Using network analysis to discern compositional patterns in ultrahigh resolution mass spectrometry data of dissolved organic matter

Authors: Krista Longnecker* and Elizabeth B. Kujawinski†

Woods Hole Oceanographic Institution, Marine Chemistry and Geochemistry, Woods Hole, MA 02543, U.S.A. *klongnecker@whoi.edu ; †ekujawinski@whoi.edu

Short title: Network analysis and mass spectrometry

Accepted: *Rapid Communications in Mass Spectrometry*

*Corresponding author. Mailing address: WHOI MS#4, Woods Hole, MA 02543, USA. Phone: (508) 289-2824. Fax: (508) 457-2164. E-mail: klongnecker@whoi.edu

Temporary citation:

Longnecker, K. and E. B. Kujawinski (accepted). Using network analysis to discern compositional patterns in ultrahigh resolution mass spectrometry data of dissolved organic matter. *Rapid Communications in Mass Spectrometry*. http://dx.doi.org/10.1002/rcm.7719.

ABSTRACT

RATIONALE

Marine dissolved organic matter (DOM) has long been recognized as a large and dynamic component of the global carbon cycle. Yet, DOM is chemical varied and complex and these attributes present challenges to the researchers interested in addressing questions about the role of DOM in global biogeochemical cycles.

METHODS

This project analyzed organic matter extracts from seawater with direct infusion with electrospray ionization into a Fourier transform ion cyclotron resonance mass spectrometer (ESI FT-ICR-MS). We used network analysis to quantify the number of chemical transformations between mass-to-charge values in each sample. The network of chemical transformations was calculated using the MetaNetter plug-in within Cytoscape. The chemical transformations serve as markers for the shared structural characteristics of compounds within complex dissolved organic matter.

RESULTS

Network analysis revealed that transformations involving selected sulfur-containing moieties and isomers of amino acids were more prevalent in the deep sea than in the surface ocean. Common chemical transformations were not significantly different between the deep sea and surface ocean. Network analysis complements existing computational tools used to analyze ultrahigh resolution mass spectrometry data.

CONCLUSIONS

This combination of ultrahigh resolution mass spectrometry with novel computational tools has identified new potential building blocks of organic compounds in the deep sea, including the unexpected importance of dissolved organic sulfur components. The method described here can be readily applied by researchers to analyze heterogeneous and complex dissolved organic matter.

INTRODUCTION

Dissolved organic matter (DOM) is a complex and heterogeneous mixture of compounds. This complexity presents analytical and computational challenges that require the continuous development of new techniques for the compositional analysis of DOM. An array of analytical platforms has been used to characterize DOM ranging from nuclear magnetic resonance instruments to mass spectrometers coupled to gas chromatography- or liquid chromatography-based pre-separation. Each of these analytical systems generates information about different components of DOM. The resulting data can include quantitative details for individual molecules which may be representative of larger processes [1, 2] or may include information on the diversity of the thousands of molecules that comprise DOM [3, 4]. The breadth of described DOM extraction methods [5, 6], further emphasizes the complexity of DOM. Thus, there is no one extraction protocol, instrumentation setup, or computational analysis tool that is able to fully resolve, identify, and quantify the compounds that comprise DOM.

Here we choose ultrahigh resolution mass spectrometry to assess the molecular level composition of DOM. This project relied on direct infusion with electrospray ionization (ESI) into a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS). The resulting

data provided thousands of mass-to-charge ratios. However, even with a hypothetical mass spectrometer with no measurement error, there are many possibilities for chemical structures and this problem is exacerbated as the mass-to-charge values increase [7]. A variety of tools exist to analyze ultrahigh resolution mass spectrometry data. The computational tools allow the calculation of elemental formulas [8] and structural isomers [9], provide models of the stability of DOM [10], and include visualization tools such as van Krevelen diagrams [11], carbon vs. mass plots [12], and two-dimensional correlation analysis [13]. Alternatively, statistical tests can be used to resolve individual, or groups of, molecules that serve as characteristic markers of environmental conditions or processes [e.g., 14, 15]. Despite these advances in the analysis of ultrahigh resolution mass spectrometry data, there is still a need for the development of novel tools which can increase our understanding of DOM structure and composition in aquatic environments.

Network analysis tools have proven valuable in characterizing links between genes and proteins [16], ecological connections within bacterial communities [17], and visualization of ultrahigh resolution mass spectrometry data [18, 19]. The use of network analysis for ultrahigh resolution mass spectrometry data builds upon earlier research in which differences between mass-to-charge values in a sample were calculated and used to determine possible elemental formulas [20, 21]. In the present study, we extend network analysis of ultrahigh resolution mass spectrometry data by quantifying the number of chemical transformations observed within water samples from surface and deep ocean water masses.

