o$				–12	–3–7	6–13
$C_{23}H_{34}(CH_2)_mO_o$				–14	–3–9	5–14
$C_{23}H_{32}(CH_2)_mO_o$	–16	–1–7	5–10	–16	0–8	5–17
$C_{23}H_{30}(CH_2)_mO_o$	–18	–3–8	3–11	–18	–3–11	5–15
$C_{23}H_{28}(CH_2)_mO_o$	–20	–3–9	3–12	–20	–3–13	5–16
$C_{23}H_{26}(CH_2)_mO_o$	–22	–3–12	2–13	–22	–1–15	5–16
$C_{23}H_{24}(CH_2)_mO_o$	–24	–3–14	2–13	–24	–1–15	5–16
$C_{23}H_{22}(CH_2)_mO_o$	–26	–2–17	2–14	–26	1–15	5–16
$C_{23}H_{20}(CH_2)_mO_o$	–28	–2–17	2–15	–28	0–13	2–16
$C_{23}H_{18}(CH_2)_mO_o$	–30	–2–18	2–15	–30	1–13	1–15
$C_{23}H_{16}(CH_2)_mO_o$	–32	–1–19	2–13	–32	3–13	1–16
$C_{23}H_{14}(CH_2)_mO_o$	–34	–1–19	3–15	–34	4–14	1–15
$C_{23}H_{12}(CH_2)_mO_o$	–36	0–20	2–15	–36	5–15	1–8
$C_{23}H_{10}(CH_2)_mO_o$	–38	1–19	2–15	–38	7–13	2–7
$C_{32}H_{26}(CH_2)_mO_o$	–40	–2–12	2–15			
$C_{32}H_{24}(CH_2)_mO_o$	–42	–8–12	2–16			
$C_{32}H_{22}(CH_2)_mO_o$	–44	–6–11	2–15			
$C_{32}H_{20}(CH_2)_mO_o$	–46	–4–11	2–13			
$C_{32}H_{18}(CH_2)_mO_o$	–48	–2–11	2–13			
$C_{32}H_{16}(CH_2)_mO_o$	–50	0–11	2–12			

Structural relationships are revealed by H-deficiency (*z*), CH_2 homologous series (*m*), and nonstructural oxygen (*o*).
[a]Calculated with Equation 2.
[b]Species of a given *m* and *z* value are structurally similar for all values of *o*.
[c]Species of a given *z* and *o* value represent a CH_2 homologous series for all values of *m*.

number of CH_2 groups (*m*) in the homologous series, and the number of oxygen atoms (*o*) in the homologous series. Double-bond equivalents are also easily obtained from Table 2 (see Equation 4). The most striking difference between APPI and ESI datasets is the difference in hydrogen deficiency (*z*) of the respective ion suites. The *zmo* table also shows that most homologous series of molecules are common to both ESI and APPI mass spectra, but within these series, different molecules are detected. Positive-ion ESI yields more highly oxidized members of homologous oxygen series (higher *o*), whereas positive-ion APPI yields species with greater numbers of CH_2 groups (broader range of *m*).

Methods comparison: elemental, structural, and molecular measurements

The intent of the following method comparison is to evaluate the extent to which molecules ionized collectively by ESI and APPI are representative of the entire NOM molecular assemblage, based on elemental, structural, and molecular compositions.

(1) Elemental composition—The first approach is simply to obtain the bulk elemental composition of the organic matter observed in the MS spectra by weighting the element stoichiometry of each species ($C_cH_hN_nO_oS_s$) by its corresponding spectral magnitude, and averaging across all molecular formulas. This approach generates an elemental composition that assumes equal ionization efficiencies for all analytes and spectral magnitudes that are proportional to analyte concentration. Table 3 lists the average molar H/C, O/C, N/C, and S/C ratios obtained by each MS technique for comparison to combustion elemental analyses. The molecular mixing model calculates the bulk H/C, O/C, and N/C values from ^{13}C NMR data based on typical elemental compositions for terrestrial biomolecules.

Table 3 suggests that APPI generates a suite of ions that are slightly H-, O-, and N-deficient relative to the bulk NOM. This is consistent with the photoionization mechanism, in which UV photons are most readily absorbed by double bonds and aromatic rings. Thus, condensed structures are preferentially ionized. The H/C, O/C, N/C, and S/C ratios of the ESI^+ and ESI^- ion suites agree (±15%) with values determined by elemental analysis. The most substantial error in elemental composition determined by ESI MS is the underestimation of organic oxygen, attributable to poor ionization efficiency of carbohydrates.

(2) Biomolecule distribution—In contrast to the previous two approaches that average across molecular information to

Table 3. Comparison of methods: bulk chemical characteristics of dissolved organic matter determined by elemental analysis, ^{13}C NMR, ESI MS, and APPI MS.

	EA[a]	^{13}C NMR	(+) APPI MS	(+) ESI MS	(–) ESI MS
H/C	1.12	1.42[b]	1.03	1.27	1.33
O/C	0.42	0.49[b]	0.26	0.35	0.37
N/C	0.016	0.020[b]	0.0042	0.0055	0.010
S/C	0.0057	nd	3.0×10^{-5}	0.0038	0.0025
Aromaticity (f_a)	0.45[c]	0.42[d]	0.53[c]	0.46[c]	0.46[c]

MS data are calculated from peak magnitude–weighted molecular formulas.
[a]Number of replicate analyses = 2.
[b]Calculated by the molecular mixing model (Baldock et al. 2004).
[c]Calculated by Equation 5.
[d]Calculated by Equation 6.

obtain bulk chemical properties, the third approach interprets NMR and MS data in terms of discrete biochemical structures commonly found in soils and natural waters. This comparison is facilitated by the NMR molecular mixing model and the van Krevelen analysis (Kim et al. 2003a; Baldock et al. 2004).

