

Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry

Elizabeth B. Kujawinski¹ (*), Krista Longnecker¹, Neil V. Blough², Rossana Del Vecchio³, Liam Finlay⁴, Joshua B. Kitner⁴ and Stephen J. Giovannoni⁴

(*) Corresponding author – Department of Marine Chemistry and Geochemistry; Woods Hole Oceanographic Institution; 360 Woods Hole Rd. MS #4; Woods Hole, MA 02543; 508-289-3493; ekujawinski@whoi.edu

1 – Department of Marine Chemistry & Geochemistry; Woods Hole Oceanographic Institution; Woods Hole, MA 02543

2 – Department of Chemistry and Biochemistry; University of Maryland; College Park MD 20742

3 – Earth System Science Interdisciplinary Center; University of Maryland; College Park MD 20742

4 – Department of Microbiology; Oregon State University; Corvallis OR 97331

Submitted to *Geochimica et Cosmochimica Acta*: 7/30/08

Manuscript #: W5976

Returned: 10/15/08

Revised and resubmitted: 2/3/09

Returned: 3/10/09

Revised and resubmitted: 4/2/09

Returned: 4/30/09

Revised and resubmitted: 4/30/09

Accepted: 4/30/09

ABSTRACT

Marine dissolved organic matter (DOM) is one of the most heterogeneous and largest pools of reactive carbon on earth, rivaling in mass the carbon in atmospheric carbon dioxide.

Nevertheless, the molecular-level composition of marine DOM has eluded detailed description, impeding inquiry into the specific mechanisms that add or remove compounds from the DOM pool. Here we describe the molecular-level composition of C₁₈-extracted DOM along an east-west transect of the North Atlantic Ocean. We examine the changes in DOM composition along this transect with ultrahigh resolution mass spectrometry and multivariate statistics. We use indicator species analysis (ISA) to identify possible source markers for photochemical

degradation and heterotrophic bacterial metabolism. The inclusion of ISA in statistical evaluation of DOM mass spectral data allows investigators to determine the m/z values associated with significant changes in DOM composition. With this technique, we observe indicator m/z values in estuarine water that may represent components of terrestrially-derived chromophoric DOM subject to photo-chemical degradation. We also observe a unique set of m/z values in surface seawater and show that many of these are present in pure cultures of the marine α -proteobacterium *Candidatus Pelagibacter* ubique when grown in natural seawater. These findings indicate that a complex balance of abiotic and biotic processes controls the molecular composition of marine DOM to produce signatures that are characteristic of different environments.

1 INTRODUCTION

Dissolved organic material (DOM) is the most heterogeneous and dynamic pool of carbon in the oceans. DOM plays a fundamental role in the global carbon cycle as one of the largest reservoirs of reduced carbon on the earth's surface. At ~700 Tg it is comparable in magnitude to atmospheric carbon dioxide (~750 Tg - HEDGES, 2002). Bulk measurements and compound-specific assays have shown that the concentration and composition of DOM are affected by numerous biotic and abiotic processes such as photosynthesis (MARANON et al., 2004), heterotrophic microbial metabolism (AZAM and CHO, 1987) and photochemistry (MOPPER et al., 1991). These processes are inextricably linked, each affecting individual components of DOM to a different extent, culminating in the observed heterogeneity of DOM (NAGATA, 2000; OBERNOSTERER and BENNER, 2004; MCCALLISTER et al., 2005).

Elucidation of the molecular structure of DOM components is critical for a mechanistic understanding of the global carbon cycle and thus has been the subject of scientific inquiry for decades (HEDGES, 2002). Spectroscopic techniques have been effectively employed to examine bulk (or aggregate) changes within DOM and its fractions. For example, nuclear magnetic resonance (NMR) spectroscopy was used to characterize the composition of functional groups in bulk DOM (HATCHER et al., 1980). Later, absorption and fluorescence spectroscopy provided key information on the wide-scale distribution of chromophoric DOM (CDOM), its effect on the aquatic light field and its photochemical fate (BLOUGH and DEL VECCHIO, 2002; NELSON and SIEGEL, 2002). These techniques, however, are limited in their ability to probe the contributions and dynamics of individual molecules.

Examination of intact individual molecules in DOM has proved challenging, leading researchers to focus on analyses of biopolymer sub-units such as lignin phenols (MEYERS-

SCHULTE and HEDGES, 1986), amino acids (AMON et al., 2001) and neutral sugars (ALUWIHARE et al., 2002) to gauge the overall quantity and reactivity of the structurally-diverse lignins, proteins and polysaccharides. However, the dynamics of biopolymer subunits does not fully represent the chemistry of the parent macromolecules. Non-polar molecules such as lipids and n-alkanes have been analyzed directly, without fragmentation, by gas chromatography (GC). Concentrations and transformation rates of these compounds have provided tantalizing insights into the DOM cycle (e.g., MANNINO and HARVEY, 1999), but these compounds are a minor component of the overall DOM pool. Until recently, comparable analytical tools for polar and semi-polar molecules within DOM have been missing.

The advent of electrospray ionization coupled to mass spectrometry has provided the opportunity to characterize intact polar molecules within DOM and to explore their reactivity within biogeochemical processes. Electrospray ionization (ESI) is a “soft” ionization technique with low incidence of fragmentation for natural organic matter (NOM) molecules (ROSTAD and LEENHEER, 2004). ESI coupled to ultrahigh resolution instruments such as Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometers has been used to characterize NOM collected from freshwater systems (e.g., KIM et al., 2006a; SLEIGHTER and HATCHER, 2008; SLEIGHTER et al., 2008), the coastal ocean (e.g., TREMBLAY et al., 2007), the open ocean (e.g., DITTMAR and KOCH, 2006; KOCH et al., 2008), and laboratory-based biogeochemical studies (e.g., KUJAWINSKI et al., 2004). Altogether, these investigations have provided unprecedented detail regarding the composition of thousands of individual compounds within the polar fraction of DOM (KUJAWINSKI et al., 2002; STENSON et al., 2003; KOCH et al., 2005). Like the others listed above, this technique has limitations as a tool for DOM characterization. Compounds that are not ions in aqueous solution are not detected, and ancillary analyses such as MS/MS fragmentation are

required to identify structural isomers. Nonetheless, the ultrahigh resolution and mass accuracy of ESI FT-ICR MS provides molecular masses that are accurate to within 1ppm, which often enables the determination of elemental formulae from the mass measurement alone (KIM et al., 2006b). Thus, ESI FT-ICR MS can be used effectively to detect mass changes within a suite of DOM molecules and subsequently to resolve the molecular-level impact of different biogeochemical processes on DOM composition. Here we examine two such processes, photochemistry and microbial metabolism, in marine DOM.

Heterotrophic bacterial metabolism and photochemistry are arguably two of the most important biogeochemical pathways for transforming organic matter in the surface oceans (HANSELL and CARLSON, 2002) and are often inter-dependent (MORAN and ZEPP, 1997; MOPPER and KIEBER, 2002). Photochemical degradation of terrestrial chromophoric DOM is an important removal process within coastal environments, but has been difficult to study on a molecular level. CDOM substantially affects the aquatic light field, but lack of structural information has limited understanding of the reactions and rates that govern CDOM distribution. Previous work has shown that CDOM along the North Atlantic margin is derived primarily from terrestrial sources and its primary sink is photodegradation (VODACEK et al., 1997; DEL VECCHIO and BLOUGH, 2004b; VAILLANCOURT et al., 2005). Terrestrially-derived CDOM is largely resistant to microbial degradation (MORAN et al., 2000). However, during photochemical degradation, low-molecular-weight compounds (KIEBER, 2000) and nutrients (BUSHAW et al., 1996; MOPPER and KIEBER, 2002) are commonly produced with a concomitant decrease in the molecular size of CDOM. Many photochemical products are readily consumed by bacteria (KIEBER et al., 1989) and thus stimulate bacterial growth (MOPPER and KIEBER, 2002). Photochemistry can also inhibit

the microbial consumption of algal-derived DOM (BENNER and BIDDANDA, 1998; TRANVIK and KOKALJ, 1998), presumably through structural modifications of existing biomolecules.

Ecological theory presupposes that diversification of microbial taxa can be a consequence of resource specialization, but very little is known about interactions between specific microorganisms and the field of compounds that comprise DOM. Some studies have examined the production of DOM by microbes with bulk measurements or compound-specific assays (see review in NAGATA, 2000). Detailed analyses of biologically-produced DOM have been constrained by analytical challenges and thus have focused on compounds such as amino acids, sugars and other biopolymer subunits. The study of bacterial utilization of DOM has been limited similarly. Some studies examined decreases in bulk DOM concentrations or the loss of particular substrates. Although these studies yielded insights into DOM cycling, they lacked the power to broadly resolve new and unforeseen interactions between marine microorganisms and specific compounds.

Ultrahigh resolution mass spectrometry such as ESI FT-ICR MS is the first tool that has the power to broadly resolve biogeochemical alteration of DOM at a molecular level. Numerous investigators have now used ESI FT-ICR MS to compare DOM from different sources in freshwater systems (e.g., TREMBLAY et al., 2007; SLEIGHTER and HATCHER, 2008; SLEIGHTER et al., 2008) and open ocean environments (e.g., DITTMAR and KOCH, 2006; KOCH et al., 2008). Here, we focus on those studies that examined photochemical or microbial degradation of DOM. For example, Kujawinski et al. (2004) showed that elemental formulae with relatively high aromatic character and low oxygen number were preferentially removed during photochemical degradation of riverine DOM. Likewise, aromatic compounds such as condensed hydrocarbons and lignin-derived humic materials were lost from mangrove DOM (TREMBLAY et al., 2007) and

riverine DOM (GONSIOR et al., 2009) during outwelling to coastal estuaries. Both sets of authors ascribe their observations to photochemical degradation during estuarine mixing.

In contrast to photochemistry, few studies on microbial utilization or production of DOM have utilized ESI FT-ICR MS. One such study showed that bacteria produce different DOM mass spectral signatures in the presence and absence of protozoan grazing (KUJAWINSKI et al., 2004). However, this work was conducted in laboratory culture and its results may not be representative of field conditions. Tantalizing evidence of the microbial impact on DOM composition was acquired with lower resolution ESI mass spectrometry (SEITZINGER et al., 2005), but this technology lacked the mass resolution to assign empirical formulae to the compounds involved in microbial-DOM interaction. In short, direct structural identification of the compounds within DOM that are utilized by bacteria, that absorb solar radiation, and that are produced as a result of microbial or photochemical processing has yet to be achieved, but is critical for a comprehensive understanding of DOM cycling within the oceans.

Here we combine ultrahigh resolution mass spectrometry with spectroscopy and microbiology to explore changes in C_{18} -extracted DOM composition along a transect of the North Atlantic Ocean. We isolated ~2200 unique m/z features with ultrahigh resolution mass spectrometry and used multivariate statistics to compare DOM composition across a gradient of terrestrial input. We adapted Indicator Species Analysis (ISA) to identify tentative markers for terrestrially-derived CDOM and microbial DOM. Elemental formulae were assigned to most of the marker m/z values and the resulting elemental compositions were consistent with previous models of photo-active molecules and microbial exudates. Many marker m/z values from the surface ocean samples were also present in DOM extracted from pure cultures of *Candidatus Pelagibacter ubique* grown in sterilized seawater.

