Unsupervised Structural Classification of Dissolved Organic Matter based on fragmentation pathways.

*Dennys Leyva[a], Muhammad Usman Tariq[b], Rudolf Jaffé[a], Fahad Saeed[b], and Francisco Fernandez Lima[*a,c].*

[a]Department of Chemistry and Biochemistry, Florida International University, Miami, FL, United States.

[b]School of Computing and Information Science, Florida International University, Miami, FL, United States.

[c]Biomolecular Sciences Institute, Florida International University, Miami, Florida, 33199, United.

Keywords: DOM, ESI-FT-ICR MS/MS, neutral loss, precursor, core fragment, network

ABSTRACT: Dissolved organic matter (DOM) is considered an essential component of the Earth's ecological and biogeochemical processes. Structural information of DOM components at the molecular level remains one of the most extraordinary analytical challenges. Advances in chemical formulae determination from molecular studies of DOM have provided limited indications on the structural signatures and potential reaction pathways. In this work, we extend the structural characterization of a wetland DOM sample using precursor and fragment molecular ions obtained by a sequential electrospray ionization – Fourier transform – ion cyclotron resonance tandem mass spectrometry (ESI-FT-ICR CASI-CID MS/MS) approach. The DOM chemical complexity

resulted in near 900 precursors (P) and 24,000 fragments (F) molecular ions over a small *m/z* 261-477 range. The DOM structural content was dissected into families of structurally connected precursors based on neutral mass loss patterns (P_{n-1}+$F_{1:n}$+C) across the 2D MS/MS space. This workflow identified over 1,000 structural families of DOM compounds based on precursor and neutral loss (H_2O, CH_4O and CO_2). Inspection of the structural families showed a high degree of isomeric content (numerous identical fragmentation pathways), not discriminable with sole precursor ion analysis. The connectivity map of the structural families allows for the visualization of potential biogeochemical processes that DOM undergoes throughout its lifetime. This study illustrates that integrating effective computational tools on a comprehensive high resolution mass fragmentation strategy further enables the DOM structural characterization.

1. INTRODUCTION

Decoding the chemical structure of dissolved organic matter remains not only as one of the most interesting but also challenging analytical tasks. Although the molecular features of DOM have been the focus of a multitude of studies over the last decades,[1-7] the elucidation of its compositional structures and a clear view of DOM isomeric complexity persist as one of the most demandingly difficult analytical problems.[8-12]

Nuclear magnetic resonance (NMR)[2,13-18] and hyphenated ultrahigh-resolution mass spectrometry (UHRMS)[9-11,19-22] have been the leading approaches in the structural characterization of DOM. Although most advanced NMR techniques have provided valuable multi-dimensional information on DOM structural characteristics, the extraordinary molecular complexity of this material is still overcoming the NMR capabilities to resolving discrete molecular structures[1,13,23]. On the other hand, analytical approaches integrating ultrahigh-resolution mass spectrometry, gas/liquid separation techniques, and tandem mass spectrometry strategies have provided much of

42 the existing information on the chemical diversity of DOM.[3,9,12,24,25] With the progressive increase
43 in computational power and the high demand in the analysis of complex data, the characterization
44 of DOM at the molecular level has been addressed by molecular dynamics and machine learning
45 approaches as complementary tools to experimental workflows.[26-29]

46 In general, the study of DOM structural complexity using UHRMS have commonly focused on
47 strategies that analyze regular patterns solely based on molecular ions[6,7,15,30-35]. The van Krevelen-
48 type diagram has been the preferred approach to map UHRMS data from complex samples[2,36,37].
49 By plotting O/C vs H/C ratios from the molecular composition, it is possible to visualize clusters
50 of compounds that exhibit similar structural characteristics. Despite that lipid, protein,
51 carbohydrate, tannin, lignin, and carboxylic-rich alicyclic (CRAM) type compound classes have
52 been routinely identified in DOM[38,39], a structural assignment cannot be accurately provided solely
53 on the basis of chemical formulas.[1,40] Kim et.al[36] additionally explored the combination of van
54 Krevelen plots with Kendrick mass defect[37,41,42] to provide structural information based on reaction
55 pathways. For instance, the replacement of 2H by an oxygen atom found along a diagonal of two
56 parallel CH_2 series, was suggested as an oxidation pathway of a primary alcohol to a carboxylic
57 acid.

58 Several parameters derived from chemical formulas are also commonly utilized to predict
59 structural signatures and compositional trends of DOM molecular species. For instance, double
60 bond equivalents (DBE) and the aromaticity index are used to estimate the degree of structure
61 unsaturation and identify aromatic/condensed species in DOM components respectively.[43,44]
62 Furthermore, the occurrence of several regular patterns in DOM and their potential correlation
63 with families of structurally related compounds has been also reported[1]. However, an explanation
64 on the origin of these regularities and the structural correlation among the compounds belonging

65 to the homologous families has not been yet provided. Reports based on tandem mass spectrometry
66 of selected molecular ions have shown promise for the identification of DOM structural
67 features[4,20,22,45-47].

68 A FT-ICR MS/MS study[47] of solid-phase-extracted (SPE)-atmospheric organic matter has
69 suggested that structural analogies could exist among members of a CH_2 homologous series since
70 they share identical neutral losses during collision induced dissociation (CID). Similarly, several
71 regular patterns in DOM chemical formulas such as H_2 and CH_2 series and a replacement of a CH_4
72 by an oxygen atom have been found using molecular level analyses by FT-ICR MS[23,45] . Although
73 the CH_4 vs O pattern has not been clearly explained from a structural perspective[48], it has been
74 attributed to potential interchanges of functionalities (e.g. C_2H_5 vs CHO). Moreover, in a different
75 contribution, the same authors[49] utilized known degradation pathways observed for lignin (a
76 possible component of DOM) to structurally explain newly found repeating patterns in DOM
77 chemical formulas. Interestingly, the O_2 and CH_4 vs O_2 regularities found in DOM components
78 were correlated with aromatic ring openings ($+O_2$) and a combination of aromatic ring openings
79 after one demethylation ($-CH_2$) and one side-chain oxidation ($-H_2$) respectively.