EXPERIMENTAL

Sample collection

Seawater samples were collected in July 2010 from two stations off the northeastern coast of South America. Samples were divided into surface water samples and water samples from the deeper, North Atlantic Deep Water (NADW, Table 1). Seawater was filtered with 0.2 μm Omnipore Filters (Millipore, Massachusetts, USA) mounted in Teflon filter holders, acidified to pH 3, and the dissolved organic matter (DOM) was extracted from the acidified seawater with 1g / 6 ml Bond Elut PPL cartridges (Varian, California, USA) [6] as previously described [22].

Ultrahigh resolution mass spectrometry data collection

All samples were analyzed on a 7T FT-ICR mass spectrometer (FT-ICR-MS, Thermo Fisher Scientific, Waltham MA, USA). For positive ion mode analyses, sample aliquots were reconstituted in 50% methanol/water with 0.1% formic acid. For negative ion mode analyses, sample aliquots were reconstituted in 50% methanol/water. MilliQ water, processed and analyzed in the same manner as the samples, indicated low overlap between the MilliQ water and the samples. For both positive and negative ion modes, samples were infused into the ESI interface at 4 μL min^{-1}, and instrument and spray parameters were optimized for each sample. The capillary temperature was set at 250°C, and the spray voltage was between 3.7 and 4 kV. At least 200 scans were collected for each sample which is a sufficient number of scans for good peak reproducibility [23]. The mass ranges for the full-scan collection was $150 < m/z < 1000$ in both positive and negative ion modes. Weekly mass calibrations were performed with an external standard (Thermo Calibration Mix) which results in mass accuracy errors < 1.5 ppm. The processed spectra are internally calibrated following the guidelines described in Bhatia et al. [24] which resulted in a mass accuracy < 1 ppm. The target average resolving power was 400,000 at

m/z 400 (where resolving power is defined as $m/\Delta m_{50\%}$ where $\Delta m_{50\%}$ is the width at half-height of peak *m*).

Peak Detection

We collected individual transients as well as a combined raw file using xCalibur 2.0 (Thermo Fisher Scientific, Waltham MA, USA). Transients were co-added and processed with custom-written MATLAB code [25]. Within each sample, only those transients whose total ion current (TIC) was greater than 20% of the maximal TIC were co-added, processed with Hanning apodisation, and zero-filled once prior to fast Fourier transformation. We retained all mass-to-charge (*m/z*) values with a signal-to-noise ratio above 5. Spectra were internally re-calibrated using a short list of *m/z* values present in a majority of the samples. The individual sample peak lists were then aligned in MATLAB [26]. Positive and negative ion mode data were aligned separately in MATLAB with an error tolerance of 1 ppm. The data are publically-available at MetaboLights (MTBLS366).

Network analysis

We used network analysis to examine differences between pairs of *m/z* values within a sample that could be ascribed to specific chemical transformations. This analysis is more robust if all the samples have the same dynamic range in peak heights in order to minimize differences in ionization efficiencies of features across a sample set. Thus, we conducted the network analysis using peak heights corrected to consider the dynamic range in each sample. The dynamic range of each sample was calculated as the ratio of the highest and lowest peak height within each sample. The detection limit for each sample was then calculated as the maximum peak height divided by the lowest dynamic range in the sample set. Any peaks with peak heights

below this detection limit were discarded. This calculation has the effect of lowering the sensitivity of a set of samples to the sample with the smallest dynamic range.

We used Cytoscape [27] with the MetaNetter plug-in [28] to conduct the network analysis. The list of *m/z* values from each sample was imported into Cytoscape where each *m/z* value was defined as a node. In networks, nodes are connected to each other by edges. We defined edges as the mass difference between two *m/z* values resulting from a chemical transformation. For example, the gain or loss of $C_6H_{12}N_4O$ ($\Delta m = 156.101111$) between two compounds would be represented as two 'nodes', or the corresponding smaller and larger *m/z* values, connected by an 'edge' named '$C_6H_{12}N_4O$'. Using these lists, MetaNetter calculated the edges for each sample within a 1 ppm error window. Figure 1 shows an example of a network with nine nodes and 15 edges. When the samples from the present project were analyzed with network analysis, the graphical presentation of the results is more complex (e.g., Figure 2). For each chemical transformation (e.g. gain or loss of a CH_2 group), there may be a series of connected *m/z* values. In addition, one *m/z* value may be connected to other *m/z* values by more than one transformation. As an example, in Figure 1 *m*/*z* 565.35567 is connected to *m*/*z* 581.35086 by an oxygen atom and to *m*/*z* 563.3766 by the substitution of an oxygen atom for a CH_2 group.

The network analysis requires pre-defined lists of *m/z* values for use as edges in MetaNetter. The first list of chemical transformations analyzed using network analysis was the most common chemical transformations within the dataset. To identify these transformations, we calculated all possible mass differences between *m*/*z* values in each sample. We then used the algorithm described by Kunenkov et al. [21], which accounts for the error in the mass spectrometer, to produce a discrete set of mass differences within each sample. From this smaller

list, we determined the most frequent chemical transformations in both positive and negative ion modes and assigned elemental compositions to each (Table 2).