2 METHODS

2.1 Cruise sample collection

Samples were collected on a cruise in September 2005 along an east-west transect from the head of the Delaware River to the Sargasso Sea (station locations in Table 1). Water was collected by Niskin bottles on a Conductivity-Temperature-Depth (CTD) rosette at selected depths (Table 1). Water was acidified to pH 2-3 with HCl and DOM was extracted with C₁₈ cartridges (Mega Bond Elut, by UTC). The cartridges were pretreated with 100 mL of high purity MeOH followed by 50 mL of acidified (pH 2) Milli-Q water prior to extraction. Each sample (20 L) was pre-filtered through a 0.2 µm bell-filter, acidified to pH 2, and pumped through the C₁₈ cartridge at 50 mL min⁻¹. Each cartridge was then rinsed with 1 L of acidified (pH 2) Milli-Q water to remove salts and stored in the refrigerator (4°C) until further processing. DOM was extracted with 50 mL of high purity MeOH: the first fraction (DOM eluted with the first 5 mL) was not employed for this analysis; the second fraction (DOM eluted with 45 mL of MeOH) was collected and evaporated to dryness under vacuum at 30-35°C. The dried material was redissolved in Milli-Q water, neutralized with diluted NaOH and stored frozen until further analysis. Other investigators have shown that 30-60% of DOM is extracted by this technique in riverine and open ocean environments (KIM et al., 2003b; TREMBLAY et al., 2007; DITTMAR et al., 2008). Higher extraction efficiencies have been reported for riverine samples compared to marine samples. We estimate a range of 30-50% extraction efficiency in our samples based on absorbance measurements (at 250-350 nm) of DOM pre- and post-extraction.

2.2 P. ubique sample collection

Candidatus Pelagibacter ubique (HTCC1062), a member of the SAR11 clade of α -proteobacteria (RAPPÉ et al., 2002), was grown in sterilized seawater (collected from the Oregon coast – LNHM medium) in 20-L polycarbonate carboys under light (12:12 light:dark cycle) and dark conditions (CONNON and GIOVANNONI, 2002). Cell growth was monitored until the culture reached maximum density at which time the cells were removed by filtration. DOM from 2 L subsamples of 0.2- μ m filtrate from each culture and a non-inoculated light control were extracted according to previously published methods (KIM et al., 2003b). The light control sample was collected at the same time as the growth culture samples. In brief, filtrate was acidified with concentrated HCl until pH values ranged between 2 and 3. The filtrate was then passed through two stacked 47-mm extraction disks; first a C₁₈-based disk and then a SDB-based disk. Extraction disks were conditioned according to manufacturer's instructions. Once the entire filtrate was passed through the disks, the disks were washed with 10-20 mL pH 2 nanopure water. DOM was collected from the C₁₈/SDB disks using 70% methanol:water. Extracts were concentrated by vacuum centrifugation, re-dissolved in a known volume of 70% methanol/water and stored frozen until analysis. Twenty liters of Milli-Q water was acidified and extracted with the combined C₁₈/SDB-disks for an extraction blank.

2.3 Optical characterization methods

A Hewlett Packard 8452A and a Shimadzu 2401-PC spectrophotometers were employed to acquire UV-visible absorption spectra. Absorption spectra were recorded against Milli-Q water over the range 200-800 nm. The absorption values at wavelengths greater than 650 nm were averaged to determine the baseline and this average was subtracted from spectra to correct for

small offsets of the baseline (GREEN and BLOUGH, 1994). Absorption coefficients at various wavelengths, $a(\lambda)$, were calculated as in Del Vecchio and Blough (2004b) and the absorption spectra were then fit to an exponential function, using a non-linear least squares fitting routine over the range 290-700 nm. DOC concentrations were determined with high-temperature combustion, following a method previously described (DEL VECCHIO and BLOUGH, 2004b).

Concentrations of lignin-derived phenols were measured on C₁₈-extracted DOM following a slightly modified protocol (HEDGES and ERTEL, 1982; GONI and MONTGOMERY, 2000; LOUCHOUARN et al., 2000). Briefly, samples were digested by CuO oxidation in a microwave oven (CEM MARS-5) at 150°C. Following digestion, a known amount of recovery standard (ethylvanillin) was added to each sample. High purity ethyl acetate (Burdick& Jackson) was used to extract lignin phenols to minimize any contamination from solvent. The ethyl acetate was carefully evaporated by rotary evaporation at 35°C. The dried material was redissolved in 100µL pyridine, amended with an internal standard (p-hydroxyphenyl acetic acid) and a silylating agent (100µL of Regisila (BSTFA) 1%TCMS (Regis Tech. Inc.)) and reacted in a water bath at 60°C for 10 min. Samples were then analyzed by gas chromatography employing a Shimadzu GC17A with a flame ionization detector and a 60m × 0.23mm (I.D.) × 0.25µm film thickness J&W DB-1 column. The flow rate of carrier gas (He) was set at 1.5 mL min⁻¹ and the split ratio was 1:13. The injector port and detector were maintained at 300°C and 280°C, respectively. The temperature program consisted of an initial temperature of 100°C, a ramp at 4°C min⁻¹ to 250°C, a ramp at 13°C min⁻¹ to 270°C, and a final hold at 270°C for 10 min.

2.4 FT-MS data acquisition

All samples and the extraction blank were analyzed on a 9.4 T electrospray ionization (ESI) Fourier-transform ion cyclotron (FT-ICR) mass spectrometer (SENKO et al., 1996b) at the National ICR Users' Facility at the National High Magnetic Field Laboratory (NHMFL) at Florida State University in Tallahassee FL. Samples were reconstituted in 75:25 MeOH/water with 1% NH₄OH and analyzed in negative ion mode. Base was added prior to analysis to promote negative ion formation and to avoid co-occurring complexes of Na⁺ and H⁺ with individual DOM molecules that are common in positive ion mode. Samples were infused into the ESI interface at 400-500 nL min⁻¹. Instrument parameters were optimized for each sample. The capillary needle voltages ranged from -1350 V to -2000 V. Ions were accumulated in the external octopole for 12-20 sec before transfer to the ion cyclotron cell. The two transfer octopole frequencies were set at 1.6 and 1.8 MHz. Data were collected (4 MWord) by a MIDAS data station (SENKO et al., 1996a). Numerous scans (200) were co-added prior to Hanning apodization, zero-fill and Fourier-transformation. The data was truncated once (to 2 MWord) due to insufficient signal at longer transient times. The instrument was calibrated daily with an external standard (ESI TOF Mix, Agilent Technologies). Relative peak height was calculated by normalization with the most abundant ion in the mass spectrum. The average resolving power was 300,000 at *m/z* 400 (calculated as $\Delta M/M$, where ΔM is the peak width at half-peak height and *M* is the *m/z* value - Marshall et al. (1998)).

2.5 FT-MS data analysis

2.5.1 Peak detection and blank correction

Peaks were considered “detected” if the peak height was greater than three times the noise level. Thresholds and noise levels were determined for each mass spectrum (Table 1). For the purposes of this study, peaks were considered “not detected” if their peak height was below the threshold. Very few peaks overlapped between the extraction blank and the DOM mass spectra. Nonetheless, all peaks found in the blank were removed from the DOM peak lists.

2.5.2 Calibration and multivariate comparisons

Daily external calibration spanned the full range of observed m/z values ($322 < m/z < 922$) and resulted in <1.0 ppm mass accuracy for all spectra collected that day. To further constrain our mass errors, spectra were internally re-calibrated with a $-CH_2$ series of m/z values present in all spectra ($393 < m/z < 519$). Elemental formulae for the calibrants were determined by best-fit with all possible elemental combinations containing ^{12}C , 1H , ^{16}O , and ^{14}N (Appendix Table 1). After internal calibration, RMS errors for the calibrants ranged from 0.05 to 0.16 ppm. Ideally, the internal calibrants would cover the full range of observed m/z values (MUDDIMAN and OBERG, 2005) but we were not able to find a series of $-CH_2$ -related m/z values that (a) would span our full range and (b) was present in all spectra. We assume that the error on mass measurement of peaks outside our calibrated range fall between the errors of the calibrated range and the 1 ppm errors set by the external calibrants. We have found that this range of mass accuracies is sufficient for assigning correct elemental formulae to the majority of peaks in DOM spectra (KUJAWINSKI and BEHN, 2006).

Differences in peak lists among the samples were assessed with cluster analysis (e.g., KOCH et al., 2005). Spectra were aligned with an in-house algorithm to generate a comprehensive list of m/z values from all spectra. For this and all subsequent analyses, we treated the resulting data matrix in two ways. In the first, all relative peak heights were transformed to presence (peak height = 1) or absence (peak height = 0) and then normalized to the total number of m/z values within each spectrum. This transformation ignores differences in relative peak height between m/z values within and among spectra. In the second transformation, the relative peak heights were retained without alteration. A distance matrix was calculated between all samples with the Bray-Curtis distance measure (code written by David Jones, University of Miami, as part of the Fathom toolbox, <http://www.rsmas.miami.edu/personal/djones/matlab/matlab.html>). Cluster analysis was performed on the presence-absence distance matrix using Ward's linkage method (Fig. 2B).

We also compared the m/z peak lists with non-metric multi-dimensional scaling (NMS - KRUSKAL, 1964; MATHER, 1976). NMS reduces the comparisons between samples from a multidimensional space to fewer dimensions, preferably two or three. These differences are then presented graphically; samples which are close together in this plot (or ordination) are more similar than samples located farther apart. We chose NMS for two reasons. First, an estimate of the ordination robustness can be calculated through comparisons between randomized datasets and the original distance matrix. Second, NMS does not assume an underlying linear relationship between variables. The statistics toolbox in Matlab was used to run the NMS analyses, and our starting configuration was the solution to classic non-dimensional scaling. Additional axes were considered if the addition of the axis resulted in a significant improvement over the randomized data (at $p \leq 0.05$) and the reduction in stress was greater than 0.05. The p-values were calculated

as the proportion of randomized runs with stress less than or equal to the observed stress calculated using Kruskal's stress formula 1; stress is a measure of goodness of fit used in NMS. In this study, we chose two dimensions based on Monte Carlo simulations that compared 20 ordinations with our distance matrix to 50 ordinations with a randomized distance matrix ($p = 0.0196$ in the presence/absence matrix). The best solution had a stress value (S_R) of 0.0947 (in the presence/absence matrix (Fig. 2A); $S_R = 100 \cdot \sqrt{S}$); where S is the scaled stress (MCCUNE and GRACE, 2002)). The proportion of variation represented by each axis was assessed with a Mantel test to calculate the coefficient of determination (r^2) between distances in the ordination space and distance in the original space.

2.5.3 *Elemental formula assignments*

The bulk of the peaks in our mass spectra correspond to singly-charged compounds because the mass difference between major peaks and their isotope peaks were always integers, rather than non-integers associated with multiply-charged compounds. We thus considered all m/z values to be equivalent to mono-isotopic masses. Elemental formulae were assigned to the aligned m/z values using a modified version of our Compound Identification Algorithm (CIA - KUJAWINSKI and BEHN, 2006). The error window on formula assignments was set at 0.5 ppm. Elemental formulae were assigned in the form, $C_cH_hO_oN_nS_s$. CIA assigns elemental formulae in a 3-step process. First, elemental formulae are determined for each m/z value below 500 Da by calculating all (mathematically) possible combinations of elements within a pre-assigned error window. Chemically unreasonable formulae are removed and a small list (usually <5) of chemically and mathematically legitimate elemental formulae remain. In previous work, this list was then sorted according to lowest deviation from the observed m/z value. We have modified

this step to sort the formulae according to the lowest number of non-oxygen heteroatoms (N+S). We validated this change with synthetic data sets (as in KUJAWINSKI and BEHN, 2006) and observed that this step significantly increases the accuracy of CIA, particularly for high-molecular weight compounds with inherently lower mass accuracies. The second step of CIA finds functional group differences between m/z values. The mass difference associated with each change in functional group is then used to assign elemental formulae for the higher m/z value as the sum of the lower formula and the appropriate functional group difference. The last step in CIA incorporates one ^{13}C atom for compounds with an isotopic isomer by finding pairs of m/z values that differ by exactly 1.0034 Da. An elemental formula containing one ^{13}C atom is then assigned to the higher m/z value. This step was validated with synthetic datasets and we concluded that this approach was superior to one in which ^{13}C was included in the brute-force assignments (Step 1) for m/z values < 500 Da.