80 New structural insights into the H_2 and CH_2 homologous series from low molecular weight
81 compounds of Suwannee River fulvic acid standard using size exclusion chromatography-
82 electrospray ionization-time-of-flight (TOF) tandem mass spectrometry, have been reported by
83 These et.al[48]. The similarity found in the fragmentation patterns of homologous isolated precursors
84 (fragments exhibiting the same H_2 or CH_2 difference as their corresponding precursors), suggested
85 that structural dissimilarities among family members presumably lied on their corresponding core
86 structures.

87 The structural complexity of marine DOM using ultrahigh-resolution tandem mass analysis
88 based on an orbitrap MS/MS workflow was explored by Cortes-Francisco et.al[20]. Although this
89 study was not oriented to the analysis of structural regularities found in DOM, the potential
90 fragmentation pathways proposed for one of the precursor ions attributed to a lipid-like compound,
91 showed the utility of integrating MS/MS data with van Krevelen information to provide new
92 structural understandings of DOM components.

93 The advantages of DOM analysis using complementary trapped ion mobility spectrometry-FT-
94 ICR MS/MS with correlated harmonic excitation field have been shown[12]; while this work
95 demonstrated the isolation by mobility and tandem MS/MS at the level of chemical formula, its
96 routine application is unviable due to the large number of isomers and isobars present in DOM
97 samples. There is a need for simplified strategies capable of establishing structural patterns based
98 on MS1 and tandem MS/MS information using shorter experimental and processing time scales.
99 In this report, we propose a systematic nominal mass UHR MS/MS follow by a computational
100 model capable of correlating structural features (or families) based on the fragmentation pathways
101 of precursor molecules.

102 Data independent acquisition (DIA) is an acquisition strategy in mass spectrometry based on
103 parallel collection of MS/MS spectra and has recently been utilized to improve the signal-to-noise
104 ratio, reproducibility, and ultimate analyte coverage.[50-52] Recent advances in computing power and
105 electronics have enabled 2D FT-ICR MS as an emerging DIA tool to analyze complex mixtures.[53-
106 55] The application of an RF pulse sequence to manipulate the ion's cyclotron radii in the ICR
107 cell[56,57], along with no ion isolation and ion-neutral collisions (infrared multi-photon dissociation
108 and electron capture dissociation are mostly used), led to the correlation of precursor and fragment
109 ion signals with enhanced resolution and sensitivity. Nevertheless, the presence of abundant

110 scintillation noise[58] and difficulties associated with data processing, are still important limitations
111 that need to be addressed to obtain comprehensive MS/MS data.
112 The introduction of continuous accumulation of selected ions (CASI) in FT-ICR MS instrument
113 by Senko et.al[59] provided a way to increase the sensitivity and dynamic range, while reducing
114 space charge effects by sequentially transmitting smaller m/z segments. More recently, this
115 strategy has also been implemented in top-down mass spectrometry[60] and protein imaging[61],
116 respectively. In the case of DOM analysis, CASI has allowed for the detection of a larger number
117 of chemical formulas when compared to traditional broadband acquisitions[23,62]. Despite of the
118 increase on the number of chemical formulas, there is a need for further CASI implementations
119 combined with sequential fragmentation (CASI MS/MS). In the case of complex mixtures
120 analysis, CASI MS/MS workflows can greatly benefit from new computational algorithms for
121 MS/MS data processing and structural correlations.
122 In this work, we extend the structural characterization of a wetland DOM sample from Pantanal,
123 Brazil, using precursor and fragment molecular ions obtained with electrospray ionization-Fourier
124 transform – ion cyclotron resonance tandem mass spectrometry (ESI-FT-ICR CASI CID MS/MS).
125 Families of structurally related DOM compounds are identified based on characteristic mass loss
126 patterns across heteroatom classes. We propose a novel graphical analysis of interconnected
127 structural families as a potential tool that helps to understand DOM biogeochemical processes.

128 **2. MATERIALS AND METHODS**

129 **2.1 Sample Preparation**. The DOM sample was obtained by SPE of surface water collected
130 from wetlands located at Pantanal National Park, Brazil. Details on sampling, sample treatment
131 and the SPE procedure are described by Hertkorn et.al[2] and Dittmar et. al[63]. Briefly, 2L of surface
132 water were collected using HCl pre-cleaned brown plastic bottles. Samples were kept refrigerated