Our second list of possible chemical transformations included chemical building blocks, amino acids, and an organo-sulfur functional group. Amino acids are a large fraction of identifiable organic molecules in marine systems [1] and they play a central role in biological processes as the basic structural units of proteins. Thus, our second list of transformations includes *m/z* values that could be isomers of the twenty essential amino acids (Table 3). We do not intend to convey that we have identified these compounds solely based on their *m/z* values. Furthermore, we are not searching for the exact mass listed in Table 3, rather we are looking for the mass difference between two measured *m/z* values that corresponds to the mass listed in the table. We also have an interest in sulfur-containing organic molecules in the deep sea, and Table 4 lists six additional elemental formulas, their exact mass, and possible isomers for each chemical transformation.

The final set of chemical transformations we assessed using MetaNetter were randomly-generated transformations. The random chemical transformations were generated in MATLAB. Each random chemical transformation was required to contain at least one carbon, hydrogen, and oxygen; however, they could also contain zero or more nitrogen, sulfur, and/or phosphorus. We used the 'rand' random number generator with MATLAB to randomly assign the number of elements within each chemical transformation. In order to prevent assessment of chemically unlikely elemental formulas, we checked the elemental formula using published bounds on elemental formulas from Kind et al. [29]. Once the elemental formulas passed these criteria, we conducted the network analysis using MetaNetter as already described.

For the network analysis, a cluster is a group of *m/z* values within one spectrum that are linked by one type of chemical transformation. The size of each cluster is equivalent to the number of *m/z* values that are linked together and thus the minimum cluster size is two. With the network analysis, we observed clusters up to a size of 21 which corresponds to 21 *m/z* values linked by one of the chemical transformations. With potentially hundreds of clusters for each chemical transformation, manually counting the number of clusters is not tractable. We used the ClusterMaker plug-in [30] in Cytoscape to quantify the total number of clusters for each chemical transformation. Within the ClusterMaker plug-in, the Markov Clustering (MCL) algorithm provided fast and accurate counts of the number of clusters. The information on the number of clusters was exported to MATLAB for further processing.

Statistical analysis

The non-parametric Spearman's rank correlation and Wilcoxon rank sum test implemented in MATLAB were used to test (1) differences in the number of *m/z* values in negative compared to positive ionization mode, (2) differences in average peak heights, (3) the correlation between formula distributions and depth of the seawater sample, and (4) differences in the number of clusters identified with network analysis.

RESULTS AND DISCUSSION

General characteristics of the ultrahigh resolution mass spectrometry data

The negative and positive ion mode spectra revealed a complex mixture of organic matter with multiple *m/z* values per nominal mass. The total number of *m/z* values (3700 to 5600) was not significantly different between positive and negative ion modes; there were also no significant differences in the number of *m/z* values between the surface and NADW samples (Wilcoxon rank sum tests, p-values > 0.05). The average peak height in each sample was also not

significantly different between surface and NADW samples in either positive or negative ion mode (Wilcoxon rank sum tests, p-values > 0.05, data not shown). The equal number of *m/z* values and lack of differences in average peak height make these data amenable to network analysis because the analysis will not be biased by different numbers of peaks or differences in the ionization strength across the sample set.

Network analysis to characterize organic compounds

We conducted the network analysis with a set of chemical transformations to provide us with new hypotheses regarding the assembly of organic molecules in the deep sea. Due to the time and effort needed to calculate the number of clusters associated with each chemical transformation in all of the samples, we had to limit the number of chemical transformations. A subset of the chemical transformations were based on a published list of chemical transformations from Breitling et al. [31] and we further added both common and random chemical transformations to assess the validity of our conclusions based on network analysis. While Table 3 and Table 4 provide one potential structural isomer associated with each transformation, given the complexity of DOM [3, 7], there are other possible isomers for each chemical transformation. Furthermore, the elemental formula listed in Table 3 and Table 4 may be added to or subtracted from an organic compound in pieces and not as a coherent molecule. Yet, this list of chemical transformations serves as a starting point to investigate new hypotheses regarding the assembly of organic molecules.

Samples were split into two groups: “surface” and “NADW” samples for the network analysis. The network analysis revealed 800 to 1400 clusters for the common chemical transformations (data not shown). However, there was no significant difference in the number of clusters between the surface and NADW samples (Wilcoxon rank sum test). The number of

clusters for the random chemical transformations was lower than for the common transformations, ranging from zero clusters up to a maximum of 300 clusters. Two of the random chemical transformations showed differences between surface and NADW samples ($C_{14}H_9O_3N_7$, $C_{10}H_{12}O_7N_6$). Both of these chemical transformations are listed as compounds with multiple structural isomers in PubChem. Given the complexity of DOM, we were not surprised to randomly find known chemical compounds, and thus the investigation of random chemical compounds is not a good means to test the strength of this analysis tool. Yet, the chemical transformations that were more prevalent in the deep sea could not be due to increased peak heights in the deep sea samples or to increased peak numbers, because neither of these parameters was significantly different between the surface and NADW samples. We also set the same dynamic range in peak heights across the dataset. Finally, only a subset of the chemical transformations showed significant differences between the surface and the deep ocean, and none of the common chemical transformations (i.e., those in Table 2) revealed such differences.