Although classifying species identified by FT-ICR MS into only five biochemical categories is an obvious oversimplification, we do so here for the sake of comparing ionization techniques to one another and to NMR mixing model results. For instance, lipids, as defined below, should include all of the following: aliphatic hydrocarbons, fatty acids, aliphatic alcohols and esters, terpenes, and sterols. Similarly, compounds classified as lignin encompass polyphenolic oligomers comprised of syringyl, guiacyl, and *p*-hydroxyphenyl units, their acid forms, and other phenolic phytochemicals such as condensed tannins, hydrolysable tannins, flavonoids, and alkaloids. Thus, we caution that element stoichiometries and aromaticity indices defined below, and summarized in Figure 3, are not infallible means of structure assignment. Rather, they are proposed as a rational set of rules that are based on molecular formulas and structures of well-characterized molecules in the natural products literature (e.g., Kaufman et al. 1999; Romero 2005).

Lignins are defined as having H/C = 1.5 to 0.7 and O/C = 0.1 to 0.67. A small portion of molecules (<5%) in that range contain a nitrogen atom, but they are too hydrogen-deficient (H/C = 0.6 to 1.0) to exist as small peptides (300–800 Da). They are classified as lignins for the purpose of comparison to the ^{13}C NMR mixing model, but perhaps they are more likely to be aromatic amines such as alkaloids or the products of polyphenol-peptide binding reactions that are known to occur in the tannin-rich acidic conditions that exist in Lake Drummond. Polyphenol-peptide reaction products have been documented in soils and sediments (Hsu and Hatcher 2005; Olk et al. 2006; Mutabaruka et al. 2007).

Proteins in the mass range 300–800 Da are more accurately described as peptides of 3 to 6 amino acid residues. Elemental composition: N ≥ 1; H/C = 1.0 to 2.2; O/C = 0.1 to 0.67. Typical amino acid distributions in forest soils indicate a protein fraction with average H/C = 1.7 and O/C = 0.4 (Friedel and Scheller 2002).

Carbohydrates are considered to be hexose and pentose glycosides, including those with aldehyde, ketone, and/or acid functionality. Carbohydrates may contain nitrogen as in chitin. Molar ratios: H/C > 1.5 and O/C > 0.67.

Lipids are defined by H/C = 1.5 to 2.0 and O/C < 0.3. However, a substantial number (8%) of species in the positive-ion ESI mass spectrum have H/C = 1.5 to 2.0 and O/C = 0.3 to 0.7, and thus cannot be classified as carbohydrates, lipids, or proteins by the foregoing criteria. Their elemental compositions suggest that they possess both carbohydrate-like and aliphatic

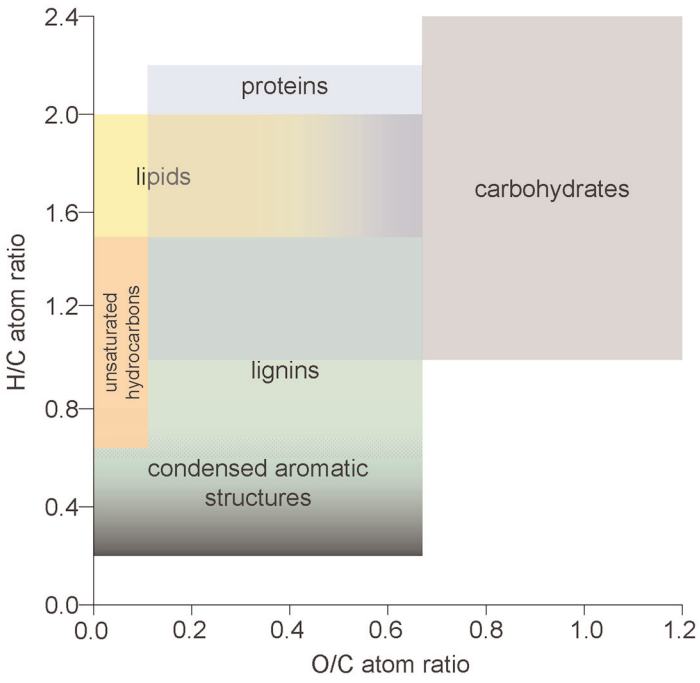

Fig. 3. Model van Krevelen diagram showing the H/C and O/C space occupied by broadly classified natural products. Regions in which structures share common element ratios (represented by faded lines) are distinguished by nitrogen content or aromaticity index.

Table 4. Comparison of methods: molecular composition of natural organic matter as observed by APPI MS, ESI MS, and ^{13}C NMR.

	(+) APPI MS	(+) ESI MS	(–) ESI MS	MS total[a] (MS mean)[b]	^{13}C NMR + model
Lipid %	3	20	14	9	28
	(9)	(26)	(16)	(16)	
Lignin %	87	71	69	81	54
	(66)	(51)	(52)	(58)	
Peptide %	6	6	14	6	7
	(16)	(15)	(26)	(18)	
Carbohydrate %	1	1	1	1	9
	(1)	(4)	(2)	(2)	
Condensed aromatic structures	3	2	2	3	2
	(8)	(4)	(4)	(6)	
Number of unique molecular species[c]	4110	2585	2276		
Relative spectral magnitude (%)[d]	65	32	3		

Uppermost values are integrated spectral magnitude. Parenthetical values are unweighted numerical averages. [a]MS total = relative spectral magnitude across all ion sources. [b]MS mean = numerical average across all ion sources. [c]Excludes compositionally redundant ions (e.g., ^{13}C versus ^{12}C, H^+ and Na^+ adducts versus radical cations) for the same neutral precursor. [d]Relative spectral magnitude is the percentage of the total integrated MS magnitude from all ion sources.

or alicyclic domains. This is the case for many common plant terpenes and sterols that contain a glycone unit (e.g., saponins). Furthermore, many of these species belong to homologous series differing only in the number of oxygen atoms. The most oxygen-deficient members of the homologous series can be classified as lipids. Thus, for the purpose of comparison to NMR data, we classify such molecules as lipids, and all others as carbohydrates. This distinction is represented in the van Krevelen diagram (Figure 3) as a soft division between lipid and carbohydrate regions.