133 on ice and filtered using GFF pre-combusted glass fiber filters (0.7 μm nominal pore size) within
134 6 h after collection. Filtered samples acidified to a pH 2 with concentrated HCl, were loaded by
135 gravity onto a 1g-Varian Bond Elut PPL cartridge using Teflon tubing. The PPL cartridge was
136 preconditioned with methanol followed by pH 2 Milli-Q water. The loaded cartridge was then
137 rinse with pH 2 Milli-Q water and dried in a N2 gas flow for five minutes prior to the elution of
138 DOM molecules with 20 mL of methanol. SPE-DOM extracts were stored in pre-combusted glass
139 vials at -20 °C until further analysis. The choice of the sample comes from its recent UHRMS and
140 IMS-UHRMS characterization (recent papers[11,12] and a 2016 report by Hertkorn et.al[2]). The SPE-
141 DOM sample was diluted ten times by dissolving it in 1 mL of denatured ethanol. All solvents
142 used were of Optima LC-MS grade or better, obtained from Fisher Scientific (Pittsburgh, PA).
143 **2.2 ESI-FT-ICR-MS**. A SolariX 9T ESI-FT-ICR MS spectrometer (Bruker Daltonics, MA)
144 equipped with an infinity ICR cell was optimized for high transmission in the 100-1200 m/z range.
145 Samples were ionized using an electrospray ionization source (Apollo II ESI design, Bruker
146 Daltonics, Inc., MA) in negative ion mode at 200 μL/h injection. Typical operating conditions
147 were 3700 - 4200 V capillary voltage, 2 L/min dry gas flow rate, 2.0 bar nebulizer gas pressure,
148 and a dry gas temperature 200 °C. Operational parameters were as follows: funnel rf amplitude
149 160 peak-to-peak voltage (Vpp), capillary exit -150 V, deflector plate -140 V, skimmer1 -20 V,
150 transfer line RF 350 Vpp, octupole RF amplitude 350 Vpp and collision cell RF 1100 Vpp. An
151 Arginine cluster ion series (173-1740 Da) was used during the instrument tuning and control
152 optimization. The broadband MS_1 spectrum (first MS dimension) of 115 co-added scans was
153 collected at 4 MW data acquisition size (mass resolution of 4M at 400 *m/z*).
154 **2.3 ESI-FT-ICR CASI CID MS/MS.** For the CASI-CID experiments, ions at odd nominal
155 masses were sequentially isolated (1 Da window) in the quadrupole (*m/z* range 261-477),

accumulated for 5-7 s in the collision cell, and subject to CID prior to the analysis in the ICR cell. Multiple CID collision voltages (15 V – 27 V) tailored to the precursor nominal *m/z* were utilized for a better coverage across low and high *m/z* fragments. The same ion optics parameters used in broadband analysis were utilized during the MS/MS experiments. Up to 100 scans were co-added for each tandem mass spectrum (MS_2) in the segmental acquisition mode. Eight predefined segments were acquired and stitched for each experiment using the serial run mode.

2.4 ESI-FT-ICR CHEF SORI MS/MS. Differences between nominal mass and chemical formula-based MS/MS were evaluated for the case of the 267.087412 *m/z* ion ($C_{13}H_{15}O_6$) using correlated harmonic excitation field (CHEF)[12,46,64,65], shots ejection of isobaric ions (~0.002% power and 0.04 pulse length) and sustained off resonance irradiation (SORI)-CID (1.4% SORI power, 0.1 s pulse length of and -500 Hz frequency offset). A sweep excitation was applied, and six hundred MS/MS scans were collected at 2 MW data size.

2.5 Data Processing. Data was processed using Data Analysis (v. 5.2, Bruker Daltonics, CA), and all other plots were created using OriginPro 2016 (Originlab Co., MA). Chemical formulas assignment was conducted using Composer software (version 1.0.6, Sierra Analytics, CA, USA) and confirmed with Data Analysis (version 5.2, Bruker Daltonics). The formulas assignment was based on lowest formula errors, the presence of isotopologue signals and the removal of isolated assignments (de-assignment of peaks belonging to classes with only a few sparsely scattered members). Theoretical formula constraints of $C_{4\text{-}50}H_{4\text{-}100}N_{0\text{–}3}O_{0\text{–}25}S_{0\text{–}2}$, S/N>3, m/z range 100-900, error <1 ppm and $0<O/C\leq2$, $0.3\leq H/C\leq2.5$, and DBE-O ≤ 10[66] were considered. The internal walking calibration performed in Composer using oxygen homologous series (O_4-O_{20}) resulted in an average error < 80 ppb for the mass range 229-890 Da. Both odd and even electron configurations were allowed in Data Analysis software. The MS/MS spectra were internally

179 calibrated using a list of exact masses of fragment ions obtained from commonly occurring neutral
180 losses in DOM and their combinations[8,22]. A four column excel file containing (1) the accurate
181 mass of assigned peaks from both MS2 and MS1 (odd masses *m/z* 261-477), (2) the isolated
182 nominal mass, (3) the intensity, and (4) the chemical formulas was created as input file for further
183 data processing using Graph-DOM, an in-house code written in Python 3.7.3.
184 Ordered fragmentation pathways were computed based on the following equation:
185 $P = [NL_1 + NL_2 + NL_3 + \cdots NL_n] + C$ (1),
186 where P corresponds to the chemical formula of the isolated precursor at nominal mass and NL
187 is the neutral specie lost during the fragmentation of precursor and fragment ions. In this study,
188 CH_4, O, H_2O, CO, CH_2O, CH_4O, and CO_2 were considered as potential neutral losses[20,46,67,68]. The
189 sequence [NL_1-NL_2-NL_3-…NL_n] in equation (1) is an ordered array of neutral losses (NL)
190 generated by an approach similar to the one recently described by Simon et.al[67]. Differently from
191 Simon's approach, we sequentially match the exact mass of the theoretical NL with the mass
192 difference of two consecutive assigned peaks with 1 mDa tolerance error. The core fragment (C)
193 was defined as the lowest mass assigned fragment in a given pathway. Note that this approach also
194 considers the search of multiple NL if the mass difference between two peaks does not match the
195 accurate mass of a single NL. Due to the large amount of fragment data collected in this study, we
196 set the NL multiple at 2. Nevertheless, the Graph-DOM code allows the user to define both the
197 type of NL and its multiples.
198 Families of structurally related compounds were identified using a conceptual model (P_{n-}
199 $_{1}$+$F_{1:n}$+C), defined in the Graph-DOM code, based on *de novo* matching of fragmentation
200 pathways. Briefly, a precursor chemical formula along with the full fragmentation pathway is
201 searched across all computed pathways in ascending order of mass. Note that the n-1 subscript in
202 the model indicates the presence of P_{n-1} precursor as first fragment of P_n's fragmentation pathway

203 (See Figure 3 panel B). The term $F_{1:n}$ defines the full match condition for all fragments in the
204 pathway to consider a precursor in a family. Cytoscape v.3.8.2[69] was used to visualize the
205 complexity of DOM in the form of structural networks formed by a neutral loss-based
206 interconnection of family members. A list of the precursors found in the structural families was
207 imported into Cytoscape and defined as nodes. Structural functionalities based on neutral loss
208 differences among precursors in a family were imported as edges.