Interpreting the network analysis results

Organic compounds may be formed through mergers of existing molecules, generating new and larger organic compounds [32]. On the other hand, fragments can be removed from existing organic molecules by biological activity such as enzymatic cleavage, as observed in numerous studies [33, 34]. While we cannot use our data to verify either of these hypotheses, the results from the network analysis can guide new ideas about complex organic molecules in the deep sea. In negative ion mode, 12 of the chemical transformations that might be associated with amino acids showed significantly higher numbers of clusters in the NADW samples compared to the surface samples (Figure 3). Furthermore, six of the chemical transformations involving sulfur were significantly different between the surface and NADW samples (Figure 4). None of the

chemical transformations in positive ion mode were significantly different in NADW compared to the surface water samples. Thus, our results suggest that chemical building blocks potentially associated with structural isomers of amino acids and sulfur-containing compounds are more prevalent in the deep ocean than in the overlying surface water masses.

Previous studies have shown that DOM is increasingly refractory with depth [35] and most identifiable biological monomers decrease in relative concentration [1]. The network analysis performed here hints at the continued presence of these monomers in the deep ocean even though they may not be easily accessible by current chemical techniques. It is possible that these transformations do not represent the biological monomer, but instead are a coincidental loss of this combination of elements between two *m/z* values. This seems unlikely due to the fact that an equivalent number of peaks is present in the surface samples but these transformations are more frequent in the NADW samples. Future research must consider if these variants of amino acids and sulfur-containing compounds are available to the microbial community found in the deep ocean. Recent meta-genomic evidence suggests that deep sea microorganisms employ unique metabolic strategies to access refractory organic matter as growth substrates [36, 37]. These metabolic strategies may alter organic matter in unpredictable ways that challenge our current understanding regarding how organic molecules are assembled.

CONCLUSIONS

This combination of network analysis with quantification of the chemical transformations can be used to identify shared structural characteristics of compounds within complex dissolved organic matter. This represents an advance over present tools that focus on the percent of compounds containing certain elements or on changes in elemental ratios (e.g. H:C, O:C) between samples. Our observation that a number of chemical transformations are more prevalent

in a subset of samples is an example of a new way to consider the factors governing the assembly of organic matter. Through the use of ultrahigh resolution mass spectrometry and the continued development of novel computational tools, we will be able to address the next generation of questions regarding dissolved organic matter, such as the nature of metabolic by-products of organic matter remineralization, and their fate in the presence of marine microorganisms.

ACKNOWLEDGEMENTS

We thank Rainer Lohmann for the opportunity to participate in this cruise. Kari Pohl, Lindsey Koren, Hilary Hamer, Rachel Cooper, and Ashley Tucker were helpful during the shipboard sample processing. The help of the captain, crew, and marine technicians of the R/V *Endeavor* was greatly appreciated. Discussions within the Kujawinski lab group helped us consider new ideas about organic matter cycling in the deep sea. The authors thank Melissa C. Kido Soule for FT-ICR-MS assistance and the funding sources for the WHOI FT-MS facility (NSF grant OCE-0619608 and the Gordon and Betty T. Moore Foundation). This work was supported by WHOI's Deep Ocean Exploration Institute (to EBK) and NSF OCE-1154320 (to EBK and KL).

REFERENCES

[1] R. Benner. in *Biogeochemistry of marine dissolved organic matter* (Eds.: D. A. Hansell, C. A. Carlson), Academic Press, **2002**, pp. 59.

[2] E. B. Kujawinski, M. C. Kido Soule, D. L. Valentine, A. K. Boysen, K. Longnecker, M. C. Redmond. Fate of dispersants associated with the Deepwater Horizon oil spill. *Environ. Sci. Technol.* **2011**, *45*, 1298.

[3] N. Hertkorn, M. Frommberger, M. Witt, B. P. Koch, P. Schmitt-Kopplin, E. M. Perdue. Natural organic matter and the event horizon of mass spectrometry. *Anal. Chem.* **2008**, *80*, 8908.

[4] B. P. Koch, T. Dittmar, M. Witt, G. Kattner. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **2007**, *79*, 1758.

[5] T. A. Vetter, E. M. Perdue, E. Ingall, J. F. Koprivnjak, P. H. Pfromm. Combining reverse osmosis and electrodialysis for more complete recovery of dissolved organic matter from seawater. *Sep Purif Technol* **2007**, *56*, 383.

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

[7] T. Kind, O. Fiehn. Metabolomics database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **2006**, *7*, 234.

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

[9] C. Benecke, R. Grund, R. Hohberger, A. Kerber, R. Laue, T. Wieland. MOLGEN(+), a generator of connectivity isomers and stereoisomers for molecular-structure elucidation. *Anal. Chim. Acta* **1995**, *314*, 141.