Char carbon is defined by an aromaticity index ($f_a > 0.7$). These molecules are thought to consist primarily of water-stable byproducts of charcoal and soot decomposition in soils. Charcoal decomposition byproducts have been widely detected in dissolved NOM and can be distinguished by their condensed aromatic structures (Kim et al. 2003a, 2004; Hockaday et al. 2006, 2007). Therefore, we refer to charcoal carbon and condensed aromatic structures interchangeably. The aromaticity index considers carbon–nitrogen bonds and carbon–oxygen double bonds. Therefore, condensed aromatic structures do not have an explicit range of H/C values. This feature is depicted in the van Krevelen diagram (Figure 3) as a soft line dividing lignins and condensed aromatic structures.

Although they are minor components of this sample (≤5% of the integrated spectral magnitude), several hundred species are identified as *unsaturated hydrocarbons*. They comprise a mixture of aliphatic and aromatic or alkene structures with relatively little oxygen (O/C < 0.1). Plausible examples include plant pigments such as chlorophylls and carotenoids which contain both aromatic and/or pyrrolic and long-chain aliphatic (e.g., phytol) moieties. For the purpose of comparison to the ^{13}C NMR mixing model in Table 4, unsaturated hydrocarbons with $f_a < 0.7$ are classified as lipids.

The van Krevelen diagrams summarizing the MS datasets are shown in Figure 4. The van Krevelen diagrams show that most molecules ionized by positive-ion APPI have lower H/C and O/C ratios than molecules ionized by positive- or negative-ion ESI. Comparison of Figures 3 and 4 reveals that most molecular species identified by positive-ion APPI MS have elemental compositions similar to proteins, lignins, and unsaturated hydrocarbons, whereas positive- and negative-ion ESI and ESI MS identify more species in the composition-space of lipids and carbohydrates. The red points in the $APPI^+$ data represent molecular cation radicals generated via direct photoionization. They clearly occupy the composition-space of aromatic compounds (condensed hydrocarbons and lignins/tannins). The majority of APPI MS data points, shown in black, are chemical ions, $(M+H)^+$, generated by proton transfer.

The van Krevelen diagrams also provide the opportunity to visualize the homologous series of CH_2 and oxygen that are identified in the *zmo* table (Table 2). The CH_2 series are manifest as points falling on conspicuous diagonal lines originating from point 2.0 on the H/C axis and sloping downward across the diagram. For a CH_2 series with constant *o* value, the slope of the diagonal is controlled by the *z* value. Increasingly negative *z* values give rise to increasingly negative slopes. Homologous oxygen series also stand out in the van Krevelen diagram as points falling along horizontal lines. For a given *m* value, the H/C intercept is controlled by *z*. Oxygen series of increasingly negative *z* values have lower H/C ratios.

Table 4 summarizes the results of the van Krevelen analyses in a quantitative fashion, for direct comparison to ^{13}C NMR mixing model results. Table 4 lists, for each MS ion source, the mass percent and number average molecular species according to biomolecule type. Discrepancies in biomolecule concentrations measured by FT-ICR MS and ^{13}C NMR can be interpreted