209 **3. Results and discussion**

210 The broadband ESI-FT-ICR MS spectrum of the SPE-DOM sample showed a typical
211 distribution of $[M-H]^-$ ion signals with a maximum around 400 *m/z* (Figure 1A). A section of the
212 spectrum (4Da mass range) depicts the characteristic DOM pattern of most abundant signals
213 located at every other odd *m/z* and lower intensity peaks at even m/z (See inset, Figure 1A). The
214 van Krevelen plot (Figure 1B) obtained after assigning near 4,000 molecular formulas, showed a
215 dominance of CHO (green) and CHON (orange) heteroatoms classes in the region 0.3<O/C<0.8 -
216 0.4<H/C<1.8 attributed to lignin and tannins type molecules followed by less abundant CHOS
217 (blue) compound classes associated to sulfonated carboxylic-rich alicyclic components[2,36,39,70].

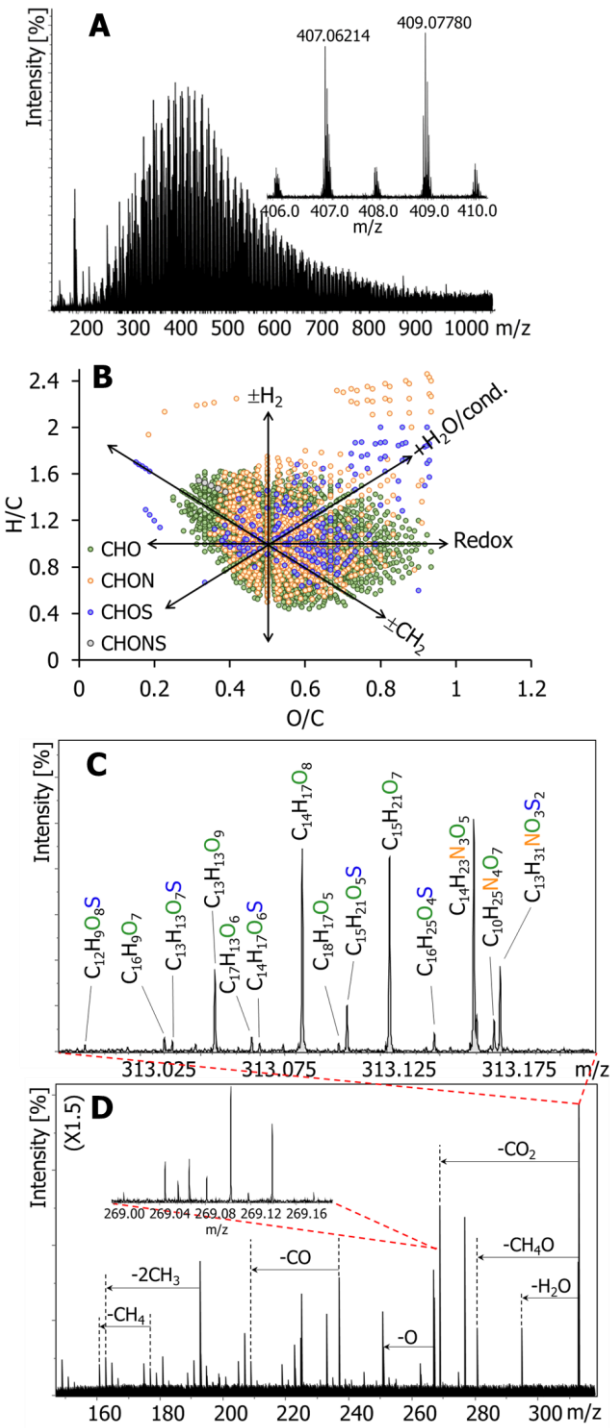


Figure 1. ESI-FT-ICR MS broadband spectrum of the SPE-DOM sample and expanded view of the *m/z* range 406-410 shown in the inset (A). Van Krevelen plot obtained after chemical formula assignment of mass signals with black arrows describing DOM reaction pathways previously

222 suggested by Kim et.al[36]. CHO, CHOS, CHON and CHONS compound classes are represented in
223 green, orange, blue and grey colors respectively (B). Section of a MS/MS spectrum showing [M-
224 H]$^-$ precursor ions isolated at nominal mass 313. Assigned molecular formulas are displayed with
225 heteroatoms indicated with the color code (C). Typical MS/MS spectrum of the precursors isolated
226 at nominal *m/z* 313 with annotated common neutral losses observed in DOM (D). Note that single
227 peaks showed at nominal masses may comprise an envelope of multiple mass signals. For instance,
228 nine peaks resulting from the CO_2 loss of precursors fragmented at m/z 313 are shown at *m/z* 269
229 (Panel D, inset).

230 A closer view of the nominal mass 313 (Figure 1C), shows the characteristic isobaric complexity
231 of the sample, where up to 14 precursor ions of the CHO, CHON, CHOS and CHONS classes
232 were co-isolated and fragmented. Similar patterns resulting in an average of 10 precursor ions per
233 MS_2 spectrum (total 110 fragment spectra collected) are found across the studied mass range (*m/z*
234 261-477). Typical fragmentation patterns showing common DOM neutral losses of CH_4, H_2O, CO,
235 CH_4O, and CO_2 were observed across the fragmentation data set (See the MS_2 profile of precursors
236 isolated at 313 *m/z* in Figure 1D). Few other less abundant neutral losses associated with sulfur
237 (SO_3) and nitrogen (NH_2OH and HNO_3) species were also observed.

