

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

3 Authors: Krista Longnecker* and Elizabeth B. Kujawinski†

4 Woods Hole Oceanographic Institution, Marine Chemistry and Geochemistry, Woods Hole, MA
5 02543, U.S.A. *klongnecker@whoi.edu ; †ekujawinski@whoi.edu

6 Short title: Network analysis and mass spectrometry

7 Accepted: *Rapid Communications in Mass Spectrometry*

8 *Corresponding author. Mailing address: WHOI MS#4, Woods Hole, MA 02543, USA. Phone:
9 (508) 289-2824. Fax: (508) 457-2164. E-mail: klongnecker@whoi.edu

10

11

12 Temporary citation:

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

17

18

19

20

21

22 **ABSTRACT**

23 **RATIONALE**

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

28 **METHODS**

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

36 **RESULTS**

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

42 **CONCLUSIONS**

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

48

49 **INTRODUCTION**

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

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

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

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

85 **EXPERIMENTAL**

86 **Sample collection**

87 Seawater samples were collected in July 2010 from two stations off the northeastern
88 coast of South America. Samples were divided into surface water samples and water samples
89 from the deeper, North Atlantic Deep Water (NADW, Table 1). Seawater was filtered with 0.2
90 μm Omnipore Filters (Millipore, Massachusetts, USA) mounted in Teflon filter holders, acidified
91 to pH 3, and the dissolved organic matter (DOM) was extracted from the acidified seawater with
92 1g / 6 ml Bond Elut PPL cartridges (Varian, California, USA) ^[6] as previously described ^[22].

93 **Ultrahigh resolution mass spectrometry data collection**

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

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

110 **Peak Detection**

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

121 **Network analysis**

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

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

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

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

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

154 Our second list of possible chemical transformations included chemical building blocks,
155 amino acids, and an organo-sulfur functional group. Amino acids are a large fraction of
156 identifiable organic molecules in marine systems ^[1] and they play a central role in biological
157 processes as the basic structural units of proteins. Thus, our second list of transformations
158 includes m/z values that could be isomers of the twenty essential amino acids (Table 3). We do
159 not intend to convey that we have identified these compounds solely based on their m/z values.
160 Furthermore, we are not searching for the exact mass listed in Table 3, rather we are looking for
161 the mass difference between two measured m/z values that corresponds to the mass listed in the
162 table. We also have an interest in sulfur-containing organic molecules in the deep sea, and Table
163 4 lists six additional elemental formulas, their exact mass, and possible isomers for each
164 chemical transformation.

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

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

184 **Statistical analysis**

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

190 **RESULTS AND DISCUSSION**

191 **General characteristics of the ultrahigh resolution mass spectrometry data**

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

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

202 **Network analysis to characterize organic compounds**

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

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

220 clusters for the random chemical transformations was lower than for the common
221 transformations, ranging from zero clusters up to a maximum of 300 clusters. Two of the random
222 chemical transformations showed differences between surface and NADW samples ($C_{14}H_9O_3N_7$,
223 $C_{10}H_{12}O_7N_6$). Both of these chemical transformations are listed as compounds with multiple
224 structural isomers in PubChem. Given the complexity of DOM, we were not surprised to
225 randomly find known chemical compounds, and thus the investigation of random chemical
226 compounds is not a good means to test the strength of this analysis tool. Yet, the chemical
227 transformations that were more prevalent in the deep sea could not be due to increased peak
228 heights in the deep sea samples or to increased peak numbers, because neither of these
229 parameters was significantly different between the surface and NADW samples. We also set the
230 same dynamic range in peak heights across the dataset. Finally, only a subset of the chemical
231 transformations showed significant differences between the surface and the deep ocean, and none
232 of the common chemical transformations (i.e., those in Table 2) revealed such differences.