[10] O. J. Lechtenfeld, G. Kattner, R. Flerus, S. L. McCallister, P. Schmitt-Kopplin, B. P. Koch. Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **2014**, *126*, 321.

[11] S. Kim, R. W. Kramer, P. G. Hatcher. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van krevelen diagram. *Anal. Chem.* **2003**, *75*, 5336.

[12] T. Reemtsma. The carbon versus mass diagram to visualize and exploit FTICR-MS data of natural organic matter. *J Mass Spectrom* **2010**, *45*, 382.

[13] H. A. N. Abdulla, R. L. Sleighter, P. G. Hatcher. Two dimensional correlation analysis of fourier transform ion cyclotron resonance mass spectra of dissolved organic matter: a new graphical analysis of trends. *Anal. Chem.* **2013**, *85*, 3895.

[14] E. B. Kujawinski, K. Longnecker, N. V. Blough, R. Del Vecchio, L. Finlay, J. B. Kitner, S. J. Giovannoni. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2009**, *73*, 4384.

[15] R. Flerus, O. J. Lechtenfeld, B. P. Koch, S. L. McCallister, P. Schmitt-Kopplin, R. Benner, K. Kaiser, G. Kattner. A molecular perspective on the ageing of marine dissolved organic matter. *Biogeosciences* **2012**, *9*, 1935.

[16] K. Faust, J. Raes. Microbial interactions: from networks to models. *Nat Rev Micro* **2012**, *10*, 538.

[17] J. A. Fuhrman, J. A. Steele. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* **2008**, *53*, 69.

[18] P. Schmitt-Kopplin, G. Liger-Belair, B. P. Koch, R. Flerus, G. Kattner, M. Harir, B. Kanawati, M. Lucio, D. Tziotis, N. Hertkorn, I. Gebefügi. Dissolved organic matter in sea spray: a transfer study from marine surface water to aerosols. *Biogeosciences* **2012**, *9*, 1571.

[19] D. Tziotis, N. Hertkorn, P. Schmitt-Kopplin. Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *Eur. J. Mass. Spectrom.* **2011**, *17*, 415.

[20] A. Reinhardt, C. Emmenegger, B. Gerrits, C. Panse, J. Dommen, U. Baltensperger, R. Zenobi, M. Kalberer. Ultrahigh mass resolution and accurate mass measurements as a tool to characterize oligomers in secondary organic aerosols. *Anal. Chem.* **2007**, *79*, 4074.

[21] E. V. Kunenkov, A. S. Kononikhin, I. V. Perminova, N. Hertkorn, A. Gaspar, P. Schmitt-Kopplin, I. A. Popov, A. V. Garmash, E. N. Nikolaev. Total mass difference statistics algorithm: a new approach to identification of high-mass building blocks in electrospray ionization fourier transform ion cyclotron mass spectrometry data of natural organic matter. *Anal. Chem.* **2009**, *81*, 10106.

[22] K. Longnecker. Dissolved organic matter in newly formed sea ice and surface seawater. *Geochim. Cosmochim. Acta* **2015**, *171*, 39.

[23] M. C. Kido Soule, K. Longnecker, S. J. Giovannoni, E. B. Kujawinski. Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter. *Org. Geochem.* **2010**, *41*, 725.

[24] M. P. Bhatia, S. B. Das, K. Longnecker, M. A. Charette, E. B. Kujawinski. Molecular characterization of dissolved organic matter associated with the Greenland ice sheet *Geochim. Cosmochim. Acta* **2010**, *74*, 3768.

[25] A. D. Southam, T. G. Payne, H. J. Cooper, T. N. Arvanitis, M. R. Viant. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier Transform Ion Cyclotron Resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal. Chem.* **2007**, *79*, 4595.

[26] D. Mantini, F. Petrucci, D. Pieragostino, P. Del Boccio, M. Di Nicola, C. Di Ilio, G. Federici, P. Sacchetta, S. Comani, A. Urbani. LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* **2007**, *8*, 101.

[27] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, T. Ideker. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2011**, *27*, 431.

[28] F. Jourdan, R. Breitling, M. P. Barrett, D. Gilbert. MetaNetter: inference and visualization of high-resolution metabolomic networks. *Bioinformatics* **2008**, *24*, 143.

[29] T. Kind, O. Fiehn. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007**, *8*, 105.

[30] J. Morris, L. Apeltsin, A. Newman, J. Baumbach, T. Wittkop, G. Su, G. Bader, T. Ferrin. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* **2011**, *12*, 436.

[31] R. Breitling, S. Ritchie, D. Goodenowe, M. L. Stewart, M. P. Barrett. *Ab initio* prediction of metabolic networks using Fourier transform mass spectrometry data. *Metabolomics* **2006**, *2*, 155.

[32] G. R. Harvey, D. A. Boran, L. A. Chesal, J. M. Tokar. The structure of marine fulvic and humic acids. *Mar. Chem.* **1983**, *12*, 119.