238 The analysis of the potential reaction pathways previously reported for DOM[36] and described by
239 black arrows in Figure 1B, suggests that compounds found along a pathway (e.g. Redox) in the
240 van Krevelen space are part of a structural family with a potential common backbone. Since
241 structural questions are difficult, if not impossible, to answer solely based on chemical composition
242 obtained from UHRMS, here we explored a fragmentation strategy that will provide new
243 information about the structural complexity of DOM as a complementary tool to the traditional
244 van Krevelen plot.

245 The application of the ESI-FT-ICR CASI CID MS/MS workflow resulted in more than 24,000
246 total assigned chemical formulas (~900 precursors). The CHO constituted the most abundant
247 compound class (80% of all the precursors assigned), followed by the CHOS (~17%) and CHON
248 (<3%) classes (Figure S1). 2D MS/MS plots generated using all identified molecular formulas (A)
249 and the filtered *m/z* signals assigned to the CHO, CHON and CHOS compound classes respectively
250 (B-D) are shown in Figure 2. A closer view to the panels B-D in Figure 2 confirmed the clear
251 dominance of the CHO compounds during fragmentation (>23, 000 chemical formulas) over the
252 less abundant CHON and CHOS compound classes. Consequently, the O-heteroatom class will
253 constitute the main focus of this study.

254 Similar to the 2D mass spectrum described by van Agthoven[54], in our 2D MS/MS plot of
255 Fragment *m/z* vs Precursor *m/z*, typical straight lines can be observed. Examination of the Figure
256 2A denotes that data points are aligned over diagonal lines described by the equation (2):

257 $$Precursor\left(\frac{m}{z}\right) = Fragment\left(\frac{m}{z}\right) + neutral\ loss \qquad (2)$$

258 The first diagonal line observed (right towards left) represents the precursor line and it contains
259 the precursor ions. Since the chemical formula assignment was based on accurate mass (error < 1
260 ppm), precursors and fragments can be directly correlated (data points horizontally aligned in the
261 2D MS/MS domain)[8,20,22]. Neutral loss lines are parallel to the precursor line and the NL mass
262 (relative to precursor line) can be determined by the intercept of the equation (2). For instance, the
263 characteristic line of one H_2O loss (first line from precursor line in Figure 2B) can be described
264 using the equation $Precursor\left(\frac{m}{z}\right) = Fragment\left(\frac{m}{z}\right) + 18$. Other typical NLs observed (e.g.,
265 CO, CH_4O, CO_2, etc.) and their corresponding multiples can be visualized in the form of their
266 characteristic lines.

267

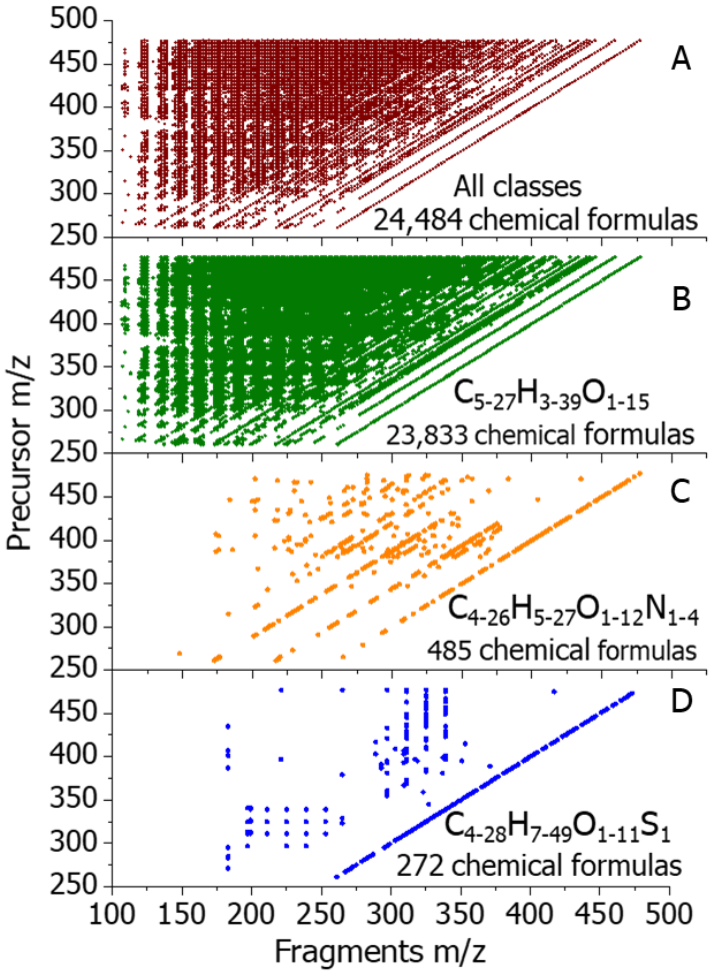


Figure 2. 2D MSMS plots generated after chemical formula assignment of ion signals obtained from the FT-ICR CASI-CID MS/MS experiments.

The alignment of the MS/MS data along unique NL lines observed in Figure 2 A-C evidenced the similarity of DOM fragmentation pathways regardless of the precursor chemical composition. These structural patterns are in good agreement with previous findings obtained from fragmentation experiments of few selected nominal masses[8,46,47,71]. The systematic occurrence of NL-line patterns in the 2D MS/MS space, suggests that DOM molecules are clustered by families of compounds that could likely share common backbone structures.