233 **Interpreting the network analysis results**

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

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

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

260 **CONCLUSIONS**

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

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

271 **ACKNOWLEDGEMENTS**

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

281

282 **REFERENCES**

- 283 [1] R. Benner. in *Biogeochemistry of marine dissolved organic matter* (Eds.: D. A. Hansell,
284 C. A. Carlson), Academic Press, **2002**, pp. 59.
- 285 [2] E. B. Kujawinski, M. C. Kido Soule, D. L. Valentine, A. K. Boysen, K. Longnecker, M.
286 C. Redmond. Fate of dispersants associated with the Deepwater Horizon oil spill.
287 *Environ. Sci. Technol.* **2011**, *45*, 1298.
- 288 [3] N. Hertkorn, M. Frommberger, M. Witt, B. P. Koch, P. Schmitt-Kopplin, E. M. Perdue.
289 Natural organic matter and the event horizon of mass spectrometry. *Anal. Chem.* **2008**,
290 *80*, 8908.
- 291 [4] B. P. Koch, T. Dittmar, M. Witt, G. Kattner. Fundamentals of molecular formula
292 assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **2007**,
293 *79*, 1758.
- 294 [5] T. A. Vetter, E. M. Perdue, E. Ingall, J. F. Koprivnjak, P. H. Pfromm. Combining reverse
295 osmosis and electro dialysis for more complete recovery of dissolved organic matter from
296 seawater. *Sep Purif Technol* **2007**, *56*, 383.
- 297 [6] T. Dittmar, B. Koch, N. Hertkorn, G. Kattner. A simple and efficient method for the
298 solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol.*
299 *Oceanogr. Meth.* **2008**, *6*, 230.
- 300 [7] T. Kind, O. Fiehn. Metabolomics database annotations via query of elemental
301 compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*
302 **2006**, *7*, 234.
- 303 [8] E. B. Kujawinski, M. D. Behn. Automated analysis of electrospray ionization Fourier-
304 transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.*
305 **2006**, *78*, 4363.
- 306 [9] C. Benecke, R. Grund, R. Hohberger, A. Kerber, R. Laue, T. Wieland. MOLGEN(+), a
307 generator of connectivity isomers and stereoisomers for molecular-structure elucidation.
308 *Anal. Chim. Acta* **1995**, *314*, 141.
- 309 [10] O. J. Lechtenfeld, G. Kattner, R. Flerus, S. L. McCallister, P. Schmitt-Kopplin, B. P.
310 Koch. Molecular transformation and degradation of refractory dissolved organic matter in
311 the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **2014**, *126*, 321.
- 312 [11] S. Kim, R. W. Kramer, P. G. Hatcher. Graphical method for analysis of ultrahigh-
313 resolution broadband mass spectra of natural organic matter, the van krevelen diagram.
314 *Anal. Chem.* **2003**, *75*, 5336.
- 315 [12] T. Reemtsma. The carbon versus mass diagram to visualize and exploit FTICR-MS data
316 of natural organic matter. *J Mass Spectrom* **2010**, *45*, 382.

- 317 [13] H. A. N. Abdulla, R. L. Sleighter, P. G. Hatcher. Two dimensional correlation analysis of
318 fourier transform ion cyclotron resonance mass spectra of dissolved organic matter: a
319 new graphical analysis of trends. *Anal. Chem.* **2013**, *85*, 3895.
- 320 [14] E. B. Kujawinski, K. Longnecker, N. V. Blough, R. Del Vecchio, L. Finlay, J. B. Kitner,
321 S. J. Giovannoni. Identification of possible source markers in marine dissolved organic
322 matter using ultrahigh resolution mass spectrometry. *Geochim. Cosmochim. Acta* **2009**,
323 *73*, 4384.
- 324 [15] R. Flerus, O. J. Lechtenfeld, B. P. Koch, S. L. McCallister, P. Schmitt-Kopplin, R.
325 Benner, K. Kaiser, G. Kattner. A molecular perspective on the ageing of marine dissolved
326 organic matter. *Biogeosciences* **2012**, *9*, 1935.
- 327 [16] K. Faust, J. Raes. Microbial interactions: from networks to models. *Nat Rev Micro* **2012**,
328 *10*, 538.
- 329 [17] J. A. Fuhrman, J. A. Steele. Community structure of marine bacterioplankton: patterns,
330 networks, and relationships to function. *Aquat. Microb. Ecol.* **2008**, *53*, 69.
- 331 [18] P. Schmitt-Kopplin, G. Liger-Belair, B. P. Koch, R. Flerus, G. Kattner, M. Harir, B.
332 Kanawati, M. Lucio, D. Tziotis, N. Hertkorn, I. Gebefügi. Dissolved organic matter in
333 sea spray: a transfer study from marine surface water to aerosols. *Biogeosciences* **2012**, *9*,
334 1571.
- 335 [19] D. Tziotis, N. Hertkorn, P. Schmitt-Kopplin. Kendrick-analogous network visualisation
336 of ion cyclotron resonance Fourier transform mass spectra: improved options for the
337 assignment of elemental compositions and the classification of organic molecular
338 complexity. *Eur. J. Mass. Spectrom.* **2011**, *17*, 415.
- 339 [20] A. Reinhardt, C. Emmenegger, B. Gerrits, C. Panse, J. Dommen, U. Baltensperger, R.
340 Zenobi, M. Kalberer. Ultrahigh mass resolution and accurate mass measurements as a
341 tool to characterize oligomers in secondary organic aerosols. *Anal. Chem.* **2007**, *79*, 4074.
- 342 [21] E. V. Kunenkov, A. S. Kononikhin, I. V. Perminova, N. Hertkorn, A. Gaspar, P. Schmitt-
343 Kopplin, I. A. Popov, A. V. Garmash, E. N. Nikolaev. Total mass difference statistics
344 algorithm: a new approach to identification of high-mass building blocks in electrospray
345 ionization fourier transform ion cyclotron mass spectrometry data of natural organic
346 matter. *Anal. Chem.* **2009**, *81*, 10106.
- 347 [22] K. Longnecker. Dissolved organic matter in newly formed sea ice and surface seawater.
348 *Geochim. Cosmochim. Acta* **2015**, *171*, 39.
- 349 [23] M. C. Kido Soule, K. Longnecker, S. J. Giovannoni, E. B. Kujawinski. Impact of
350 instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-
351 ICR mass spectra of natural organic matter. *Org. Geochem.* **2010**, *41*, 725.