[33] M. McCarthy, T. Pratum, J. Hedges, R. Benner. Chemical composition of dissolved organic nitrogen in the ocean. *Nature* **1997**, *390*, 150.

[34] R. M. W. Amon, R. Benner. Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **1996**, *41*, 41.

[35] C. A. Carlson. in *Biogeochemistry of marine dissolved organic matter* (Eds.: D. A. Hansell, C. A. Carlson), Academic Press, **2002**, pp. 91.

[36] E. F. DeLong, C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N.-U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, D. M. Karl. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **2006**, *311*, 496.

[37] J. McCarren, J. W. Becker, D. J. Repeta, Y. Shi, C. R. Young, R. R. Malmstrom, S. W. Chisholm, E. F. DeLong. Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16420.

393 Table 1. Station and depth information for the samples collected from the equatorial Atlantic
394 Ocean. The temperature, salinity, and oxygen concentrations are given for each sample in
395 addition to the water mass assigned to each sample.

Station	Depth (m)	Temperature (ºC)	Salinity	Oxygen (mg L^{-1})	Water mass
3	5	29.5	28.8	6.4	Surface
3	2500	3.0	34.9	8.2	NADW
3	4500	2.1	34.9	7.9	NADW
5	2	29.6	34.1	6.2	Surface
5	60	28.1	36.2	6.1	Surface
5	700	5.9	34.6	3.9	AAIW
5	1500	4.4	34.9	7.1	NADW
5	2100	3.4	34.9	8.1	NADW
5	2800	2.9	34.9	8.0	NADW
5	3500	2.5	34.9	8.0	NADW

396
397

Table 2. The most commonly observed chemical transformations in the present project. The table reveals the measured mass difference, the elements involved, the rank of the mass difference in negative or positive ion mode. In the Elements column, those elements separated by a "/" indicate an exchange between oxygen (O) and either a methylene (CH_2) or two hydrogens (H_2).

Mass difference	**Elements**	**Negative ion mode**	**Positive ion mode**
12	C	1	2
27.995	CO	2	4
15.995	O	3	1
14.016	CH_2	4	3
44.026	$C_2H_4O_1$	5	5
56.026	C_4H_4O	6	7
28.031	C_2H_4	7	9
1.979	O / CH_2	8	6
13.979	O / H_2	9	11
2.016	H_2	10	10

403 Table 3. Chemical transformations examined using the MetaNetter plugin within Cytoscape.
404 These chemical transformations are a partial version of the list within [31]. The table provides the
405 elemental formulas and exact masses used in MetaNetter and one potential structural isomer for
406 each elemental formula.

Elemental formula	Exact mass	One possible isomer
C_3H_5NO	71.03711384	alanine
$C_6H_{12}N_4O$	156.1011111	arginine
$C_4H_6N_2O_2$	114.0429275	asparagine
$C_4H_5NO_3$	115.0269431	aspartic acid
C_3H_5NOS	103.0091856	cysteine
$C_6H_{10}N_2O_3S_2$	222.0132859	cystine
$C_5H_7NO_3$	129.0425932	glutamic acid
$C_5H_8N_2O_2$	128.0585776	glutamine
C_2H_3NO	57.02146376	glycine
$C_6H_7N_3O$	137.0589119	histidine
$C_6H_{11}NO$	113.0840641	(iso)leucine
$C_6H_{12}N_2O$	128.0949631	lysine
C_5H_9NOS	131.0404858	methionine
C_9H_9NO	147.068414	phenylalanine
C_5H_7NO	97.05276391	proline
$C_3H_5NO_2$	87.03202848	serine
$C_4H_7NO_2$	101.0476785	threonine
$C_{11}H_{10}N_2O$	186.079313	tryptophan
$C_9H_9NO_2$	163.0633286	tyrosine
C_5H_9NO	99.06841398	valine

407

Table 4**.** Additional sulfur-containing chemical transformations assessed using the MetaNetter plugin in Cytoscape. Note that three amino acids (cysteine, cystine, methionine) also contain sulfur.

Elemental formula	Exact mass	One possible isomer
$C_{10}H_{15}N_2O_3S$	243.0803393	biotinyl (-H)
$C_{10}H_{14}N_2O_2S$	226.0775996	biotinyl (-H_2O)
$C_{10}H_{15}N_3O_5S$	289.0732426	glutathione (-H_2O)
$C_{10}H_{15}N_3O_6S$	305.0682	
SO_3	79.95681572	
SO	15.9772	

Figure legends

Figure 1. Simplified version of the results from a network analysis; a complete network is given in Figure 2. Each circle represents a node which is an *m/z* value; as an example, *m/z* values are given for three of the nodes. The nodes are connected by edges which correspond to one of the chemical transformations described in the text.