Samples that are rich in carboxylic acid moieties are prone to esterification when extracted and stored in methanol (MCINTYRE and MCRAE, 2005). McIntyre and McRae (2005) showed that approximately 1-10% of carboxylic acids were converted to methyl esters in the presence of acid and methanol. We have not constrained the degree of methylation in our samples and we assume that up to 10% of the carboxylic acid sites may have been transformed; however this was not formally incorporated into our data analysis or interpretation. Although the effect of methylation on spectrum quality and compound diversity is minimized in negative-ion mode (ROSTAD and LEENHEER, 2004; MCINTYRE and MCRAE, 2005), methylation would explain the observation that some of the terrestrial marker m/z values have lower O:C and higher H:C ratios than unmodified lignin.

2.5.4 *Indicator species analysis*

Indicator species analysis (ISA) was adapted from Dufrene and Legendre (1997) to determine indicator m/z values in our samples. In ISA, indicator values (IVs) are calculated for all m/z values within our mass spectra from the North Atlantic Ocean samples. These indicator values are the multiplication product of the relative abundance and the relative frequency of a m/z value within a pre-defined group. In order to have a high IV, an m/z value must have high abundance *and* occur in most (if not all) samples within the group. Statistical significance of IVs is calculated by comparison with Monte-Carlo simulations of randomized data.

This analysis requires the *a priori* assignment of samples to groups. We identified the ideal number of groups in our sample set by conducting ISA with all possible group combinations of our individual samples (protocol outlined in MCCUNE and GRACE, 2002). For each case, we calculated the average p -value of all IVs and recorded the number of indicator m/z values. The (statistically) best number of groups is the case in which the average p -value is minimized and the number of indicator m/z values is maximized. For our dataset, this occurred when we had three groups of samples (Fig. 2): Group 1 = (surface / near-surface) marine samples, Group 2 = riverine / estuarine samples, and Group 3 = deep marine sample, i.e., the same groups that are visually evident within our cluster and NMS analyses. We used this group assignment to find indicator compounds in Groups 1 and 2 with both data transformations described above. In the presence/absence data transformation, all m/z values have equal weight and so the variability among the samples is driven by m/z diversity only. In contrast, the variability in the second data transformation will be determined by a combination of m/z value *and* (relative) peak height.

We culled our list of indicator compounds in both data transformations according to the following criteria:

- (1) Indicator value (IV) must be greater than or equal to 50.
- (2) The *p*-value associated with the indicator *m/z* value must be less than 0.07 when compared to randomized data from Monte Carlo simulations.
- (3) The average peak height for an indicator *m/z* value must be 1.5X greater in the assigned group than in the other group(s).
- (4) If the indicator *m/z* value is assigned an elemental formula that contains ¹³C, the *m/z* value associated with the full ¹²C-isotopomer must meet criterion 3.

3 RESULTS AND DISCUSSION

3.1 Elemental formula assignments

We assigned elemental formulae to the majority of *m/z* values in our spectra (1837 of 2201; 83%) within 0.5 ppm error. The general elemental composition was consistent with previous assessments of aquatic DOM by ESI FT-ICR MS (Fig. 3 - KOCH et al., 2005; KIM et al., 2006a; SLEIGHTER and HATCHER, 2008). For comparison with recent studies, we calculated the magnitude-averaged H:C, O:C, N:C, S:C and double-bond equivalence (DBE) values for our spectra (Table 2), using the equations in Sleighter et al. (2008). These parameters combine the relative peak height and assigned elemental formula for each *m/z* value in individual samples to derive a bulk chemistry assessment of the observed compounds in each spectrum. Thus, they represent the “chemical character” of the observed DOM. Differences in sample preparation among numerous investigators, in ionization efficiencies among various FT-MS instruments, and in ion number among disparate sample matrices preclude strict comparison of these parameters

among studies. Nonetheless, the comparison provides confirmation that DOM of similar characteristics has been observed in different investigations.

Here, values for each parameter were comparable to other studies and fit within the range of values observed for both freshwater (SLEIGHTER and HATCHER, 2008) and marine DOM (KOCH et al., 2008). In general, the freshwater samples (Group 2) appear to have slightly higher H:C values and lower O:C values than the surface marine DOM samples (Group 1), consistent with the observations of Sleighter and Hatcher (2008). Similar to previous studies, elemental formulae containing only C, H and O dominate the assigned elemental formulae (Table 3). The relative contribution of CHO formulae is higher in the riverine / estuarine samples (Group 2) than in the marine samples, although this contribution is lower when the magnitude-averaged value is considered. All marine DOM samples have higher contributions of N- and S-containing elemental formulae than the riverine / estuarine DOM samples.

We were intrigued to find a substantial contribution of sulfur relative to nitrogen in the deep marine sample. Although this observation must be confirmed with additional deep ocean samples, it is not surprising that sulfur would be preferentially observed in negative ion mode spectra. Oxygen and sulfur have acidic character and readily lose protons to form negative ions in aqueous solution. Thus, S-containing moieties such as $-\text{SO}_2$ and $-\text{SH}$ have good ionization efficiencies under negative ion mode (HUGHEY et al., 2004). In contrast, nitrogen is a basic element and prefers to gain a proton to form positive ions in aqueous solution. Therefore, N-containing moieties such as amines ($-\text{NH}_2$) have better ionization efficiencies under positive ion mode. Further, N functional groups are not extracted with high efficiency during C_{18} extraction because they are ionized ($-\text{NH}_3^+$) during acidification (SLEIGHTER and HATCHER, 2008). The relatively high contribution of S-containing formulae is an interesting observation and further

work with MS/MS fragmentation is needed to characterize their structure. Many structural isomers are likely within each peak and additional work with chromatographic pre-separation and MS/MS fragmentation is needed to define the dominant isomer(s) present.

3.2 Linkage / NMS analysis

We examined the similarity among our six samples with cluster analysis and non-metric multidimensional scaling (NMS - MCCUNE and GRACE, 2002; Fig. 2). This combination of cluster analysis and NMS has also been used successfully by Dittmar et al. (2007) for lower-resolution mass spectral data. Both cluster analysis and NMS showed similarity among the three surface or near-surface marine DOM samples (Station 2 surface, Station 5 surface and 43m – Figs. 2A and 2B) and among the estuarine / riverine DOM samples (Station 7 and Station 9). The deep marine DOM sample (Station 2 1000m) was quite different from the samples in the other groups. We are confident in this similarity assessment due to a low stress value ($S_R = 0.095$) and a high r^2 value of 0.98 between the presence / absence NMS ordination and the original Bray-Curtis distance matrix. The NMS ordinations from both data transformations are the same, suggesting that the underlying variability in the samples is due to m/z diversity within our samples, rather than to changes in relative peak height. We hypothesized that the environmental parameter underlying Dimension 1 was terrestrial input. We calculated the linear regression between salinity and Dimension 1 values, which yielded an inverse relationship with an r^2 value of 0.97. Thus Dimension 1 may represent increasing terrestrial contribution to DOM. We could not assign an environmental factor to Dimension 2 with parameters measured in this study.

3.3 Indicator species analysis

Indicator species analysis has been used in the ecological community to determine the species of organisms that can be considered characteristic or “indicators” for particular groups of samples (PETT-RIDGE and FIRESTONE, 2005). This analysis has two components – (1) an indicator value (IV) is calculated for each component within a dataset and (2) a *p*-value is calculated by comparison with randomized data. No linearity among the sample components or their underlying variables is assumed during ISA. Due to these three qualities, this analysis can be a powerful tool for resolving characteristic (or “indicator”) components within a complex dataset and can be used to focus further investigations on those components which drive sample variability. To our knowledge, this tool has not yet been applied to ultrahigh resolution mass spectral data for DOM, but it may highlight those components of DOM which are involved in key biogeochemical processes. With this approach, we must assume that reproducibility of peak presence within ESI FT-MS spectra is high. We are not aware of any studies that have systematically examined peak reproducibility within repeat sample analyses (for DOM) and this work is clearly needed to constrain our (and others’) conclusions about DOM composition. To mitigate this assumption until reproducibility studies have been completed, we developed a series of conservative criteria to cull our indicator *m/z* values and considered two data transformations.

In the first data transformation, we removed all peak heights and examine only the presence or absence of *m/z* values in different spectra. This transformation minimizes the effects of matrix ions and of relative ionization efficiencies inherent within electrospray ionization mass spectrometry. In the second transformation, we retained the relative peak height for each *m/z* value. Electrospray ionization mass spectrometry is not fully quantitative and small changes in

ion abundance in FT-ICR mass spectra cannot be used to infer similar changes in the parent solution. Our criteria for indicator m/z values help avoid concerns associated with small changes in ion abundance. For Group 1 (surface marine) samples, the indicator values (IV – Appendix Table 2) for both data transformations are identical, suggesting that the similarity between these samples is based on ion diversity, rather than on peak height. In contrast, the indicator m/z values identified in Group 2 (terrestrial) are present in most of the samples but had higher peak heights in Group 2 mass spectra than in all other mass spectra (Appendix Table 3). We conclude that the similarity in Group 2 samples is thus based on a combination of ion diversity and peak height.

Indicator species analysis yielded 32 marker m/z values for Group 1 in both data transformations. Elemental formulae were assigned to all (Appendix Table 2). Peaks associated with these elemental formulae were not detected in the riverine / estuarine DOM (Group 2) or in the deep marine DOM (Group 3). The elemental compositions of most Group 1 indicator m/z values (27/32: 84%) occur within a relatively H-rich, O-rich region that overlaps substantially with the region of proteins and carbohydrates (Figs. 3 and 4). These compound classes do not ionize as well in negative ion mode as lignin-derived products, our proposed source for the Group 2 indicator values. Thus, it is possible that Group 1 indicators were present in Group 2 spectra but were suppressed during ionization. Nonetheless, this result indicates that the concentration of lignin-derived products must be sufficiently decreased in the marine samples to allow detection of ions with lower ionization efficiencies. Significant peak suppression of Group 1 indicator values is unlikely in the deep marine DOM sample since the number of peaks in this sample was approximately the same as in the samples of Group 1 (Table 1). It seems more likely that these peaks are less abundant or not present in the deep marine sample and are produced in the surface ocean.

Twenty markers for Group 2 were identified by ISA when presence/ absence data were used and an additional 57 markers were identified when the relative peak heights were included (Appendix Table 3). Elemental formulae were assigned to all but two of the 77 m/z values. In contrast to Group 1 markers, most Group 2 markers were detected in all samples, suggesting a common origin such as terrestrial organic matter. However, a number of markers (20 / 57) were not detected in some marine samples, implying a removal process such as photo-degradation or microbial degradation. It seems unlikely that peak suppression would explain the absence of these m/z values in marine samples, due to the similarity of elemental composition between all Group 2 indicator values. The elemental compositions of Group 1 and Group 2 markers do not overlap extensively on the van Krevelen diagram (Fig. 4), suggesting that these elemental formulae represent different sources of DOM in surface marine (Group 1) and riverine / estuarine (Group 2) organic matter.

3.4 Group 2 indicator m/z values (riverine / estuarine DOM)

The riverine / estuarine samples in Group 2 were characterized by relatively low salinity and high DOC concentrations compared to the marine samples in Groups 1 and 3. Absorption (at 355 nm) by CDOM and lignin phenol concentrations decreased dramatically outside Delaware Bay (Table 1) to a low constant value in the marine samples, consistent with previous observations (VODACEK et al., 1997; DEL VECCHIO and BLOUGH, 2004b). Marker m/z values had higher peak heights in the riverine /estuarine samples (Group 2) and were either present at uniformly lower lower peak heights in the marine samples (Groups 1 and 3) or not detected, consistent with the hypothesis of a DOM source that is removed during transit from its terrestrial source to offshore waters. This removal process could be (1) a physical one, such as dilution by

mixing with the marine end-member water, (2) a chemical one, such as photochemical degradation, or (3) a biological one, such as microbial metabolism.