- 352 [24] M. P. Bhatia, S. B. Das, K. Longnecker, M. A. Charette, E. B. Kujawinski. Molecular
353 characterization of dissolved organic matter associated with the Greenland ice sheet
354 *Geochim. Cosmochim. Acta* **2010**, *74*, 3768.
- 355 [25] A. D. Southam, T. G. Payne, H. J. Cooper, T. N. Arvanitis, M. R. Viant. Dynamic range
356 and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier Transform Ion
357 Cyclotron Resonance mass spectrometry-based metabolomics increased by the spectral
358 stitching method. *Anal. Chem.* **2007**, *79*, 4595.
- 359 [26] D. Mantini, F. Petrucci, D. Pieragostino, P. Del Boccio, M. Di Nicola, C. Di Ilio, G.
360 Federici, P. Sacchetta, S. Comani, A. Urbani. LIMPIC: a computational method for the
361 separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* **2007**, *8*,
362 101.
- 363 [27] M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, T. Ideker. Cytoscape 2.8: new features
364 for data integration and network visualization. *Bioinformatics* **2011**, *27*, 431.
- 365 [28] F. Jourdan, R. Breitling, M. P. Barrett, D. Gilbert. MetaNetter: inference and
366 visualization of high-resolution metabolomic networks. *Bioinformatics* **2008**, *24*, 143.
- 367 [29] T. Kind, O. Fiehn. Seven Golden Rules for heuristic filtering of molecular formulas
368 obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007**, *8*, 105.
- 369 [30] J. Morris, L. Apeltsin, A. Newman, J. Baumbach, T. Wittkop, G. Su, G. Bader, T. Ferrin.
370 clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*
371 **2011**, *12*, 436.
- 372 [31] R. Breitling, S. Ritchie, D. Goodenowe, M. L. Stewart, M. P. Barrett. *Ab initio* prediction
373 of metabolic networks using Fourier transform mass spectrometry data. *Metabolomics*
374 **2006**, *2*, 155.
- 375 [32] G. R. Harvey, D. A. Boran, L. A. Chesal, J. M. Tokar. The structure of marine fulvic and
376 humic acids. *Mar. Chem.* **1983**, *12*, 119.
- 377 [33] M. McCarthy, T. Pratum, J. Hedges, R. Benner. Chemical composition of dissolved
378 organic nitrogen in the ocean. *Nature* **1997**, *390*, 150.
- 379 [34] R. M. W. Amon, R. Benner. Bacterial utilization of different size classes of dissolved
380 organic matter. *Limnol. Oceanogr.* **1996**, *41*, 41.
- 381 [35] C. A. Carlson. in *Biogeochemistry of marine dissolved organic matter* (Eds.: D. A.
382 Hansell, C. A. Carlson), Academic Press, **2002**, pp. 91.
- 383 [36] E. F. DeLong, C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N.-U. Frigaard, A.
384 Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, D. M. Karl.
385 Community genomics among stratified microbial assemblages in the ocean's interior.
386 *Science* **2006**, *311*, 496.