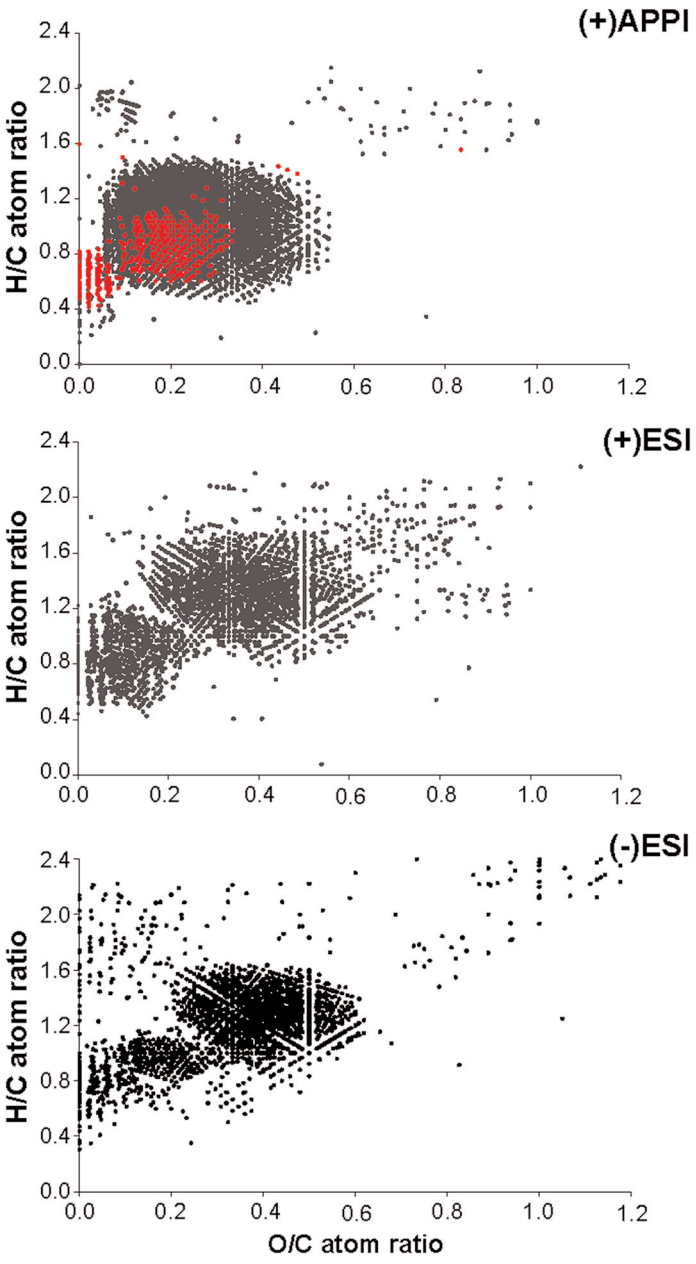

Fig. 4. Van Krevelen diagrams of Great Dismal Swamp NOM molecular components determined by (+)APPI MS (upper), (+)ESI MS (middle), and (–)ESI MS (lower). Red points in the APPI data represent radical cations ($M^{\cdot+}$), and black points are chemical ions ($M + H^+$).

as differences in the relative ionization efficiencies of the biomolecules. The most obvious discrepancy is the underrepresentation of carbohydrates by all 3 ion sources, relative to ^{13}C NMR. The second obvious trend is the preferential ionization of polyphenols (lignins), especially by APPI. Nevertheless, Table 4 conveys an important result that the combined APPI and ESI MS results (MS mean in Table 4) agree with ^{13}C NMR in the relative concentrations of non-carbohydrate biomolecules (lignins > lipids > peptides).

Spectral magnitude is an important factor controlling the relative abundances of different classes of biomolecules in Table 4. We define spectral magnitude as the ion current generated by molecular species that could be assigned unique molecular formulas. As explained previously, spectral magnitude can be affected by various phenomena that ultimately determine ionization efficiency. Therefore, Table 4 also provides a comparison of relative biomolecule abundance in a manner independent of ionization efficiency. Parenthetical values in Table 4 are numerical averages that do not consider ion magnitudes (peak heights). Comparison of the numerical average values minimizes the effects of intermolecular charge competition during ionization and more directly reflects the selectivity of the ionization mechanisms. However, information about the relative concentrations of individual molecular species is excluded from the numerical average values. Nevertheless, for all three ionization techniques, biomolecule number averages are in closer agreement with ^{13}C NMR data than abundance-weighted values. The positive-ion ESI MS number average biomolecule concentrations are exceptionally similar to values determined by ^{13}C NMR and the molecular mixing model. These results suggest that peak heights in APPI and ESI mass spectra of NOM samples are not reliable for the estimation of the relative concentrations of individual molecular species.

(3) Structural characteristics—A second means of evaluating ESI and APPI MS data draws on NMR-generated structural information. Integration of the ^{13}C NMR spectrum in Figure 5 suggests that the sample is composed of 30% H-substituted alkyl C (CH_3, CH_2, and CH), 28% O- and N-substituted alkyl carbon, 15% aromatic C, 7% O-substituted aromatic C, 15% carboxyl/amide C, and 5% carbonyl C. Though FT-ICR MS data can be treated by KMD analysis to identify molecules that differ by the exact mass of a CH_2 or CO_2 group, they provide no information about the rest of the molecule. Therefore, KMD analysis does not allow for a rigorous comparison to NMR functional group data. On the other hand, one piece of bulk structural information that can be calculated from MS data is aromaticity. In the determination of chemical structures from molecular formulas (e.g., $C_cH_hN_nO_oS_s$), it is often useful to calculate the number of DBEs via Equation 4. The total aromaticity of the sample, or the aromatic fraction (f_a), can then be obtained by summing the DBE of all molecular species, *i*, and dividing by the total number of carbon atoms, *c*, and nitrogen atoms, *n* (Equation 5).

$$f_a = \frac{\sum_i DBE}{\sum_i c + n} \quad (5)$$

Similarly, in the ^{13}C NMR spectrum, carbon atoms in double bonds (C=C, C=O, and C=N) and aromatic rings have ^{13}C NMR chemical shift values of 110–220 ppm. Thus, total aromaticity is directly obtained by integration of those NMR peak areas and normalizing to total area (Equation 6).

$$f_a = \frac{\text{integrated } ^{13}\text{CNMR peak area from 110 to 220 ppm}}{\text{total integrated } ^{13}\text{C NMR peak area from 0 to 220 ppm}} \quad (6)$$

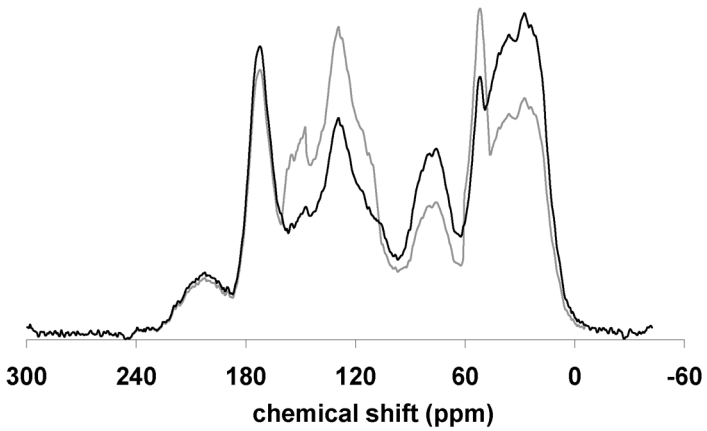

Fig. 5. ^{13}C CPMAS NMR spectrum of Lake Drummond NOM (black). The spectrum shown in gray is calculated from the sum of MS data (from Table 4) by running the molecular mixing model (Baldock et al. 2004) in reverse.