Group 2 marker m/z values occur within the region of the van Krevelen diagram that is representative of lignin and its methylated derivatives (region shown in Fig. 3; approximated from Figure 5 of KIM et al., 2003a). Lignin and its degradation products are common proxies for terrestrial organic matter in marine systems (OPSAHL and BENNER, 1997) and are responsible in part for the optical properties of terrestrial CDOM (HERNES and BENNER, 2003; DEL VECCHIO and BLOUGH, 2004a). Lignin-derived phenols were abundant in Group 2 samples (Table 1) and decreased in the open ocean (Groups 1 and 3) samples. The Group 2 marker m/z values have low occurrence of S or N in their elemental formulae (Appendix Table 3), as would be expected for lignin-derived materials. Further, Group 2 marker compounds are characterized by low O content and high double-bond equivalency (DBE – a proxy for aromatic character; Fig. 5). Previous work has shown that this compound class is selectively degraded by photochemistry, relative to compounds with high O content and low DBE (SCHMITT-KOPPLIN et al., 1998; KUJAWINSKI et al., 2004; GONSIOR et al., 2009). Therefore, it seems possible that Group 2 indicator m/z values are susceptible to photo-degradation and thus could represent lower-molecular-weight components of terrestrially-derived chromophoric DOM. The presence of Group 2 marker m/z values in the deep ocean suggests a terrestrial component within deep ocean DOM which is consistent with our and previous observations of lignin phenols in these environments (MEYERS-SCHULTE and HEDGES, 1986; OPSAHL and BENNER, 1997). More work is needed to ascertain whether these m/z values are indeed components of terrestrially-derived CDOM and are degraded by photochemical activity. Nonetheless, ISA has allowed us to narrow our ongoing work to 77 target m/z values for additional structural characterization and laboratory study.

3.5 Group 1 indicator m/z values (surface marine DOM)

Group 1 markers are detected only in the mass spectra of Group 1 samples, in contrast to the majority of Group 2 markers, suggesting that Group 1 markers are produced in the surface ocean and are degraded before or during export to the deep ocean. The elemental formulae of most Group 1 marker m/z values occur in a relatively H-rich and O-rich region of the van Krevelen diagram, similar to that of proteins and carbohydrates (KIM et al., 2003a; KUJAWINSKI and BEHN, 2006). Modifications of biomolecules are likely in aquatic environments that would shift the elemental formulae outside the strict protein and carbohydrate van Krevelen regions. Thus, we hypothesize that the marker m/z values for Group 1 are biologically-derived. Interestingly, Loh et al. (2004) showed that the carbohydrate-like and protein-like fractions of DOM were the youngest components of surface DOM in the North Atlantic and were recycled on 60-90 yr time-scales. Furthermore, carbohydrate and protein subunits have been shown to be good markers for DOM lability (AMON et al., 2001; CAMMACK et al., 2004). Group 1 markers, therefore, may represent recent biologically-produced DOM in the surface ocean. The absence of these m/z values in the deep ocean sample (section 3.2) is consistent with a possible biological source of Group 1 markers since enhanced DOM concentrations are repeatedly observed in the surface ocean (<200 m), most likely the labile products of intense photosynthetic and heterotrophic microbial metabolism (CARLSON, 2002).

3.6 Comparison of North Atlantic markers with DOM produced by *P. ubique*

Experimental work with pure cultures of a marine heterotrophic bacterium belonging to the SAR11 clade indicates that these abundant organisms may be sources of some Group 1

markers. The SAR11 clade is ubiquitous in gene-based surveys of seawater (GIOVANNONI et al., 1990; RUSCH et al., 2007) and is the most abundant group of bacteria measured in surface seawater by in situ hybridization (MORRIS et al., 2002; ALONSO-SAEZ et al., 2007). As a dominant heterotrophic member of bacterioplankton communities in marine environments, the SAR11 clade is likely to have a significant effect on DOM composition (GIOVANNONI et al., 2005). To our knowledge, no exudate DOM from mono-culture isolates has been examined by ultrahigh resolution mass spectrometry. However, many studies have examined small biopolymer sub-units such as amino acids and sugars that are exuded by single species and mixed assemblages, in both field and laboratory settings. Some of these studies highlighted the uptake of several monomeric compounds (e.g., glucose, amino acids and dimethyl sulfoniopropionate (DMSP)) by SAR11 (MALMSTROM et al., 2005; ALONSO and PERNTHALER, 2006; MOU et al., 2007).

We examined the DOM in *P. ubiquus* cultures grown in coastal Oregon seawater in either light or constant darkness. We used the extraordinary resolving power of FT-ICR MS to compare this DOM with DOM from the cruise samples. Most of the marker m/z values from both groups of North Atlantic Ocean samples were also detected in the *P. ubiquus* cultures (Appendix Tables 2 and 3). The peak heights of Group 2 (riverine) markers that were observed in the cultures (61 of 77: 79%) were generally unaffected by the presence of SAR11, indicating that these peaks were not degraded by light or by the growing cells. This observation is consistent with our hypothesis that the Group 2 markers represent refractory terrestrially-derived organic matter that is highly resistant to further degradation.

In contrast, the Group 1 (surface marine) markers that were detected in the *P. ubiquus* cultures (22 of 32: 69%) were always observed in one of the *P. ubiquus* cultures but not always in

the control treatment. Furthermore, the relative peak heights of 50% of the Group 1 markers were often enhanced ($\geq 2X$ increase over the control) in the presence of *P. ubiquus* under at least one of the two growth conditions. Conversely, only 10% of the Group 2 markers were enhanced during *P. ubiquus* growth. It is highly probable that markers with identical m/z values have the same elemental formulae due to the precision of m/z measurement by FT-ICR MS. However, we cannot be certain that the m/z values detected in the *P. ubiquus* cultures represent the same compounds as the corresponding m/z values detected in the North Atlantic Ocean samples. Nonetheless, the detection of Group 1 marker m/z values in the *P. ubiquus* cultures, together with the compositional similarity to labile microbial intermediates (KIM et al., 2003a; ROSSELLO-MORA et al., 2008), provides initial evidence that these peaks may be produced by biological activity. The SAR11 ecotype used in these experiments is common and highly abundant in the surface ocean, the same ocean region associated with the Group 1 markers. Our analysis does not enable us to determine whether the Group 1 markers that were produced in *P. ubiquus* cultures are unique to this organism or are general products of marine microbial activity. However, our study does not rule out the possibility that Group 1 markers represent compounds that are produced by a variety of microorganisms, including SAR11, in the surface ocean. Structural characterization of these peaks and studies with other cultured marine bacteria should further constrain their origins.

4 OVERVIEW

Marine DOM has so far eluded comprehensive chemical description, veiling the complex interactions between abiotic and biotic processes that control this vast pool of reactive carbon. Ultrahigh resolution mass spectrometry is the first technology that has the ability to detect and to

identify thousands of compounds, potentially revealing complex temporal and spatial patterns in DOM composition (HERTKORN et al., 2008). Adaptation and development of multi-variate statistics is an important step in the analysis of these vast datasets. Of critical importance will be statistical tools that help determine the critical compounds (or m/z features) within different DOM sources or biogeochemical processes.

The present project focused on the subset of DOM compounds that are extracted by C_{18} /SDB resin and are detected by negative ion mode ultrahigh resolution mass spectrometry (ESI FT-ICR MS). We resolved thousands of DOM compounds by their m/z values and used multivariate statistics to identify markers that are characteristic of terrestrial input and surface ocean sources. With statistical tools adapted from community ecology, we observed that the inferred elemental compositions for terrestrially-derived peaks are consistent with previous predictions for photochemical lability of DOM. We further showed that markers characteristic of ocean surface samples were also present in an axenic culture of a marine bacterial SAR11 isolate. Ongoing work in our laboratories is now focused on the structural characterization and dynamics of these 109 peaks to assess their geochemical significance.

Taken together, the marker compounds illustrate the complex interactions between abiotic and biotic processes that control the spatial and temporal variability of DOM. Here we have identified distinct m/z features within marine and estuarine DOM that participate in the biogeochemical processes of photochemical degradation and heterotrophic microbial metabolism. Ultrahigh resolution mass spectrometry, coupled to multivariate statistical tools, was critical to the identification of these features, highlighting the power of these techniques for elucidating the important components of biogeochemical cycles. Additional work with laboratory and field DOM is needed to confirm these results and to assess their applicability to

other marine and terrestrial regimes. Nonetheless, this approach offers not only the possibility of tracking the molecular-level distribution and dynamics of aquatic DOM, but also of obtaining detailed structural information by other advanced mass spectrometric techniques (MS/MS) on compounds linked to a specific environment or process. New markers for biogeochemical processes will be identified by this combined approach and new quantitative methods can be developed to examine the dynamics of these markers in aquatic systems, leading ultimately to novel insights into the aquatic carbon cycle.

Acknowledgements. The authors gratefully acknowledge the funding sources for this work: the National Science Foundation (OCE-0443217 (EBK, NVB, RDV), CAREER-OCE-0529101(EBK)), the Gordon and Betty Moore Foundation Marine Microbiology Initiative (SJG) and WHOI startup funds (EBK). Data were acquired with the help of Drs. A. Marshall, C. Nilsson, R. Rodgers and G. Klein at the National Ion Cyclotron Resonance Mass Spectrometry Users' Facility at the National High Magnetic Field Laboratory in Tallahassee, FL. The manuscript was improved by comments from two anonymous reviewers and Dr. D. Burdige and by discussions with Drs. C. Reddy and O. Zafiriou.

Citations

- Alonso-Saez, L., Aristegui, J., Pinhassi, J., Gomez-Consarnau, L., Gonzalez, J. M., Vaque, D., Agusti, S., and Gasol, J. M., 2007. Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean. *Aquat. Microb. Ecol.* **46**, 43-53.
- Alonso, C. and Pernthaler, J., 2006. Roseobacter and SAR11 dominate microbial glucose uptake in coastal North Sea waters. *Env. Microbiol.* **8**, 2022-2030.
- Aluwihare, L. I., Repeta, D. J., and Chen, R. F., 2002. Chemical composition and cycling of dissolved organic matter in the Mid-Atlantic Bight. *Deep Sea Res. II* **49**, 4421-4437.
- Amon, R. M. W., Fitznar, H. P., and Benner, R., 2001. Linkages among the bioreactivity, chemical composition, and diagenetic state of marine dissolved organic matter. *Limnol. Oceanogr.* **46**, 287-297.
- Azam, F. and Cho, B. C., 1987. Bacterial utilization of organic matter in the sea, *Ecology of Microbial Communities*. Cambridge University Press, Cambridge.