387 [37] J. McCarren, J. W. Becker, D. J. Repeta, Y. Shi, C. R. Young, R. R. Malmstrom, S. W.
388 Chisholm, E. F. DeLong. Microbial community transcriptomes reveal microbes and
389 metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc.*
390 *Natl. Acad. Sci. USA* **2010**, *107*, 16420.

391

392

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 (°C)	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

398 Table 2. The most commonly observed chemical transformations in the present project. The table
 399 reveals the measured mass difference, the elements involved, the rank of the mass difference in
 400 negative or positive ion mode. In the Elements column, those elements separated by a “/”
 401 indicate an exchange between oxygen (O) and either a methylene (CH₂) or two hydrogens (H₂).

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 ₂ H ₄ O ₁	5	5
56.026	C ₄ H ₄ O	6	7
28.031	C ₂ H ₄	7	9
1.979	O / CH ₂	8	6
13.979	O / H ₂	9	11
2.016	H ₂	10	10

402

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 ₃ H ₅ NO	71.03711384	alanine
C ₆ H ₁₂ N ₄ O	156.1011111	arginine
C ₄ H ₆ N ₂ O ₂	114.0429275	asparagine
C ₄ H ₅ NO ₃	115.0269431	aspartic acid
C ₃ H ₅ NOS	103.0091856	cysteine
C ₆ H ₁₀ N ₂ O ₃ S ₂	222.0132859	cystine
C ₅ H ₇ NO ₃	129.0425932	glutamic acid
C ₅ H ₈ N ₂ O ₂	128.0585776	glutamine
C ₂ H ₃ NO	57.02146376	glycine
C ₆ H ₇ N ₃ O	137.0589119	histidine
C ₆ H ₁₁ NO	113.0840641	(iso)leucine
C ₆ H ₁₂ N ₂ O	128.0949631	lysine
C ₅ H ₉ NOS	131.0404858	methionine
C ₉ H ₉ NO	147.068414	phenylalanine
C ₅ H ₇ NO	97.05276391	proline
C ₃ H ₅ NO ₂	87.03202848	serine
C ₄ H ₇ NO ₂	101.0476785	threonine
C ₁₁ H ₁₀ N ₂ O	186.079313	tryptophan
C ₉ H ₉ NO ₂	163.0633286	tyrosine
C ₅ H ₉ NO	99.06841398	valine

407

408 Table 4. Additional sulfur-containing chemical transformations assessed using the MetaNetter
409 plugin in Cytoscape. Note that three amino acids (cysteine, cystine, methionine) also contain
410 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 ₂ O)
$C_{10}H_{15}N_3O_5S$	289.0732426	glutathione (-H ₂ O)
$C_{10}H_{15}N_3O_6S$	305.0682	
SO ₃	79.95681572	
SO	15.9772	

411

412 *Figure legends*

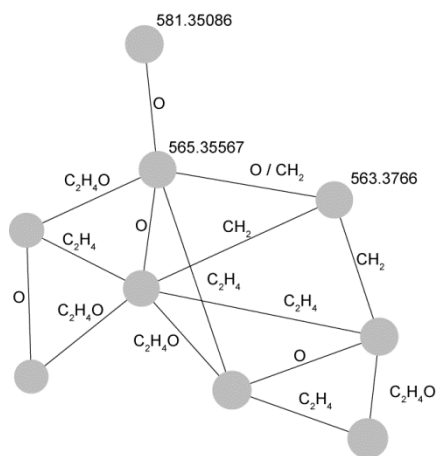
413 Figure 1. Simplified version of the results from a network analysis; a complete network is given
414 in Figure 2. Each circle represents a node which is an m/z value; as an example, m/z values are
415 given for three of the nodes. The nodes are connected by edges which correspond to one of the
416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