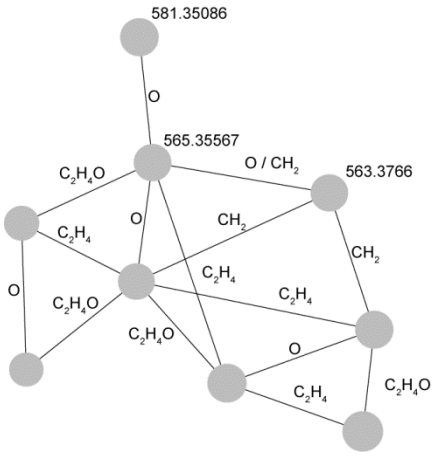
Figure 2. Example of a network calculated during the network analysis using the full set of chemical transformations. Each dot within the figure is an *m*/*z* value found within a sample. The lines connecting the *m/z* values, defined as edges, are chemical transformations. The complete list of chemical transformations is given in Table 3 and Table 4.

Figure 3. Boxplots showing the number of clusters in surface (orange, group 1) and NADW samples (blue, group 2) for transformations with statistically significant differences between the surface and NADW samples. The bottom line in the figure lists the elemental formulas corresponding to the mass differences tested using MetaNetter.

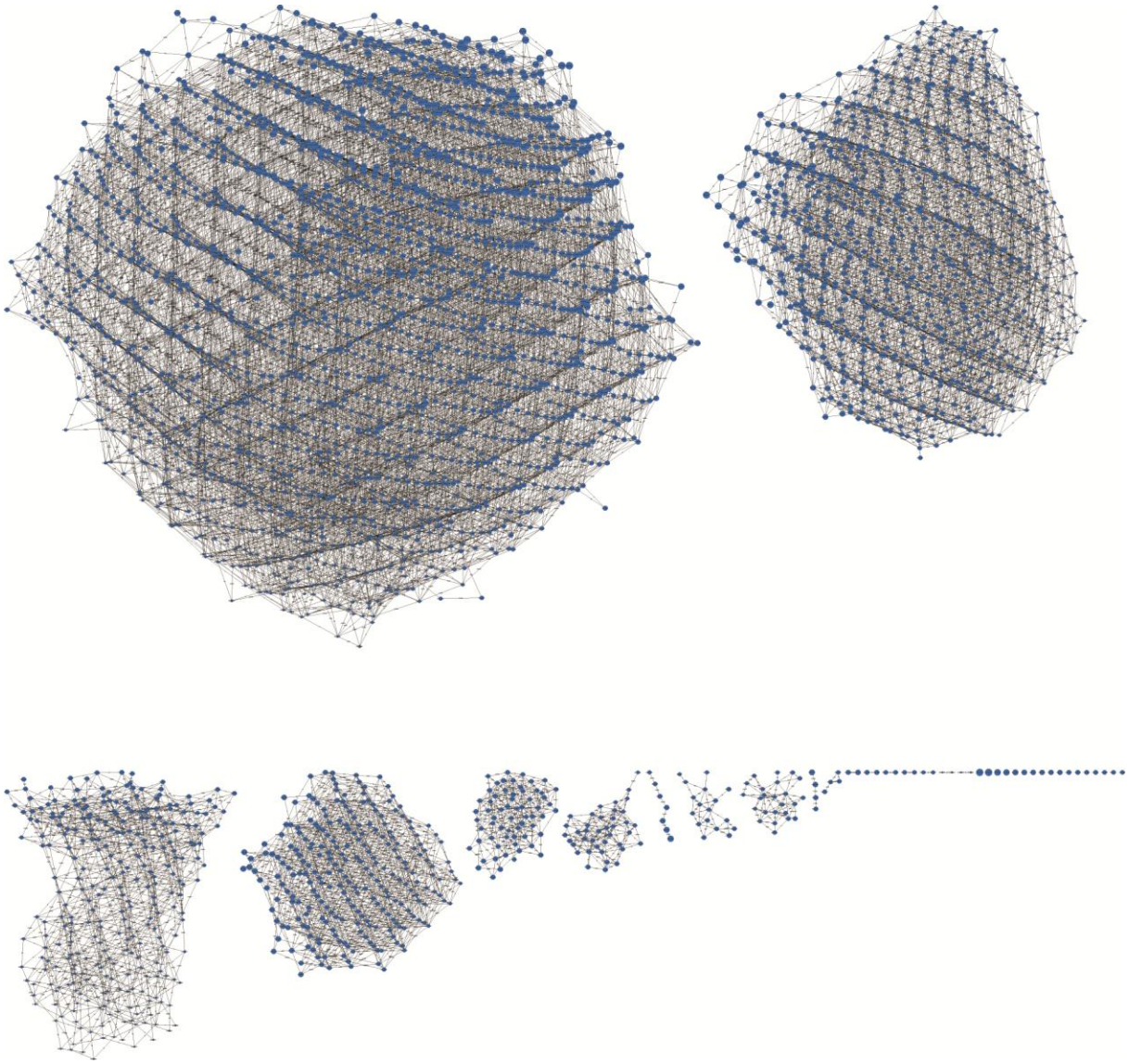
Figure 4. Boxplots showing the number of clusters in surface (orange, group 1) and NADW (blue, group 2) for chemical transformations involving sulfur. The bottom line in the figure lists the elemental formulas corresponding to the mass differences tested using MetaNetter. Only chemical transformations with statistically significant differences between surface and NADW samples are plotted.