- Benner, R. and Biddanda, B., 1998. Photochemical transformations of surface and deep marine dissolved organic matter: Effects on bacterial growth. *Limnol. Oceanogr.* **43**, 1373-1378.
- Blough, N. V. and Del Vecchio, R., 2002. Chromophoric DOM in the Coastal Environment. In: Hansell, D. A. and Carlson, C. A. (Eds.), *Biogeochemistry of marine dissolved organic matter*. Academic Press, San Diego, CA.
- Bushaw, K. L., Zepp, R. G., Tarr, M. A., Schulz-Janders, D., Bourbonniere, R. A., Hodson, R. E., Miller, W. L., Bronk, D. A., and Moran, M. A., 1996. Photochemical release of biologically available nitrogen from aquatic dissolved organic matter. *Nature* **381**, 404-407.
- Cammack, W. K. L., Kalff, J., Prairie, Y. T., and Smith, E. M., 2004. Fluorescent dissolved organic matter in lakes: Relationships with heterotrophic metabolism. *Limnol. Oceanogr.* **49**, 2034-2045.
- Carlson, C. A., 2002. Production and removal processes. In: Hansell, D. A. and Carlson, C. A. (Eds.), *Biogeochemistry of marine dissolved organic matter*. Elsevier Science, San Diego CA.
- Connon, S. A. and Giovannoni, S. J., 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied Env. Microbiol.* **68**, 3878-3885.
- Del Vecchio, R. and Blough, N. V., 2004a. On the origin of the optical properties of humic substances. *Env. Sci. Tech.* **38**, 3885-3891.
- Del Vecchio, R. and Blough, N. V., 2004b. Spatial and seasonal distribution of chromophoric dissolved organic matter and dissolved organic carbon in the Middle Atlantic Bight. *Mar. Chem.* **89**, 169-187.
- Dittmar, T., Koch, B., Hertkorn, N., and Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography-Methods* **6**, 230-235.
- Dittmar, T. and Koch, B. P., 2006. Thermogenic organic matter dissolved in the abyssal ocean. *Mar. Chem.* **102**, 208-217.
- Dittmar, T., Whitehead, K., Minor, E. C., and Koch, B. P., 2007. Tracing terrigenous dissolved organic matter and its photochemical decay in the ocean by liquid chromatography / mass spectrometry. *Mar. Chem.* **107**, 378-387.
- Dufrene, M. and Legendre, P., 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs* **67**, 345-366.
- Giovannoni, S. J., B., B. T., Moyer, C. L., and Field, K. G., 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60-63.
- Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappé, M. S., Short, J. M., Carrington, J. C., and Mathur, E. J., 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242-1245.
- Goni, M. A. and Montgomery, S., 2000. Alkaline CuO oxidation with a microwave digestion system: Lignin analyses of geochemical samples. *Anal. Chem.* **72**, 3116-3121.
- Gonsior, M., Peake, B. M., Cooper, W. T., Podgorski, D., D'Andrilli, J., and Cooper, W. J., 2009. Photochemically induced changes in dissolved organic matter identified by ultrahigh resolution Fourier transform ion cyclotron resonance mass spectrometry. *Environmental Science & Technology* **ASAP**.

- Green, S. A. and Blough, N. V., 1994. Optical-Absorption and Fluorescence Properties of Chromophoric Dissolved Organic-Matter in Natural-Waters. *Limnol. Oceanogr.* **39**, 1903-1916.
- Hansell, D. A. and Carlson, C. A., 2002. *Biogeochemistry of Marine Dissolved Organic Matter*. Academic Press, San Diego, CA.
- Hatcher, P. G., Rowan, R., and Mattingly, M. A., 1980. 1-H and 13-C NMR of marine humic acids. *Org. Geochem.* **2**, 77-85.
- Hedges, J. I., 1990. Compositional indicators of organic acid sources and reactions in natural environments. In: Perdue, E. M. and Gjessing, E. T. (Eds.), *Organic Acids in Aquatic Ecosystems*. John Wiley & Sons, Ltd.
- Hedges, J. I., 2002. Why dissolved organics matter. In: Hansell, D. A. and Carlson, C. A. (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter*. Elsevier Science, San Diego, CA.
- Hedges, J. I. and Ertel, J. R., 1982. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. *Anal. Chem.* **54**, 174-178.
- Hernes, P. J. and Benner, R., 2003. Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. *J. Geophys. Res.* **108**, 3921.
- Hertkorn, N., Frommberger, M., Witt, M., Koch, B. P., Schmitt-Kopplin, P., and Perdue, E. M., 2008. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Anal. Chem.* **80**, 8908-8919.
- Hughey, C. A., Rodgers, R. P., Marshall, A. G., Walters, C. C., Qian, K. N., and Mankiewicz, P., 2004. Acidic and neutral polar NSO compounds in Smackover oils of different thermal maturity revealed by electrospray high field Fourier transform ion cyclotron resonance mass spectrometry. *Org. Geochem.* **35**, 863-880.
- Kieber, D. J., 2000. Photochemical production of biological substrates. In: de Mora, S. J., Demers, S., and Vernet, M. (Eds.), *The effect of UV radiation in the marine environment*. Cambridge University Press, Cambridge.
- Kieber, D. J., McDaniel, J., and Mopper, K., 1989. Photochemical source of biological substrates in sea water: implications for carbon cycling. *Nature* **341**, 637-639.
- Kim, S., Kaplan, L. A., and Hatcher, P. G., 2006a. Biodegradable dissolved organic matter in a temperate and a tropical stream determined from ultra-high resolution mass spectrometry. *Limnol. Oceanogr.* **51**, 1054-1063.
- Kim, S., Kramer, R. W., and Hatcher, P. G., 2003a. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. *Anal. Chem.* **75**, 5336-5344.
- Kim, S., Rodgers, R. P., and Marshall, A. G., 2006b. Truly "exact" mass: Elemental composition can be determined uniquely from molecular mass measurement at similar to 0.1 mDa accuracy for molecules up to similar to 500 Da. *International Journal of Mass Spectrometry* **251**, 260-265.
- Kim, S., Simpson, A. J., Kujawinski, E. B., Freitas, M. A., and Hatcher, P. G., 2003b. Non-invasive advanced spectroscopic methods (electrospray ionization mass spectrometry and 2D solution NMR) for analysis of DOM isolated by C18 solid phase disk extraction. *Org. Geochem.* **34**, 1325-1335.
- Koch, B. P., Ludwichowski, K. U., Kattner, G., Dittmar, T., and Witt, M., 2008. Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Mar. Chem.* **111**, 233-241.

- Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T., and Kattner, G., 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.
- Kruskal, J. B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1-27.
- Kujawinski, E. B. and Behn, M. D., 2006. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **78**, 4363-4373.
- Kujawinski, E. B., Del Vecchio, R., Blough, N. V., Klein, G. C., and Marshall, A. G., 2004. Probing molecular-level transformations of dissolved organic matter: Insights from electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry. *Mar. Chem.* **92**, 23-37.
- Kujawinski, E. B., Hatcher, P. G., and Freitas, M. A., 2002. High-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) of humic and fulvic acids: Improvements and comparisons. *Anal. Chem.* **74**, 413-419.
- Loh, A. N., Bauer, J. E., and Druffel, E. R. M., 2004. Variable ageing and storage of dissolved organic components in the open ocean. *Nature* **430**, 877-881.
- Louchouart, P., Opsahl, S., and Benner, R., 2000. Isolation and quantification of dissolved lignin from natural waters using solid-phase extraction and GC/MS. *Anal. Chem.* **72**, 2780-2787.
- Malmstrom, R. R., Cottrell, M. T., Elifantz, H., and Kirchman, D. L., 2005. Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Applied Env. Microbiol.* **71**, 2979-2986.
- Mannino, A. and Harvey, H. R., 1999. Lipid composition in particulate and dissolved organic matter in the Delaware Estuary: Sources and diagenetic patterns. *Geochim. Cosmochim. Acta* **63**, 2219-2235.
- Maranon, E., Cermeno, P., Fernandez, E., Rodriguez, J., and Zabala, L., 2004. Significance and mechanisms of photosynthetic production of dissolved organic carbon in a coastal eutrophic ecosystem. *Limnol. Oceanogr.* **49**, 1652-1666.
- Marshall, A. G., Hendrickson, C. L., and Jackson, G. S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* **17**, 1-35.
- Mather, P. M., 1976. *Computational methods of multivariate analysis in physical geography*. J. Wiley & Sons, London.
- McCallister, S. L., Bauer, J. E., Kelly, J., and Ducklow, H. W., 2005. Effects of sunlight on decomposition of estuarine dissolved organic C, N and P and bacterial metabolism. *Aquat. Microb. Ecol.* **40**, 25-35.
- McCune, B. and Grace, J. B., 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, Oregon.
- McIntyre, C. and McRae, C., 2005. Proposed guidelines for sample preparation and ESI-MS analysis of humic substances to avoid self-esterification. *Org. Geochem.* **36**, 543-553.
- Meyers-Schulte, K. J. and Hedges, J. I., 1986. Molecular evidence of a terrestrial component of organic matter dissolved in ocean water. *Nature* **321**, 61-63.
- Mopper, K. and Kieber, D. J., 2002. The photochemistry and cycling of carbon, sulfur, nitrogen and phosphorus. In: Hansel, D. A. and Carlson, C. A. (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter*. Academic Press, San Diego, CA.

- Mopper, K., Zhou, X., Kieber, R. J., Kieber, D. J., Sikorski, R. J., and Jones, R. D., 1991. Photochemical degradation of dissolved organic carbon and its impact on the oceanic carbon cycle. *Nature* **353**, 60-62.
- Moran, M. A., Sheldon, W. M., and Zepp, R. G., 2000. Carbon loss and optical property changes during long-term photochemical and biological degradation of estuarine dissolved organic matter. *Limnol. Oceanogr.* **45**, 1254-1264.
- Moran, M. A. and Zepp, R. G., 1997. Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Limnol. Oceanogr.* **42**, 1307-1316.
- Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., and Giovannoni, S. J., 2002. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806-810.
- Mou, X. Z., Hodson, R. E., and Moran, M. A., 2007. Bacterioplankton assemblages transforming dissolved organic compounds in coastal seawater. *Env. Microbiol.* **9**, 2025-2037.
- Muddiman, D. C. and Oberg, A. L., 2005. Statistical evaluation of internal and external mass calibration laws utilized in Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **77**, 2406-2414.
- Nagata, T., 2000. Production mechanisms of dissolved organic matter. In: Kirchman, D. L. (Ed.), *Microbial ecology of the oceans*. Wiley-Liss, Inc., New York NY.
- Nelson, N. B. and Siegel, D. A., 2002. Chromophoric DOM in the open ocean. In: Hansell, D. A. and Carlson, C. A. (Eds.), *Biogeochemistry of marine dissolved organic matter*. Academic Press, San Diego, CA.
- Obernosterer, I. and Benner, R., 2004. Competition between biological and photochemical processes in the mineralization of dissolved organic carbon. *Limnol. Oceanogr.* **49**, 117-124.
- Opsahl, S. and Benner, R., 1997. Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature* **386**, 480-482.
- Pett-Ridge, J. and Firestone, M. K., 2005. Redox fluctuation structures microbial communities in a wet tropical soil. *Applied Env. Microbiol.* **71**, 6998-7007.
- Rappé, M. S., Connon, S. A., Vergin, K. L., and Giovannoni, S. J., 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630-633.
- Rossello-Mora, R., Lucio, M., Pena, A., Brito-Echeverria, J., Lopez-Lopez, A., Valens-Vadell, M., Frommberger, M., Anton, J., and Schmitt-Kopplin, P., 2008. Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *Isme Journal* **2**, 242-253.
- Rostad, C. E. and Leenheer, J. A., 2004. Factors that affect molecular weight distribution of Suwannee river fulvic acid as determined by electrospray ionization mass spectrometry. *Analytica Chimica Acta* **523**, 269-278.
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D. Y., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y. H., Falcon, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Neilson, K., Friedman, R., Frazier, M., and Venter, J. C., 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* **5**, 398-431.