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

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

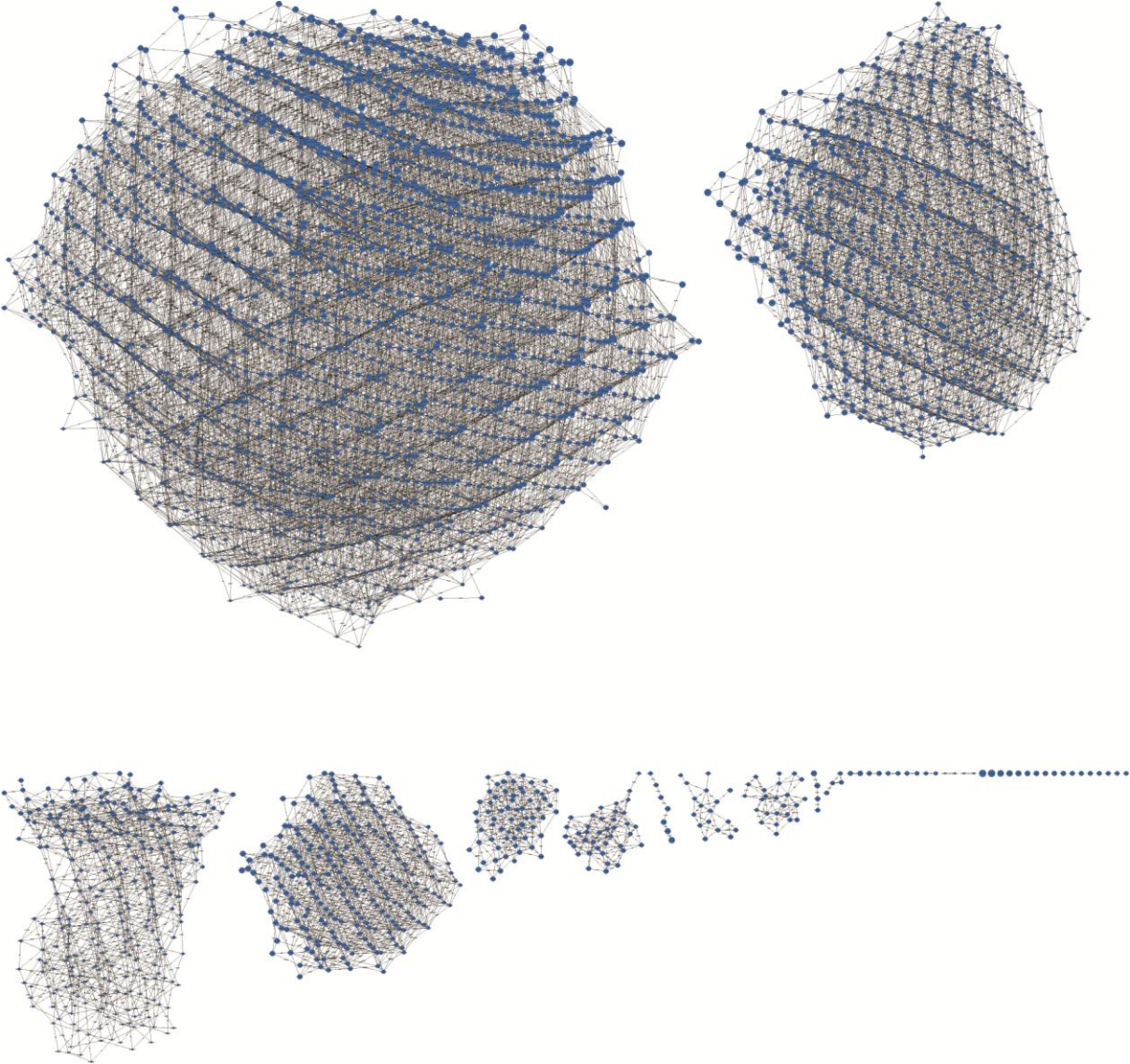
425 Figure 4. Boxplots showing the number of clusters in surface (orange, group 1) and NADW
426 (blue, group 2) for chemical transformations involving sulfur. The bottom line in the figure lists
427 the elemental formulas corresponding to the mass differences tested using MetaNetter. Only
428 chemical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