277 The analysis of the complex fragmentation data generated from the FT-ICR CASI-CID MS/MS
278 experiments was performed by designing an efficient data mining approach implemented in the
279 Graph-DOM (Figure 3). The first step consisted of computing all possible ordered fragmentation
280 pathways for the assigned precursors (Figure 3, panel A) using equation (1). For instance, a
281 fragmentation pathway for the precursor $C_{16}H_{19}O_9$ is described as $C_{16}H_{19}O_9$ =
282 $[H_2O+CO_2+CH_4O+CO_2]$ + $C_{13}H_{13}O_3$. Since NLs are directly correlated with structural
283 functionalities, the H_2O, CH_4O and CO_2 chemical units could be interpreted as CID fragments
284 associated with hydroxyl, methoxy and carboxylic moieties, respectively. Since a precursor
285 formula comprises a variety of isomeric species, multiple fragmentation pathways (with the same
286 or different core fragments) can be associated with the same precursor[11]. Note that the core
287 fragment chemical formula could be interpreted as the backbone of the precursor structure and can
288 also hold isomeric diversity. In the examples shown, the core fragment is limited by the lower m/z
289 experimentally observed (in this instrument and settings, *m/z* below 100 are not detected).
290 Over 10^7 ordered fragmentation pathways were computed for the CHO compound class
291 following the workflow described in Figure 3A. Precursor compounds within the mass range 395-
292 477 exhibited the highest number of fragmentation pathways (>100,000), in agreement with the
293 extensive amount of product ions detected (See Figure S1). An average of 7 million pathways was
294 found for precursor molecules containing 11-13 oxygens in their composition. On the other hand,
295 less oxygenated DOM compounds (8-12 oxygens) generated a larger number of core fragments
296 (Figure S2). The relative high abundance of fragmentation pathways and core fragments for O-
297 rich molecules suggests that the degree of oxygenation plays a key role on DOM structural
298 diversity.

299 Assuming that the fragmentation pathway of a precursor $C_xH_yO_z$ is fully matched to the pathway
300 of another precursor $C_{x+a}H_{y+b}O_{z+c}$, we could presume that they are structurally related.
301 Consequently, the compositional difference between these two precursors will be the chemical unit
302 $C_aH_bO_c$. Since many of the fragments assigned in the MS/MS spectra are also observed in the MS_1
303 domain, other precursors will likely show the same behavior as both $C_xH_yO_z$ and $C_{x+a}H_{y+b}O_{z+c}$.
304 Therefore, they can be grouped into families characterized by a NL-based sequence resulting from
305 the difference in chemical units among precursors.

306 The computation of structural families of DOM was conducted by implementing the conceptual
307 model $P_{n-1}+F_{1:n}+C$ graphically described in Figure 3B. An overlapping strategy of the
308 fragmentation pathways in the form of $P=[F_1+F_{2+}\ldots+Fn]+C$ was utilized. The overlap step
309 consisted of matching both the initial lowest mass precursor P_1 and its fragmentation pathway in
310 the database generated from the previous step (Figure 3, panel A) in ascending order of mass. The
311 initial precursor P_1 is further grouped into the family $[P_1, P_{n-1}]$ with the newly matched precursor
312 P_{n-1} and the chemical unit difference $NL_{Pn-1\rightarrow P1}$ is stored as the structural difference between P_1
313 and P_{n-1}. The resulting pathway $P_{n-1}= P_1+[F_1+F_{2+}\ldots+Fn] + C_1$ is searched again for a new match
314 and the loop is repeated until no further match is found. Finally, the family ($[P_1, P_2,\ldots, P_{n-1}, P_n]$) is
315 created as an array of the precursors sharing the same fragmentation pathways. The chemical unit
316 difference identified as a neutral loss among precursors within a family represents the functionality
317 that is being added /subtracted to/from the family members. This array of neutral loss-based
318 moieties illustrates the potential biogeochemical transformation processes experienced by DOM
319 molecules. Once a family is retrieved, a new precursor higher in mass than P_1 is reset as initial
320 lowest mass precursor and the pathway matching algorithm is repeated until all potential families
321 are computed. Note that since various fragmentation pathways might be common to different

322 precursors, multiple identical structural families will be expected. We define these sequences of
323 analogous precursors as isomeric families, and they are an important indication of the confidence
324 during the computation of the families.

325 The model performance to retrieve CHO structural families (coverage of precursors,
326 intermediate fragments and core fragments) is described in Figure 4. Although the coverage of
327 precursors in the families was 60%, over 2,000 DOM structural families were identified. A higher
328 coverage was found for both the intermediate (~90%) and core fragments (> 80%). Note that since
329 the same core and intermediate fragments might be found at different nominal masses, those
330 fragments were counted in the families every time they were linked with a different precursor.
331 These results suggests that there are potential structural families that remain undetected under the
332 current conceptual model. While our workflow provides higher confidence, in grouping
333 structurally related DOM compounds, than previous approaches, there are still limitations
334 associated with the considerations of the proposed model.

335 Structural families containing two-to-four precursor members of the CHO class were the most
336 abundant (261-477 *m/z* range). A decrease in the number of families was also observed as the
337 family size (number of precursors in a family) increased from four to six members (Figure 4B).
338 Up to five precursors were found in over 300 structural families and the lowest abundant family
339 (< 100) contained six DOM compounds. The relatively high number of 2-members families (>400)
340 could be attributed to the limited mass range analyzed in the current study, preventing the match
341 of fragmentation pathways from precursors with higher mass (>477).