- Schmitt-Kopplin, P., Hertkorn, N., Schulten, H.-R., and Kettrup, A., 1998. Structural changes in a dissolved soil humic acid during photochemical degradation processes under O₂ and N₂ atmosphere. *Env. Sci. Tech.* **32**, 2531-2541.
- Seitzinger, S. P., Hartnett, H., Lauck, R., Mazurek, M. A., Minegishi, T., Spyres, G., and Styles, R. M., 2005. Molecular-level chemical characterization and bioavailability of dissolved organic matter in stream water using electrospray ionization mass spectrometry. *Limnol. Oceanogr.* **50**, 1-12.
- Senko, M. W., Canterbury, J. D., Guan, S., and Marshall, A. G., 1996a. A high-performance modular data system for FT-ICR mass spectrometry. *Rapid Comm. Mass Spectrom.* **10**, 1839-1844.
- Senko, M. W., Hendrickson, C. L., Pasa-Tolic, L., Marto, J. A., White, F. M., Guan, S., and Marshall, A. G., 1996b. Electrospray ionization Fourier transform ion cyclotron resonance at 9.4 T. *Rapid Comm. Mass Spectrom.* **10**, 1824-1828.
- Sleighter, R. L. and Hatcher, P. G., 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.
- Sleighter, R. L., McKee, G. A., and Hatcher, P. G., 2008. Direct Fourier transform mass spectral analysis of natural waters with low dissolved organic matter. *Org. Geochem.*
- Stenson, A. C., Marshall, A. G., and Cooper, W. T., 2003. Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **75**, 1275-1284.
- Tranvik, L. and Kokalj, S., 1998. Decreased biodegradability of algal DOC due to interactive effects of UV radiation and humic matter. *Aquat. Microb. Ecol.* **14**, 301-307.
- Tremblay, L. B., Dittmar, T., Marshall, A. G., Cooper, W. J., and Cooper, W. T., 2007. Molecular characterization of dissolved organic matter in a North Brazilian mangrove porewater and mangrove-fringed estuaries by ultrahigh resolution Fourier Transform-Ion Cyclotron Resonance mass spectrometry and excitation/emission spectroscopy. *Mar. Chem.* **105**, 15-29.
- Vaillancourt, R. D., Marra, J., Prieto, L., Houghton, R. W., Hales, B., and Hebert, D., 2005. Light absorption and scattering by particles and CDOM at the New England shelfbreak front. *Geochem. Geophys. Geosystems* **6**.
- Vodacek, A., Blough, N. V., DeGrandpre, M. D., Peltzer, E. T., and Nelson, R. K., 1997. Seasonal variation of CDOM and DOC in the Middle Atlantic Bight: Terrestrial inputs and photooxidation. *Limnol. Oceanogr.* **42**, 674-686.

TABLES

Table 1. Data for samples used for ESI FT-ICR MS analysis in this study. Lignin phenol concentrations are the average of replicate analyses with one standard deviation. The number of unique m/z values in each spectrum indicates the number of m/z values with a signal-to-noise ratio of 3 or greater. Threshold values for peak heights are defined as three times the noise level and were individually calculated for each spectrum. Assignment of samples to group numbers (column 11) is described in the text.

Sample set	Station	Location	Depth (m)	Salinity	Temp (°C)	[DOC] ($\mu\text{mol L}^{-1}$)	a(355) (m^{-1})	[Lignin phenol] ($\mu\text{g L}^{-1}$)	# m/z values	Threshold peak height	Group
Cruise, September 2005	2 Sargasso Sea	36.74°N	0	36.1	26.9	112	0.059	0.36 \pm 0.04	1392	4.5	1
		71.02°W	--	1000	35.0	4.9	92	0.079	0.83 \pm 0.10	1264	6.0
	5 shelf-break	38.18°N	0	32.5	24.7	124	0.235	0.70 \pm 0.16	1190	6.0	1
		74.25°W	--	43	32.5	9.8	129	0.402	0.82 \pm 0.03	1115	6.0
	7 mouth of Delaware River	39.31°N	0	14.4	26.1	275	3.39	7.11 \pm 0.91	923	9.0	2
		75.38°W									
	9 head of Delaware River	40.18°N	0	0.12	27.1	184	1.99	3.17 \pm 1.30	785	9.0	2
		74.75°W									
<i>P. ubique</i> data set	Non-inoculated control	--	--	--	--	106	--	--	1425	0.45	n/a
	<i>P. ubique</i> (dark)	--	--	--	--	109	--	--	2097	0.75	n/a
	<i>P. ubique</i> (light)	--	--	--	--	99	--	--	1918	0.6	n/a

Table 2. Elemental ratios for bulk samples, calculated as magnitude-averaged values (SLEIGHTER et al., 2008) for m/z values with assigned elemental formulae. All samples from this study and from Sleighter and Hatcher (2008) are bulk C18 extracts, analyzed by direct infusion negative ion mode ESI FT-MS. Samples from Koch et al. (2008) are HPLC-separated fractions 1-4 that were subsequently analyzed by negative ion mode ESI FT-MS. “n.r.” = not reported.

Group / Ref	Sample	H:C_w	O:C_w	N:C_w	S:C_w	DBE_w
1: Surface ocean	Station 2 0m	1.30	0.36	0.004	0.003	9.1
	Station 5 0m	1.32	0.34	0.003	0.003	9.1
	Station 5 43m	1.30	0.35	0.003	0.003	9.4
Koch et al. (2008)	Antarctic Surface Sea Water	1.31	0.43			8.2
		1.31	0.43			8.3
		1.37	0.37	n.r.	n.r.	8.1
		1.39	0.33			7.7
2: Fresh water	Station 7 0m	1.34	0.33	0.002	0.001	9.2
	Station 9 0m	1.31	0.30	0.002	0.002	8.6
Sleighter & Hatcher (2008)	Dismal Swamp	1.25	0.39	n.r.	n.r.	9.6
	Great Bridge	1.29	0.35	n.r.	n.r.	8.5
	Town Point	1.37	0.35	n.r.	n.r.	7.6
	Chesapeake Bay Bridge	1.40	0.35	n.r.	n.r.	7.4
3: Deep ocean	Station 2 1000m	1.12	0.33	0.007	0.013	8.9
Koch et al. (2008)	Weddell Sea Deep Water	1.19	0.55			8.7
		1.29	0.45			8.5
		1.41	0.37	n.r.	n.r.	7.7
		1.57	0.32			5.6

Table 3. Formula classes for each mass spectrum and comparison with previous studies. All samples represent C₁₈-extracts of DOM analyzed by direct infusion negative ion mode ESI FT-MS. Number- and magnitude-averaged values determined by equations from Sleighter et al. (2008). (*) Values from Sleighter & Hatcher (2008). Values from Great Bridge, Town Point, and Chesapeake Bay Bridge were not reported as individual formula classes and are shown here as the sum of non-CHO formulae (as in reference). (#) Values from Sleighter et al. (2008).

Group / type	Sample	# <i>m/z</i> values	# Formulae	CHO	CHON	CHOS	CHONS
				<i>Number-averaged</i>			
1: Surface ocean	Station 2 0m	1392	1330 (96%)	81.1	6.1	10.8	1.7
	Station 5 0m	1190	1154 (97%)	86.2	2.0	10.1	1.5
	Station 5 43m	1115	1084 (97%)	84.5	3.6	10.6	1.1
2: Rivers	Station 7 0m	923	905 (98%)	93.7	0.4	4.8	0.8
	Station 9 0m	785	741 (94%)	93.5	0.4	5.1	0.7
3: Deep ocean	Station 2 1000m	1264	1052 (83%)	72.2	2.4	19.1	5.4
				<i>Magnitude-averaged</i>			
1: Surface ocean	Station 2 0m			89.5	2.3	5.1	0.6
	Station 5 0m			91.0	0.8	5.5	0.6
	Station 5 43m			89.6	1.4	6.4	0.5
2: Rivers	Station 7 0m			95.2	0.2	2.2	0.4
	Station 9 0m			91.0	0.2	3.1	0.7
	Dismal Swamp*			97.4	0.3	0.6	0.5
	Great Bridge*			95.7	Sum: 4.3%		
	Town Point*			91.9	Sum: 8.1%		
	Chesapeake Bay Bridge*			90.5	Sum: 9.5%		
	Pamunkey River [#]			86.4	1.1	7.3	2.4
	Dothan Run [#]			89.9	3.5	4.2	1.0
Conodoguinet Creek [#]			85.3	2.4	4.4	4.3	
3: Deep ocean	Station 2 1000m			74.3	0.9	10.7	2.8

Figure Captions

Figure 1. Representative mass spectra from the three groups identified in NMS and cluster analysis: Group 1: surface ocean (Station 2, 0m – A); Group 2: riverine / estuarine (Station 7, 0m – B); and Group 3: deep ocean (Station 2, 1000m – C). (*) = contaminants observed in all spectra and removed from all peak lists. For each spectrum, the inset shows the region $499.0 \leq m/z \leq 499.4$ and the indicator m/z values for each group. Group 1 markers are shown with solid triangles and Group 2 markers are shown with open circles. The peak detection threshold for each spectrum is shown with a dotted line in the inset. Peak heights below this threshold are considered “not detected”.

Figure 2. (A) Ordination plot for non-metric multi-dimensional scaling of 6 samples in this study. The ordination was calculated with the presence/absence data transformation. Samples that are close together are more similar than those which occur farther apart. (B) Linkage diagram of 6 samples in this study calculated from original Bray-Curtis distance matrix and Ward’s method.

Figure 3. Van Krevelen diagrams of all formulae assigned to peaks within all spectra in this study. The dots represent 1837 elemental formula assignments out of 2201 total peaks. Elemental formula assignments were constrained to ^{12}C , ^{13}C , ^1H , ^{16}O , ^{14}N and ^{32}S . Compound class regions are provided, as approximated from Kim et al. (2003a) and Hedges (1990).

Figure 4. Van Krevelen diagrams with indicator peaks determined by Indicator Species

Analysis. Top: Indicator peaks (as defined and described in text) were identified based on mass spectral data after removal of relative abundance (data transformation #1: presence/ absence).

Bottom: Indicator peaks were identified by inclusion of relative abundance in each mass spectrum (data transformation #2). In both diagrams, Group 1 represents the surface ocean samples in solid triangles and Group 2 represents the riverine samples in open circles.

Figure 5. Oxygen number vs. Double bond equivalence (DBE) for all formulae assigned in this study. Marker compounds for each group are shown as either solid triangles (Group 1 – A) or open circles (Group 2 – B). Ellipses indicate the region that is characteristic of the two groups of marker compounds (Group 1 – black; Group 2 – grey).