Aromaticity, f_a, can be also calculated from combustion elemental analysis data (with Equations 3–5).

The NOM aromaticity values determined by positive- and negative-ion ESI, shown in Table 3, agree rather closely with NMR and EA (0.42–0.46). The positive ions generated by APPI are significantly more aromatic (f_a = 0.53). This is to be expected because double bonds and aromatic rings are strong absorbers of UV photons.

As a means of quantitative comparison between MS and ^{13}C NMR, we use the MS data in Table 4 to run the molecular mixing model in reverse (the final step in the analysis scheme, Figure 1). Starting with the number average molecular composition determined collectively by APPI and ESI MS (parenthetical values in Table 4), the molecular mixing model was used to predict the signal distribution for a ^{13}C NMR spectrum. The result is shown as the gray NMR trace in Figure 5. The results again indicate that the APPI and electrospray techniques ionize molecules with aromatic functionality more effectively than molecules with alkyl and *O*-alkyl functionality. Nevertheless, approximately 74% of the NMR signal calculated from MS data coincides with the measured ^{13}C NMR signal. Assuming that ^{13}C NMR provides a reliable measure of the carbon functional group distribution, the differences in peak areas between the modeled and the measured ^{13}C NMR spectra can be interpreted as the *relative quantitation error* of the APPI and ESI MS data. Thus, mass spectrometry overestimates aromatic carbon content of the NOM by 13% and coincidentally underestimates alkyl and *O*-alkyl carbon content by 13%.

Discussion

With the exception of underestimating carbohydrates, ESI MS identified a suite of molecules whose elemental and structural composition agree closely with bulk elemental and structural properties measured by combustion elemental analysis and ^{13}C NMR. APPI MS revealed some species with the same molecular weight as identified by ESI, and an additional 3000 molecular species that were not identified by positive- or negative-ion ESI. These results reveal a dichotomy. Although APPI provides important insight to the composition of NOM by gaining access to species not ionized by electrospray, both ESI and APPI are highly selective in the types of molecules ionized. Species of the same molecular weight and molecular formula observed in both APPI and ESI mass spectra can (and often do) have different molecular structures. Thus, further study of ionization efficiencies of reference compounds and matrix effects is needed before quantitative interpretations of relative ion abundance are warranted.

Although we cannot place an exact value on the fraction of NOM detected in FT-ICR mass spectra, comparison to ^{13}C NMR data treated with the molecular mixing model provides some constraints on quantitation. Neither APPI nor ESI MS provide substantial information on the carbohydrate fraction that comprises ~9% of the organic carbon in the Lake Drummond NOM, according to ^{13}C NMR mixing model results, suggesting that at least 9% of the C_{18}-extractable NOM remains undetected. The inverse mixing model approach indicates that about 13% of the NOM is either undetected or underrepresented, and comprises mostly alkyl and *O*-alkyl carbon. Nevertheless, the method comparison reveals that collectively, ESI and APPI MS data are representative of a large portion of non-carbohydrate NOM in Lake Drummond.

Comments and recommendations

Due to the complementary natures of APPI and ESI, we recommend that both ion sources be applied in future investigations of NOM structure and composition. FT-ICR MS provides valuable information about NOM that was previously uncharacterized at the molecular level. However, a conundrum remains. If the composition of a complex mixture is unknown, how does one evaluate the performance of the spectroscopy used to characterize the unknown? Our approach in this article has been to compare results with more quantitative bulk techniques. This is a valuable first step, but a more systematic approach to performance evaluation is needed. Even if quantitation is not necessarily the goal, characterizing NOM at the molecular level requires, first and foremost, characterization of spectroscopic performance at the molecular level.

Our preliminary investigation reveals that selective ionization is an important issue for NOM samples. We have evaluated a single NOM sample, and selective ionization is likely to have different effects on other NOM samples of differing compositions. Therefore, in an effort to better interrogate the FT-ICR database on NOM in natural waters, selective ionization should be investigated systematically across a range of different samples. We recommend a standard addition approach whereby a suite of pure chemical compounds with systematic variations in molecular structures (analogous to

those of biomolecules) are added in known proportions to NOM samples before FT-ICR MS analysis. The relative abundances of internal standards would facilitate a more rigorous evaluation of ion source behavior and selective ionization. Because charge competition among analytes is an unavoidable reality, NOM samples of differing compositions will be differentially influenced by matrix effects. Therefore, the eventual creation of a widely available internal standard (spike) for FT-ICR MS analysis of NOM may be warranted—borrowing from the approach taken in isotope ratio mass spectrometry. This could serve the added purpose of mass calibration. Such an approach would make more robust the comparison of datasets collected by different investigators or different instruments.