430 Longnecker and Kujawinski
431 Figure 1



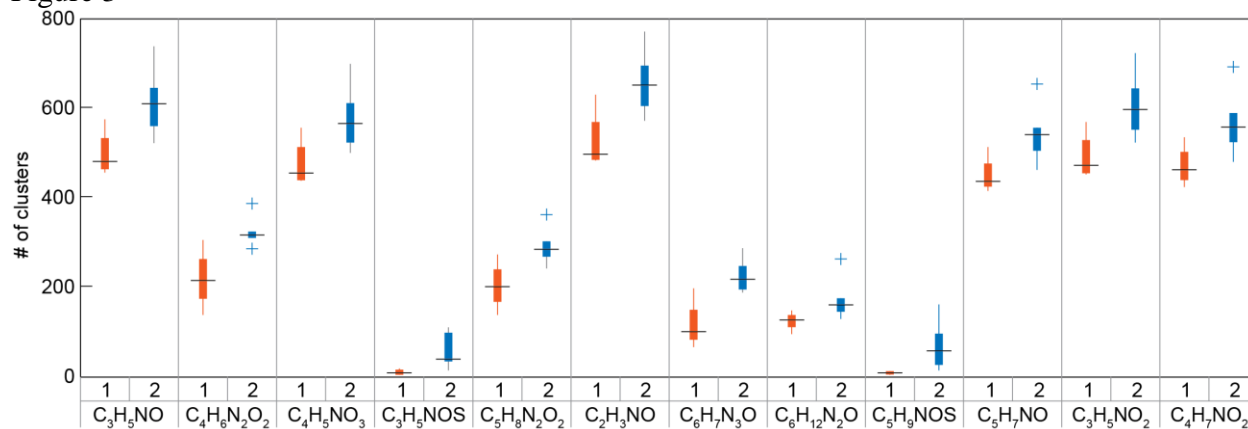
432 Longnecker and Kujawinski

433 Figure 2



434 Longnecker and Kujawinski

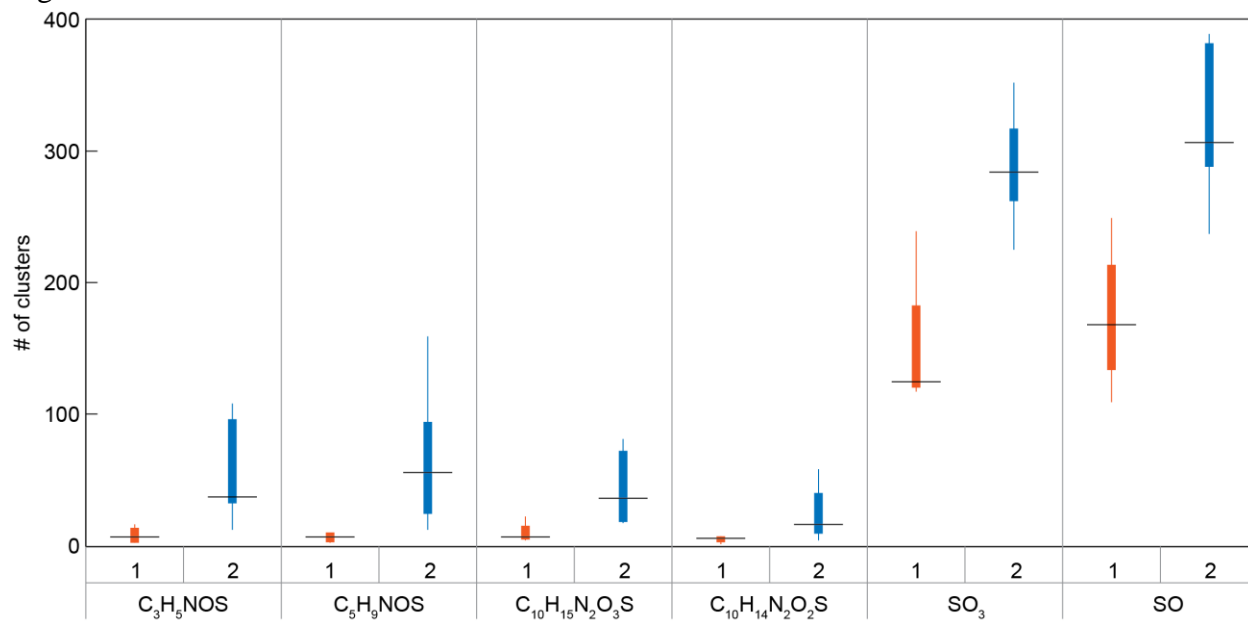
435 Figure 3



436

437 Longnecker and Kujawinski

438 Figure 4



439

440