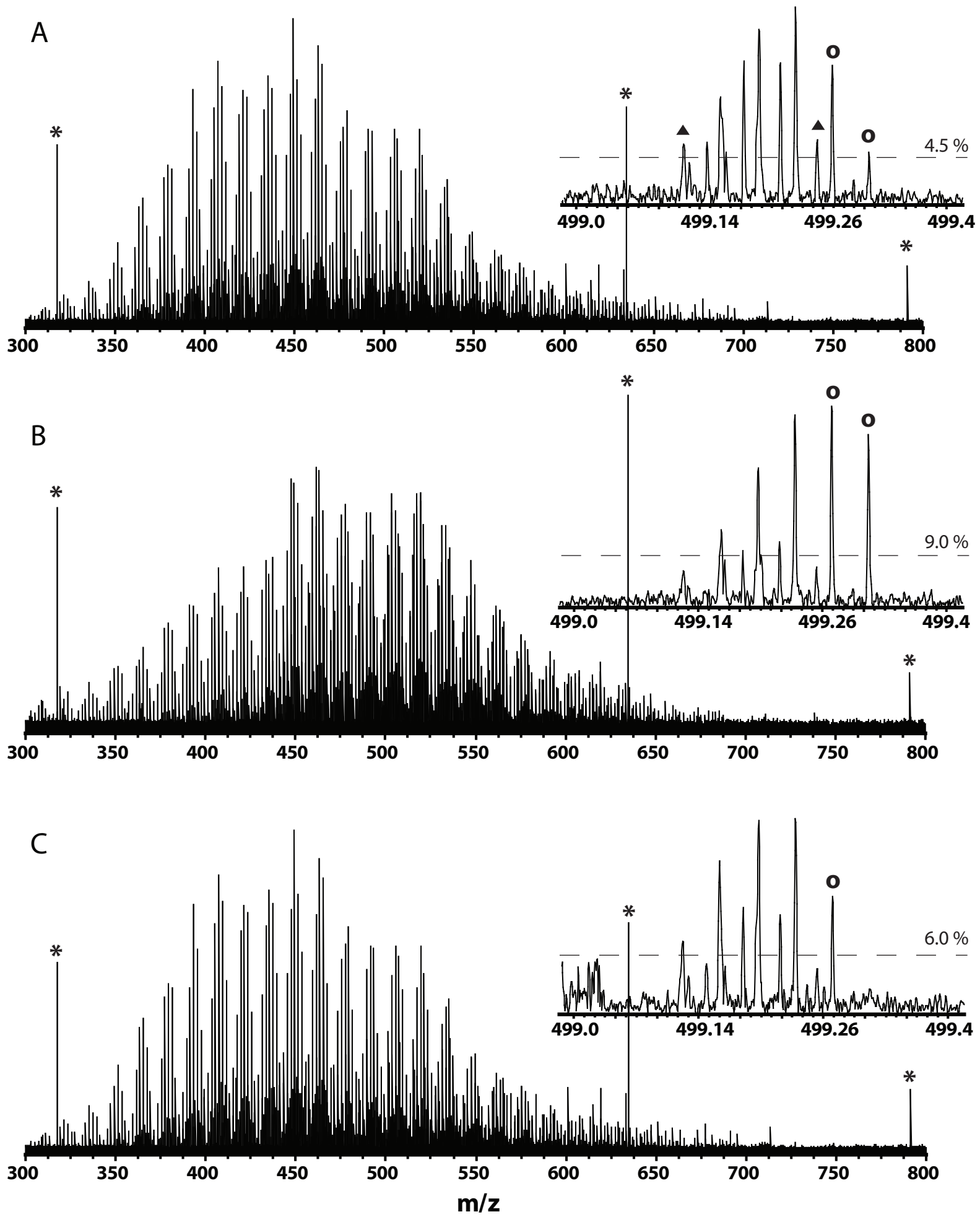


Figure 1.

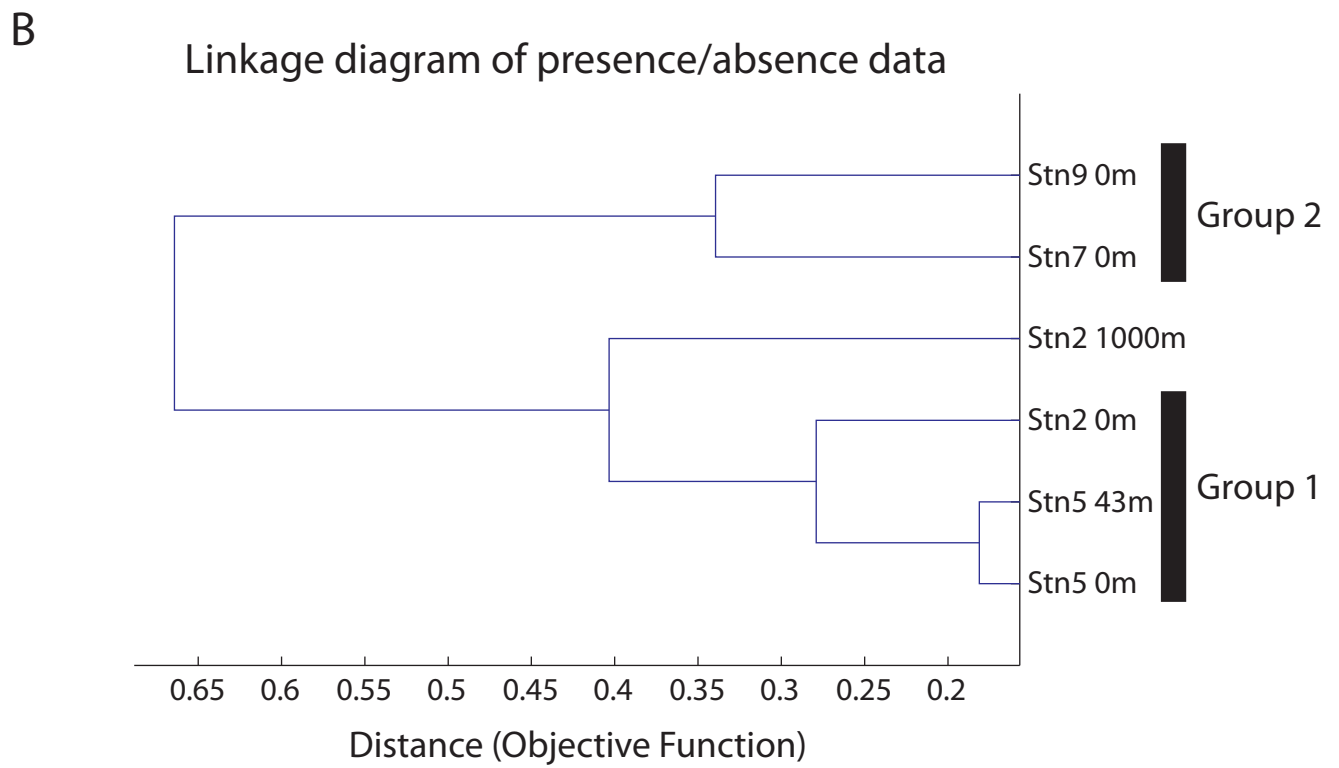
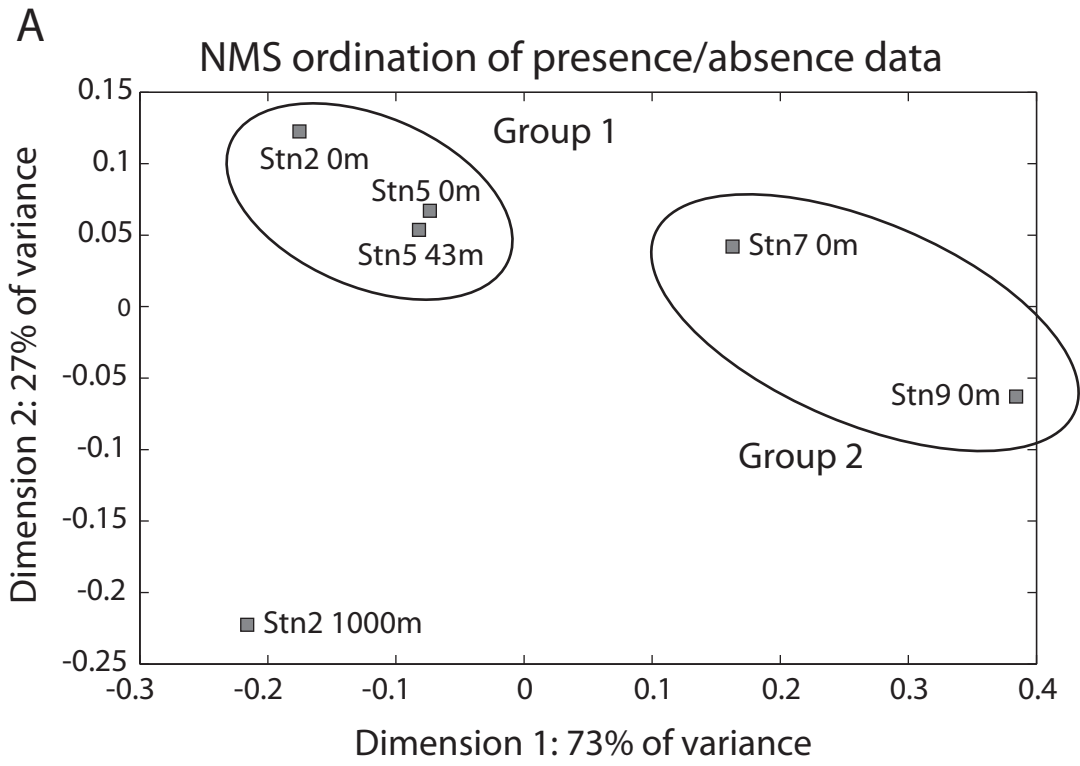


Figure 2.

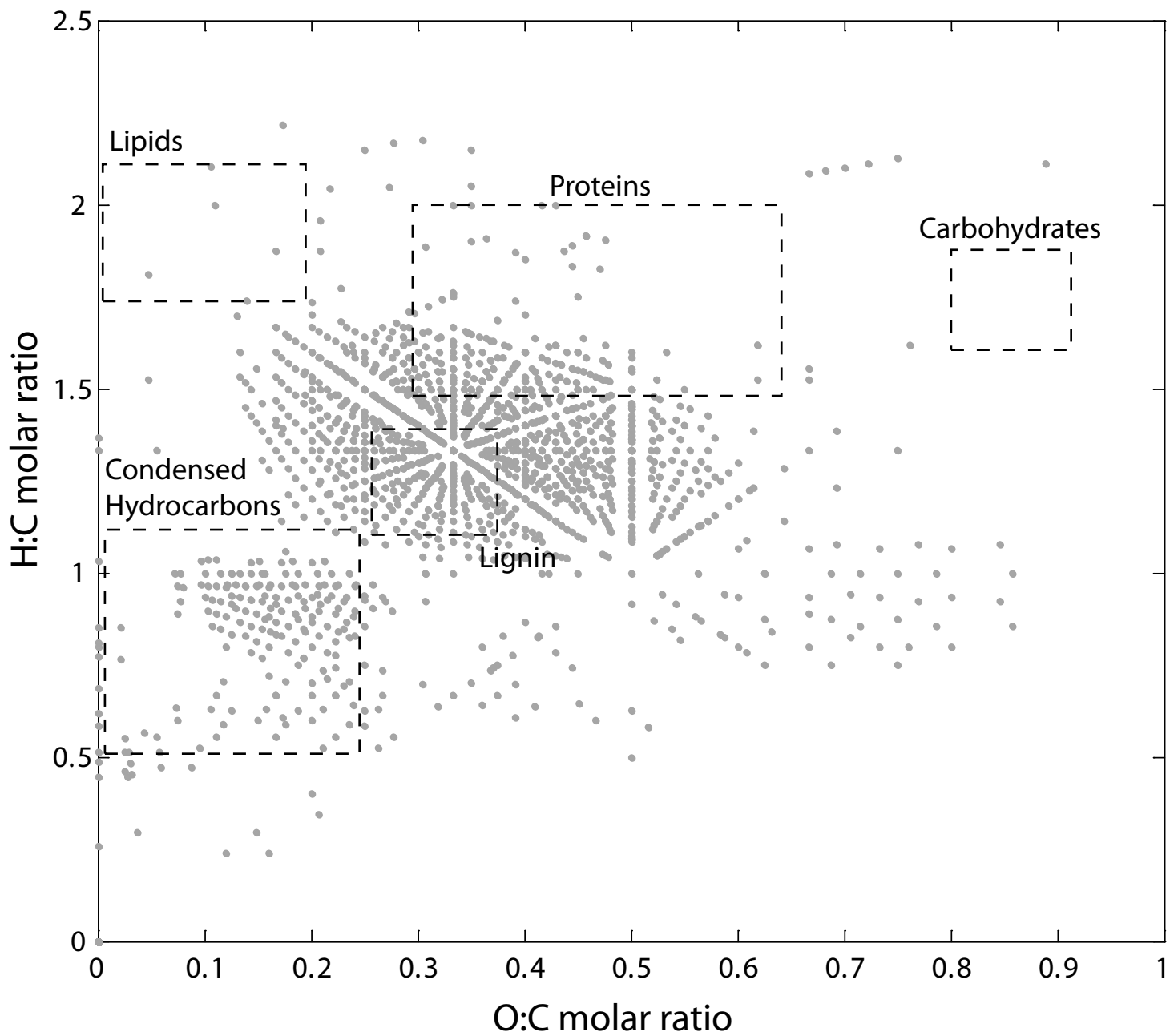


Figure 3.