In addition to selective ionization, the process of extracting and concentrating organic matter from natural waters can be selective. The C_{18} solid-phase extraction (SPE) was used here because it is relatively efficient in NOM extraction from freshwater samples (60% to 80% recovery) (Roubeuf et al. 2000; Kim et al. 2003b). SPE is also an effective means of removing inorganic ions, which can give rise to matrix effects and ion suppression. However, all SPE resins are inherently selective because retention is based on intermolecular forces, as was recently demonstrated by a study of Dismal Swamp water (Sleighter and Hatcher 2008) analyzed directly by ESI without preconcentration. It was clear that water-soluble tannins and amides/amines are excluded by SPE. Therefore, SPE extracts do not necessarily collect the most representative sample of NOM at the molecular level (Louchouarn et al. 2000). More quantitative (and therefore less selective) means of isolating dissolved NOM are now available. Reverse osmosis and electrodialysis have been combined to quantitatively extract organic matter while also deionizing the sample (Koprivnjak et al. 2006; Vetter et al. 2007; Ouellet et al. 2008). Understanding the molecular-level fractionation that occurs during sample preparation and analysis becomes increasingly important as we move toward a more quantitative understanding of the molecular composition and reactivity of NOM in natural waters.

References

Baldock, J. A., C. A. Masiello, Y. Gelinas, and J. I. Hedges. 2004. Cycling and composition of organic matter in terrestrial and marine ecosystems. Mar. Chem. 92:39-64.

Brown, T. L., and J. A. Rice. 2000. Effect of experimental parameters on the ESI FT-ICR mass spectrum of fulvic acid. Anal. Chem. 72:384-390.

Burdon, J. 2001. Are the traditional concepts of the structures of humic substances realistic? Soil Sci. 166:752-769.

Cai, S. S., and J. A. Syage. 2006. Comparison of Atmospheric Pressure Photoionization, Atmospheric Pressure Chemical Ionization, and Electrospray Ionization Mass Spectrometry for Analysis of Lipids. 78:1191-1199.

Cai, Y. X., D. Kingery, O. McConnell, and A. C. Bach. 2005. Advantages of atmospheric pressure photoionization mass spectrometry in support of drug discovery. Rapid Comm. Mass Spectrom. 19:1717-1724.

Cardoza, L. A., A. K. Korir, W. H. Otto, C. J. Wurrey, and C. K. Larive. 2004. Applications of NMR spectroscopy in environmental science. Prog. Nucl. Magn. Res. Spectr. 45:209-238.

Dickens, A. F., and others 2006. Solid-state ^{13}C NMR analysis of size and density fractions of marine sediments: insight into organic carbon sources and preservation mechanisms. Geochim. Cosmochim. Acta 70:666-686.

Dittmar, T., and B. P. Koch. 2006. Thermogenic organic matter dissolved in the abyssal ocean. Mar. Chem. 102:208-217.

Friedel, J. K., and E. Scheller. 2002. Composition of hydrolysable amino acids in soil organic matter and soil microbial biomass. Soil Biol. Biochem. 34:315-325.

Gomez-Ariza, J. L., T. Garcia-Barrera, and F. Lorenzo. 2006. Anthocyanins profile as fingerprint of wines using atmospheric pressure photoionisation coupled to quadrupole time-of-flight mass spectrometry. Anal. Chim. Acta 570:101-108.

Grannas, A. M., W. C. Hockaday, P. G. Hatcher, L. G. Thompson, and E. Mosley-Thompson. 2006. New revelations on the nature of organic matter in ice cores. J. Geophys. Res. Atmos. 111, D04304, doi:10.1029/2005JD006251. Swift 1999. 164:790-802.

Hanold, K. A., S. M. Fischer, P. H. Cormia, C. E. Miller, and J. A. Syage. 2004. Atmospheric Pressure Photoionization. Chapter 1, General Properties for LC/MS, p. 2842-2851.Q1

Hedges, J. I., J. A. Baldock, Y. Gelinas, C. Lee, M. L. Peterson, and S. G. Wakeham. 2002. The biochemical and elemental compositions of marine plankton: a NMR perspective. Mar. Chem. 78:47-63.

——— and others. 2000. The molecularly-uncharacterized component of nonliving organic matter in natural environments. Organ. Geochem. 31:945-958.

———, R. G. Keil, and R. Benner. 1997. What happens to terrestrial organic matter in the ocean? Organ. Geochem. 27:195-212.

Hertkorn, N., and others 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochim. Cosmochim. Acta 70:2990-3010.

Hiroshi, M., I. Miho, Y. Shisaku, M. Hidezaku, and A. Jean-Francois. 2004. Determination of polycyclic aromatic hydrocarbons in sediment by liquid chromatography atmospheric pressure photoionization mass spectrometry. Anal. Sci. 20:375-377.

Hockaday, W. C., A. M. Grannas, S. Kim, and P. G. Hatcher. 2006. Direct molecular evidence for the degradation and mobility of black carbon in soils from ultrahigh-resolution mass spectral analysis of dissolved organic matter from a fire-impacted forest soil. Organ. Geochem. 37:501-510.

———, A. M. Grannas, S. Kim, and P. G. Hatcher. 2007. The transformation and mobility of charcoal in a fire-impacted watershed. Geochim. Cosmochim. Acta 71:3432-3445.

Hsu, C. S., K. Qian, and Y. C. Chen. 1992. An innovative

approach to data analysis in hydrocarbon characterization by on-line liquid chromatography mass spectrometry. Anal. Chim. Acta 264:79-89.

Hsu, P.-H., and P. G. Hatcher. 2005. New evidence of covalent coupling of peptides to humic acids based on 2D NMR spectroscopy: a means for preservation. Geochim. Cosmochim. Acta 69:4521-4533.

Johannesson, K. H., J. W. Tang, J. M. Daniels, W. J. Bounds, and D. J. Burdige. 2004. Rare earth element concentrations and speciation in organic-rich blackwaters of the Great Dismal Swamp, Virginia, USA. Chem. Geol. 209:271-294.