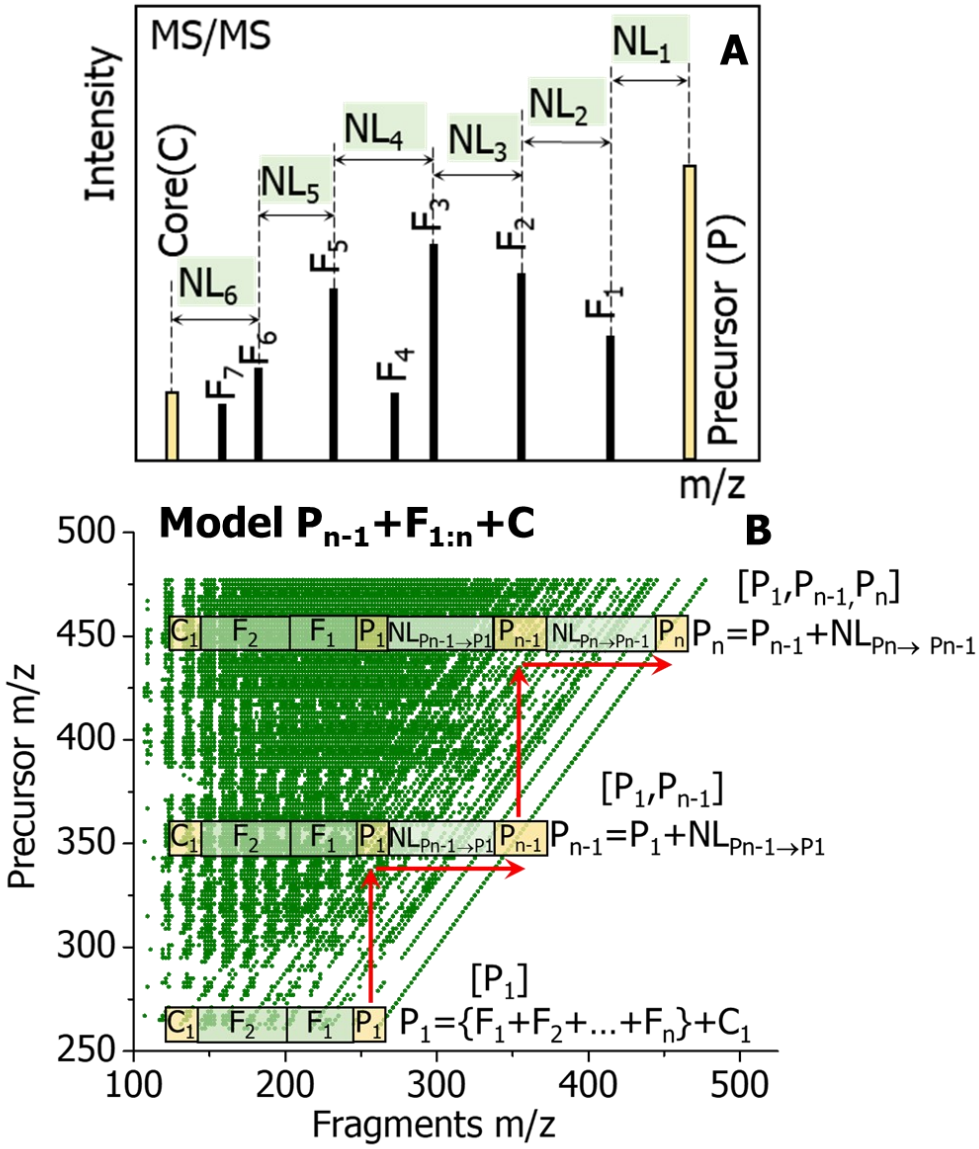
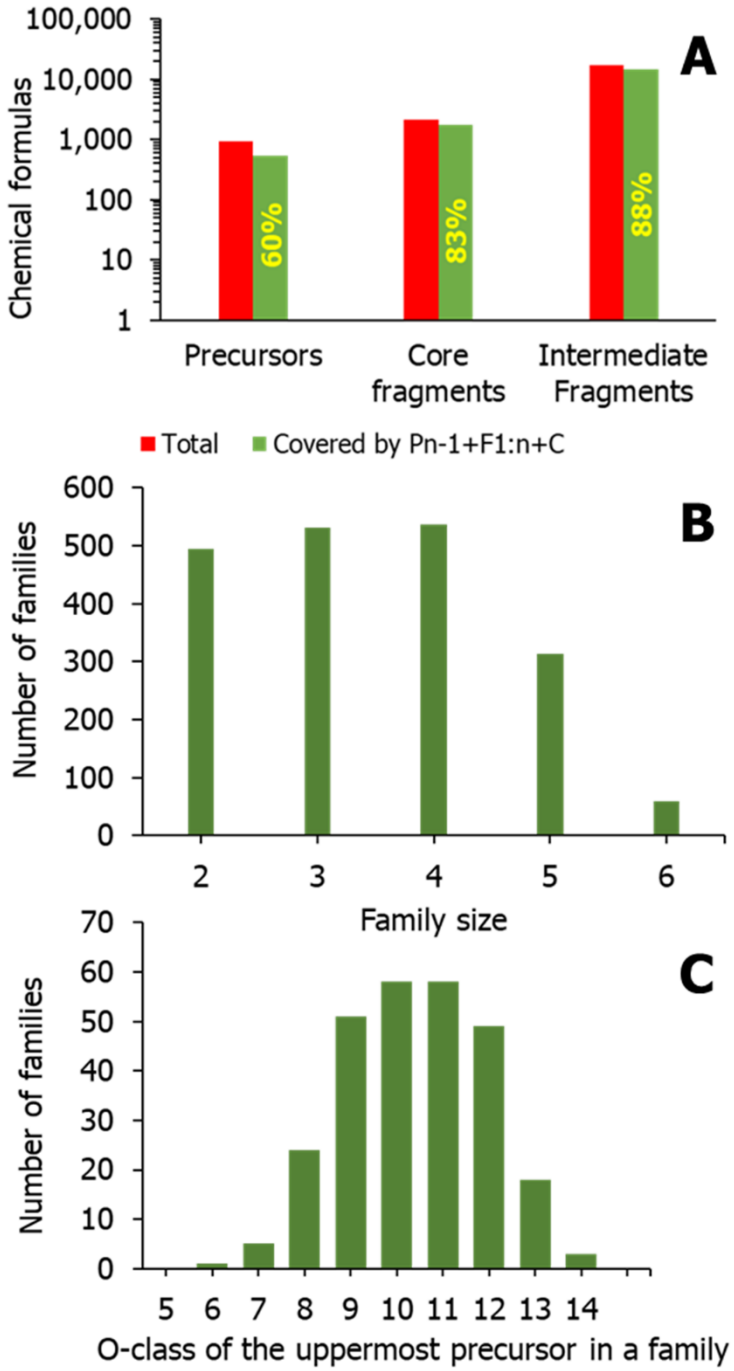