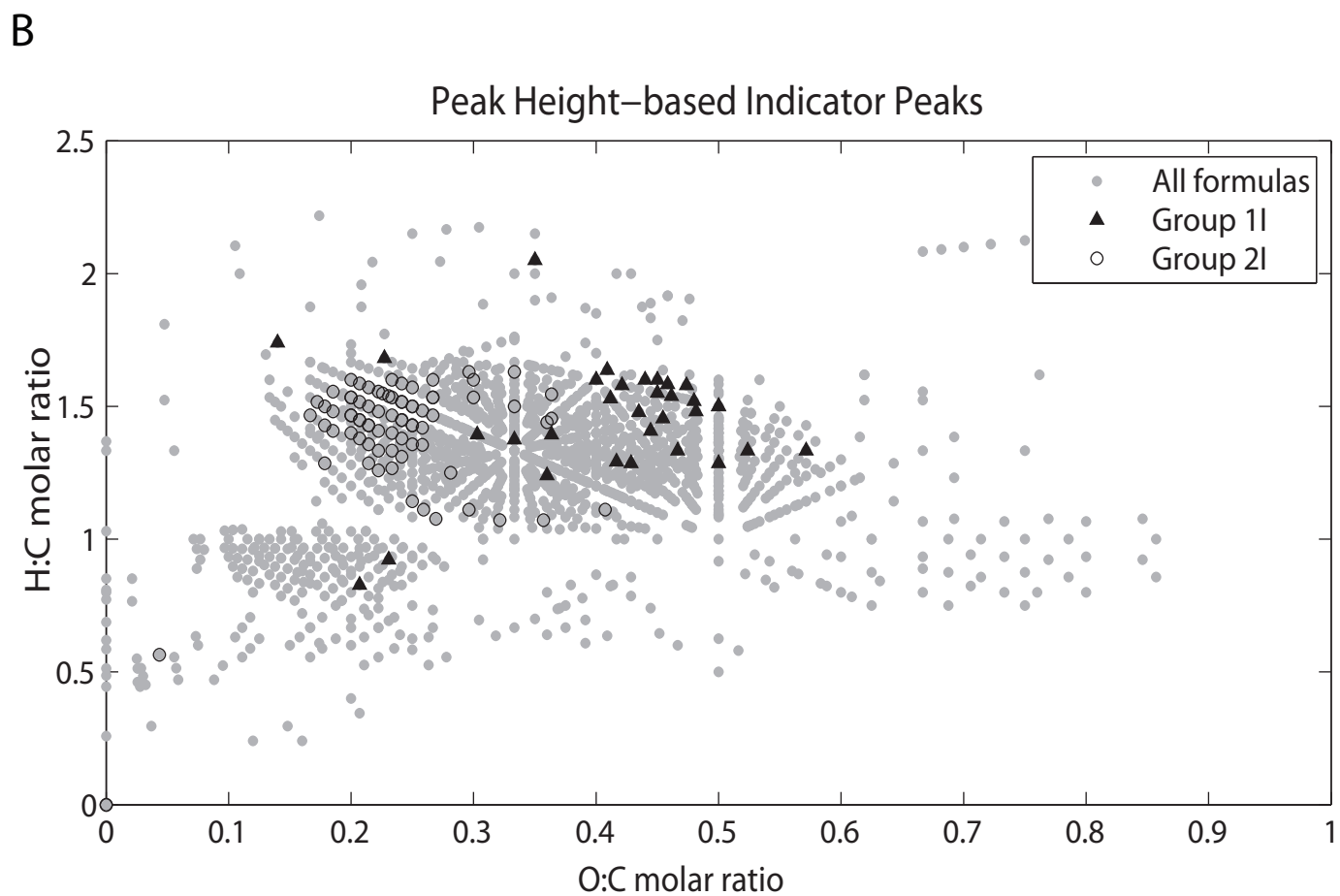
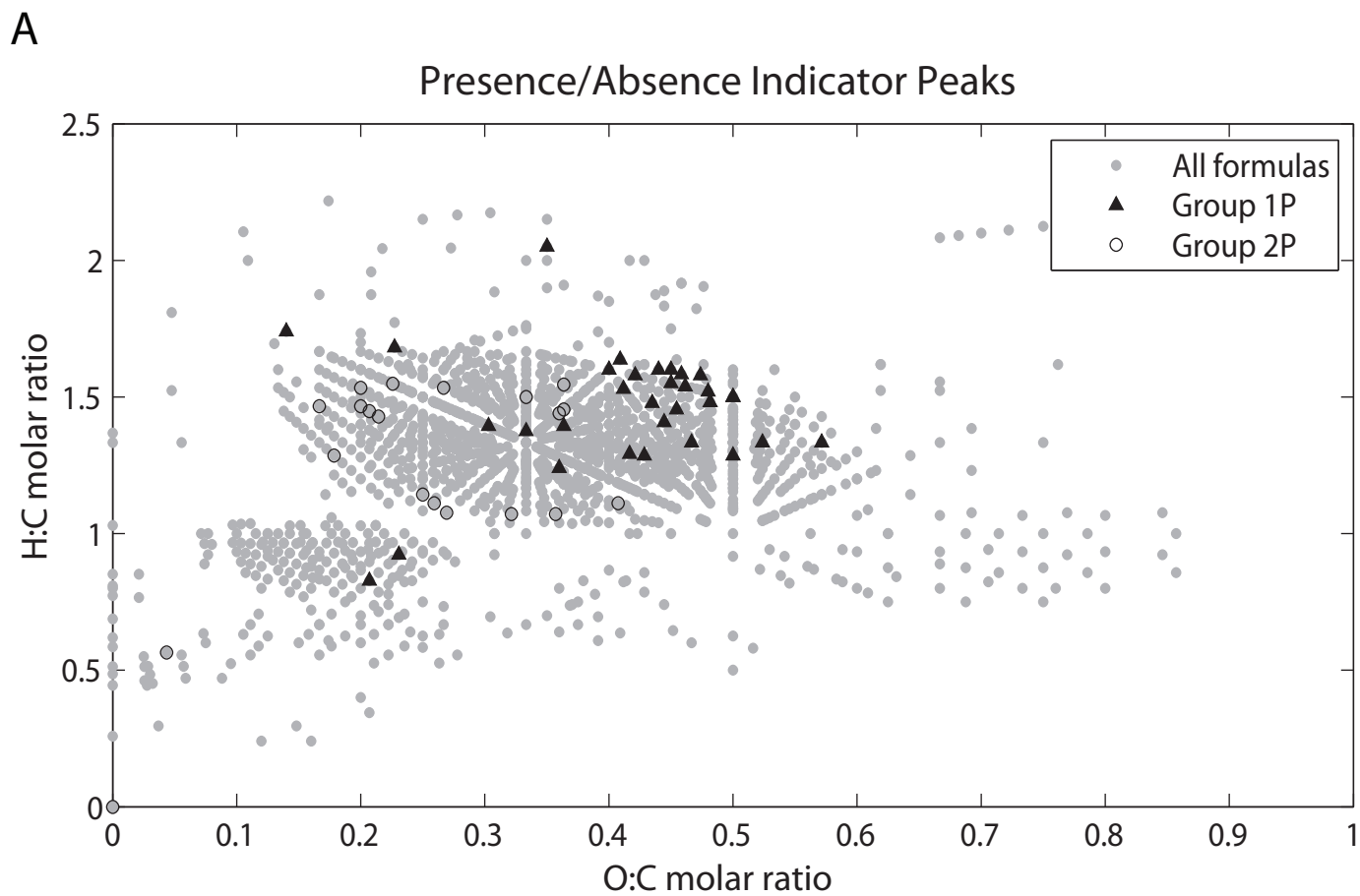


Figure 4.

Peak Height-based Indicator Peaks

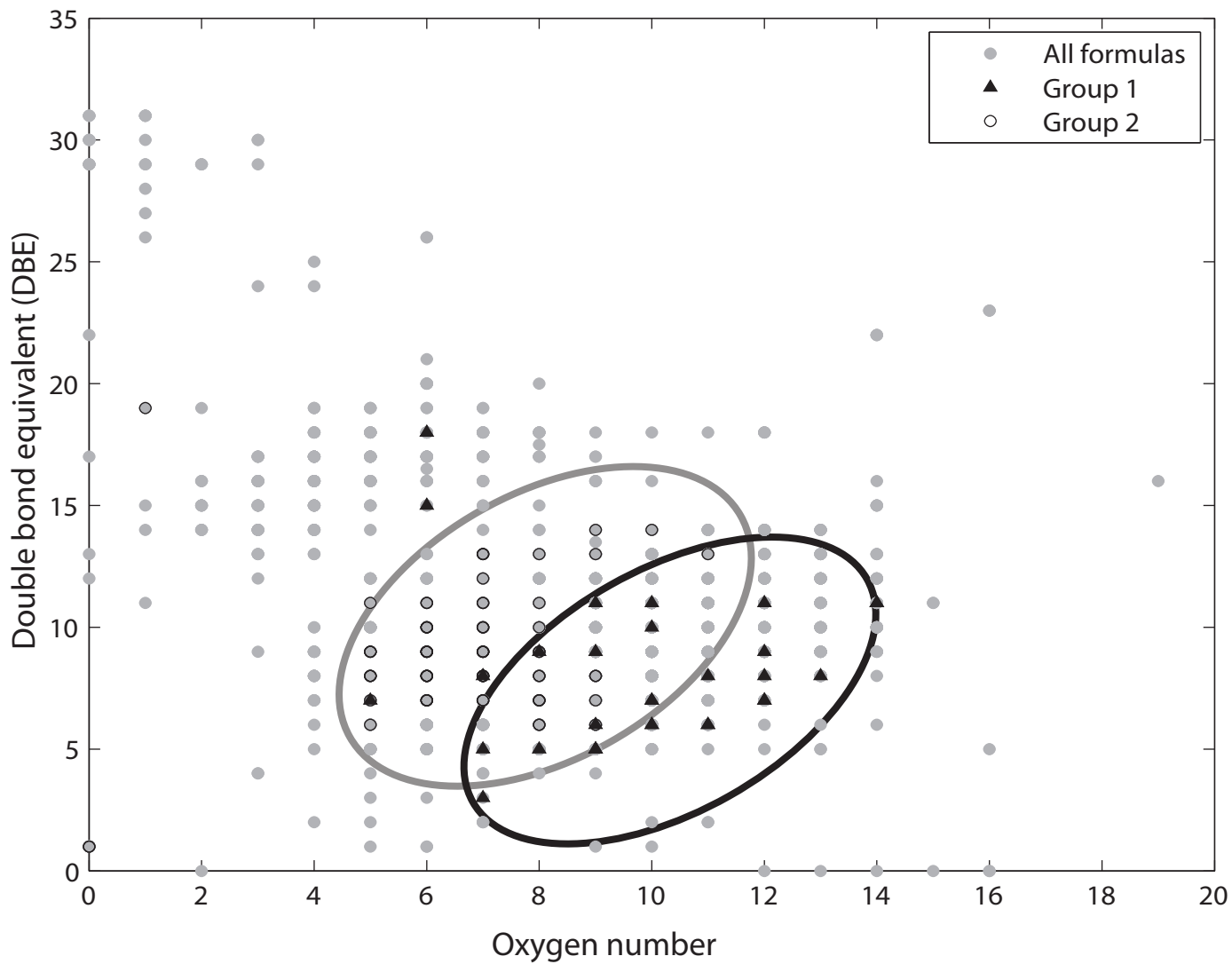


Figure 5.