Kaufman, P. B., L. J. Cseke, S. Warber, J. A. Duke, and H. L. Brielmann. 1999. Natural Products from Plants. CRC Press, Boca Raton, FL.

Kebarle, P., and M. Pesche. 2000. On the mechanism by which the charged droplets produced by electrospray lead to gas phase ions. Anal. Chim. Acta 406:11-35.

Kendrick, E. 1963. A mass scale based on CH_2 = 14.0000 for high resolution mass spectrometry of organic compounds. Anal. Chem. 35:2146-2154.

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

———, R. W. Kramer, and P. G. Hatcher. 2003a. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. Anal. Chem. 75:5336-5344.

———, A. J. Simpson, E. B. Kujawinski, M. A. Freitas, and P. G. Hatcher. 2003b. High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C-18 solid phase disk. Organ. Geochem. 34:1325-1335.

Kim, S. W., L. A. Kaplan, R. Benner, and P. G. Hatcher. 2004. Hydrogen-deficient molecules in natural riverine water samples: evidence for the existence of black carbon in DOM. Mar. Chem. 92:225-234.

Koch, B. P., T. Dittmar, M. Witt, and G. Kattner. 2007. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. Anal. Chem. 79:1758-1763.

———, M. R. Witt, R. Engbrodt, T. Dittmar, and G. Kattner. 2005. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Geochim. Cosmochim. Acta 69:3299-3308.

Koprivnjak, J. F., E. M. Perdue, and P. H. Pfromm. 2006. Coupling reverse osmosis with electrodialysis to isolate natural organic matter from fresh waters. Water Res. 40:3385-3392.

Kramer, R. W., E. B. Kujawinski, and P. G. Hatcher. 2004. Identification of black carbon derived structures in a volcanic ash soil humic acid by Fourier transform ion cyclotron resonance mass spectrometry. Env. Sci. Technol. 38:3387-3395.

Kujawinski, E. B., and M. D. Behn. 2006. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. Anal. Chem. 78:4363-4373.

———, R. Del Vecchio, N. V. Blough, G. C. Klein, and A. G. Marshall. 2004. 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. 92:23-37.

———, M. A. Freitas, X. Zang, P. G. Hatcher, K. B. Green-Church, and R. B. Jones. 2002. The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. Organ. Geochem. 33:171-180.

Ledford, E. B., D. L. Rempel, and M. L. Gross. 1984. Space charge effects in Fourier transform mass spectrometry mass calibration. Anal. Chem. 56:2744-2748.

Leri, A. C., M. A. Marcus, and S. C. B. Myneni. 2007. X-ray spectromicroscopic investigation of natural organochlorine distribution in weathering plant material. Geochim. Cosmochim. Acta 71:5834-5846.

Louchouarn, P., S. Opsahl, and R. Benner. 2000. Isolation and quantification of dissolved lignin from natural waters using solid-phase extraction and GC/MS. Anal. Chem. 72:2780-2787.

Marshall, A. G., C. L. Hendrickson, and G. Jackson. 1998. Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom. Rev. 17:1-35.

Metz, G., X. Wu, and S. O. Smith. 1994. Ramped amplitude cross polarization in magic angle spinning NMR. J. Magn. Res. 110:219-227.

Mopper, K., A. Stubbins, J. D. Ritchie, H. M. Bialk, and P. G. Hatcher. 2007. Advanced instrumental approaches for characterization of marine dissolved organic matter: extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. Chem. Rev. 107:419-442.

Mullins, O. C., B. Martinex-Haya, and A. G. Marshall. 2008. Contrasting perspective on asphaltene molecular weight. Energy Fuels 22:1765-1773.

Mutabaruka, R., K. Hairiah, and G. Cadisch. 2007. Microbial degradation of hydrolysable and condensed tannin polyphenol-protein complexes in soils from different land-use histories. Soil Biol. Biochem. 39:1479-1492.

Myneni, S. C. B. 2002. Formation of stable chlorinated hydrocarbons in weathering plant material. Science 295:1039-1041.

Nelson, P. N., and J. A. Baldock. 2005. Estimating the molecular composition of a diverse range of natural organic materials from solid-state C-13 NMR and elemental analyses. Biogeochemistry 72:1-34.

———, J. A. Baldock, P. Clarke, J. M. Oades, and G. J. Churchman. 1999. Dispersed clay and organic matter in soil: their natura and associations. Austral. J. Soil Res. 37:289-315.

Olk, D. C., K. G. Cassman, K. Schmidt-Rohr, M. M. Anders, J. D. Mao, and J. L. Deenik. 2006. Chemical stabilization of soil organic nitrogen by phenolic lignin residues in anaerobic agroecosystems. Soil Biol. Biochem. 38:3303-3312.

O'Loughlin, E., and Y. P. Chin. 2001. Effect of detector wavelength on the determination of the molecular weight of humic substances by high-pressure size exclusion chromatography. Water Res. 35:333-338.

Ouellet, A., C. Dragosh, J.-P. Plouhinec, M. Lucotte, and Y. Gelinas. 2008. Elemental, isotopic, and spectroscopic assessment of chemical fractionation of dissolved organic matter sampled with a portable reverse osmosis system. Env. Sci. Technol. 42:2490-2495.

Perdue, E. M., N. Hertkorn, and A. Kettrup. 2007. Substitution patterns in aromatic rings by increment analysis: model development and application to natural organic matter. Anal. Chem. 79:1010-1021.

Piccolo, A. 2001. The supramolecular structure of humic substances. Soil Sci. 166:810-832.