Longnecker and Kujawinski

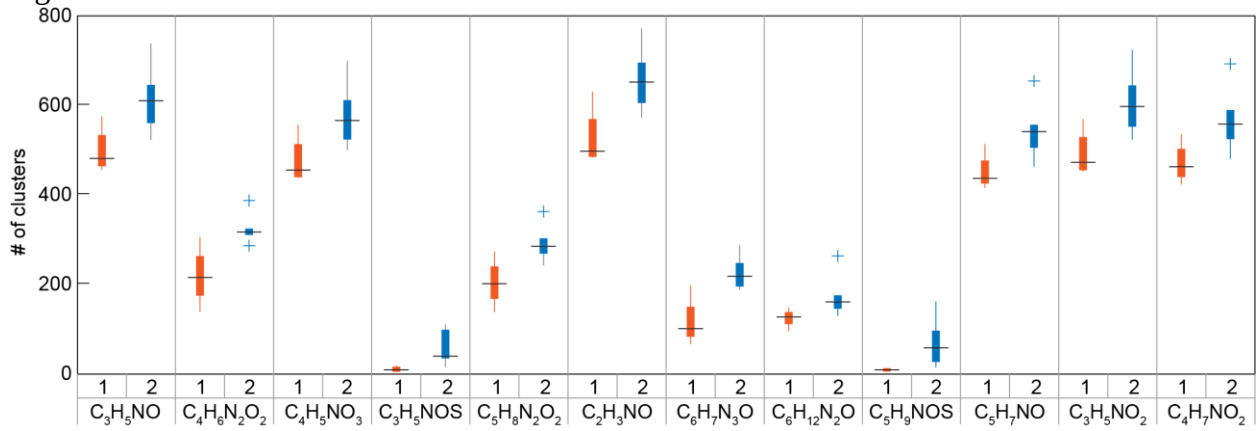
Figure 1



433 Figure 2



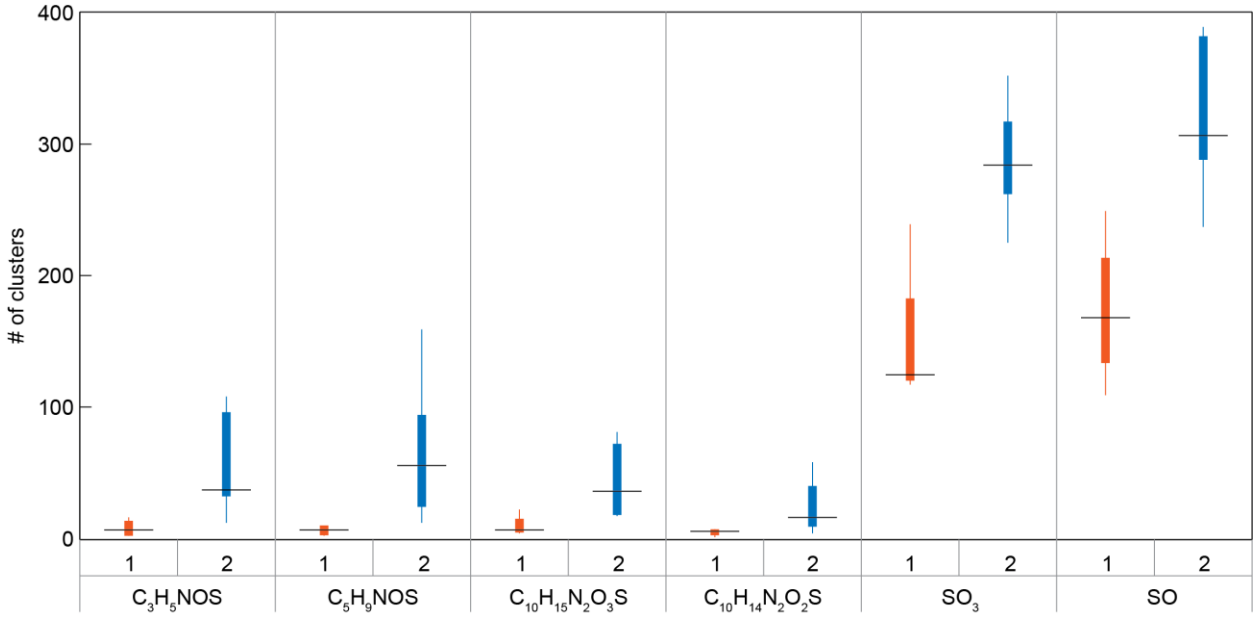
435 Figure 3



436

437 Longnecker and Kujawinski

438 Figure 4



439

440