———, P. Conte, E. Trivellone, and B. Van Lagen. 2002. Reduced heterogeneity of a lignite humic acid by preparative HPSEC following interaction with an organic acid: characterization of size-separates by Pyr-GC-MS and H-1-NMR spectroscopy. Env. Sci. Technol. 36:76-84.

Purcell, J. M., C. L. Hendrickson, R. P. Rodgers, and A. G. Marshall. 2006. Atmospheric pressure photoionization Fourier transform ion cyclotron resonance mass spectrometry for complex mixture analysis. Anal. Chem. 78:5906-5912.

———. 2007. Atmospheric pressure photoionization proton transfer for complex organic mixtures investigated by Fourier transform ion cyclotron resonance mass spectrometry. J. Am. Soc. Mass Spectrom. 18:1682-1689.

Reemtsma, T., A. These, M. Linsheid, J. A. Leenheer, and A. Spitzy. 2008. Molecular and structural characterization of dissolved organic matter from the deep ocean by FTICR-MS, including hydrophilic nitrogenous organic molecules. Env. Sci. Technol. 42:1430-1437.

Repeta, D. J., N. T. Hartman, S. John, A. D. Jones, and R. Goericke. 2004. Structure elucidation and characterization of polychlorinated biphenyl carboxylic acids as major constituents of chromophoric dissolved organic matter in Seawater. Env. Sci. Technol. 38:5373-5378.

Robb, D. B., T. R. Covey, and A. P. Bruins. 2000. Atmospheric pressure photoionization: an ionization method for liquid chromatography. Anal. Chem. 72:3653-3659.

Romero, J. T. [ed]. 2005. Chemical Ecology and Phytochemistry of Forest Ecosystems. Elsevier, Amsterdam.

Roubeuf, V., S. Mounier, and J. Y. Benaim. 2000. Solid phase extraction applied to natural waters: efficiency and selectivity. Organ. Geochem. 31:127-131.

Roy, S., and others 2006. Liquid chromatography on porous graphitic carbon with atmospheric pressure photoionization and tandem mass spectrometry for the analysis of glycosphingolipids. J. Chromatogr. A 1117:154-132.

Senko, M. W., J. D. Canterbury, S. H. Guan, and A. G. Marshall. 1996. A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry. Rapid Comm. Mass Spectrom. 10:1839-1844.

Shi, S. D.-H., J. J. Drader, M. A. Freitas, C. L. Hendrickson, and A. G. Marshall. 2000. Comparison and interconversion of the two most common frequency-to-mass calibration functions for Fourier transform ion cyclotron resonance mass spectrometry. Int. J. Mass Spectrom. 196:591-598.

Sleighter, R. L., and P. G. Hatcher. 2008. 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. 110:140-152.

Smejkalova, D., and A. Piccolo. 2008. Aggregation and disaggregation of humic supramolecular assemblies by NMR diffusion ordered Spectroscopy (DOSY-NMR). Env. Sci. Technol. 42:699-706.

Stenson, A. C., W. M. Landing, A. G. Marshall, and W. T. Cooper. 2002. Ionization and fragmentation of humic substances in electrospray ionization Fourier transform-ion cyclotron resonance mass spectrometry. Anal. Chem. 74:4397-4409.

Straube, E. A., W. Dekant, and W. Volkel. 2004. Comparison of electrospray ionization, atmospheric pressure chemical ionization, and atmospheric pressure photoionization for the analysis of dinitropyrene and aminonitropyrene LC-MS/MS. J. Am. Soc. Mass Spectrom. 15:1853-1862.

Sutton, R., and G. Sposito. 2005. Molecular Structure in Soil Humic Substances: The New View. 39:9009-9015.

Swift, R. S. 1999. Macromolecular properties of soil humic substances: fact, fiction, and opinion. Soil Sci. 164:790-802.

Takino, M., S. Daishima, and T. Nakahara. 2003. Determination of chloramphenicol residues in fish meats by liquid chromatography-atmospheric pressure photoionization mass spectrometry. J. Chromatogr. A 1011:67-75.

Teuten, E. L., and C. M. Reddy. 2007. Halogenated organic compounds in archived whale oil: a pre-industrial record. Env. Pollut. 145:668-671.

Van Krevelen, D. W. 1950. Graphical-statistical method for the study of structure and reaction processes of coal. Fuel 26:269-284.

Vetter, T. A., E. M. Perdue, E. Ingall, J. F. Koprivnjak, and P. H. Pfromm. 2007. Combining reverse osmosis and electrodialysis for more complete recovery of dissolved organic matter from seawater. Separ. Purif. Technol. 56:383-387.

Wrobel, K., B. B. M. Sadi, K. Wrobel, J. R. Castillo, and J. A. Caruso. 2003. Effect of metal ions on the molecular weight distribution of humic substances derived from municipal compost: ultrafiltration and size exclusion chromatography with spectrophotometric and inductively coupled plasma-MS detection. Anal. Chem. 75:761-767.

Yoshioka, N., Y. Akiyama, and K. Teranishi. 2004. Rapid simul-

taneous determination of *o*-phenylphenol, diphenyl, thiabendazole, imazalil and its major metabolite in citrus fruits by liquid chromatography-mass spectrometry using atmospheric pressure photoionization. J. Chromatogr. A 1022:145-150.

Zheng, J., and S. A. Shamsi. 2006. Capillary Electrochromatography Coupled to Atmospheric Pressure Photoionization Mass Spectrometry for Methylated Benzo[a]pyrene Isomers. 78:6921-6927.

Submitted 21 May 2008
Revised 19 November 2008
Accepted 15 December 2008