Figure 3. Conceptual models designed to compute ordered fragmentation pathways (panel A) and find structural families in DOM based on sequential matching of fragmentation pathways (Panel B). Note that for the precursor P1 to be considered in a family, its ion mass should match (1 mDa tolerance) the mass of the first fragment in P_2's fragmentation pathway.

The number of families per oxygen class of the uppermost precursor within a family depicts a gaussian-type distribution centered in the O-class 10 (Figure 4C). This pattern is in good agreement with the distribution of pathways and core fragments per O-class found for the CHO compound

class (Figure S2). Nevertheless, a closer view of the Figure 4C evidenced a shift of the distribution towards less oxygenated family parents and an increase in the number of these uppermost precursors with 8-9 oxygens.



354

355 Figure 4. Number of covered precursors, core and intermediate fragments by the model P_{n-}
356 $_{1}$+$F_{1:n}$+C (A), distribution of the number of families per family size (B) and families per oxygen
357 class of the uppermost precursor (compound with the highest oxygen content in the family) (C)
358 respectively for the CHO class.

359 Overall, the families retrieved from the FT-ICR CASI-CID MS/MS data collected in the studied
360 mass range showed that the structural transformation of CHO components in DOM depends on
361 oxygenation/deoxygenation processes driven by both single and mixed addition or subtraction of
362 H_2O, CH_4O and CO_2 chemical units (Figure S3). This finding suggests that the structural alteration
363 of DOM involves complex mechanisms compared to the uniform trends (e.g. hydration and
364 carboxylation) previously observed from broadband FT-MS data[36,45].Although our findings are
365 constrained to O-compounds negatively ionized, the proposed approach allows the structural
366 analysis of other molecular classes (e.g., CHOS and CHON) upon availability of substantial
367 fragmentation data.

368 A closer view at the compositional relationship among members within a family revealed that
369 oxygenation (increase in O/C ratio) through the addition of carboxylic (CO_2) moieties, increase de
370 unsaturation degree of the resulting species (+1 DBE). Conversely, hydroxyl (H_2O) and methoxy
371 (CH_4O) additions are accompanied by a decrease in one DBE unit of the subsequent molecule.

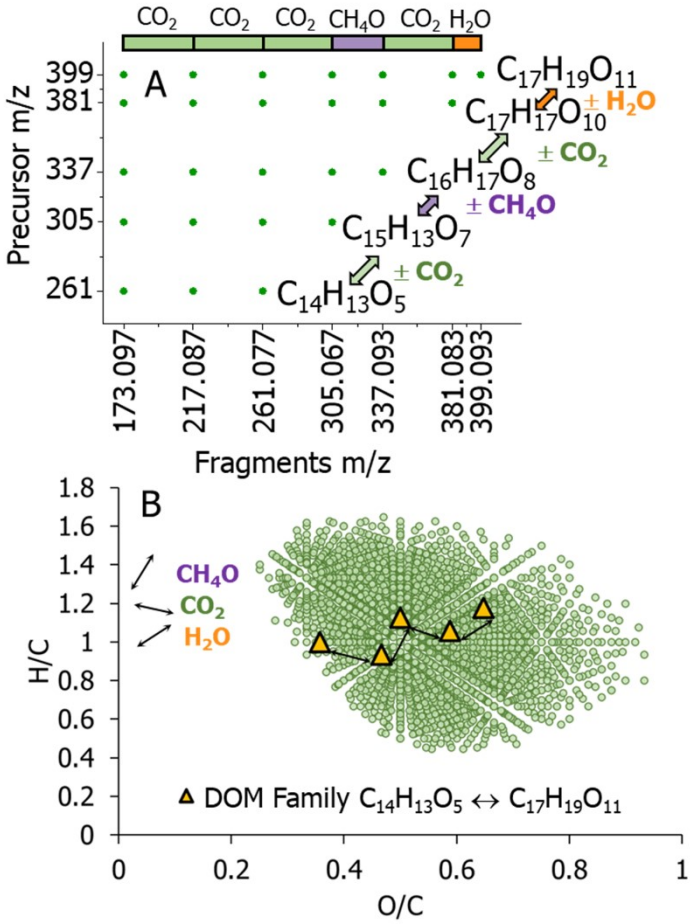


Figure 5. 2D MS/MS visualization of a characteristic DOM family of 6 precursor (Panel A). Chemical unities (H_2O, CH_4O and CO_2) differences among precursors are shown using a color code. Fragmentation pathways described as neutral losses are also shown as colored bars. Van Krevelen plot (B) of the CHO class compounds obtained from the MS_1 experiment highlighting the compositional nature of the structural family.

A 2D MS/MS representation of the molecular transformations exhibited by the structural family [$C_{14}H_{13}O_5$-$C_{15}H_{13}O_7$-$C_{16}H_{17}O_8$-$C_{17}H_{17}O_{10}$-$C_{17}H_{19}O_{11}$] identified in the SPE-DOM sample is shown in Figure 5A. The double arrows placed between family members indicate that potential biogeochemical transformations of DOM can be viewed from a bidirectional perspective. For instance, a sequential addition (synthesis-like) of carboxylic, hydroxyl and methoxy moieties (-

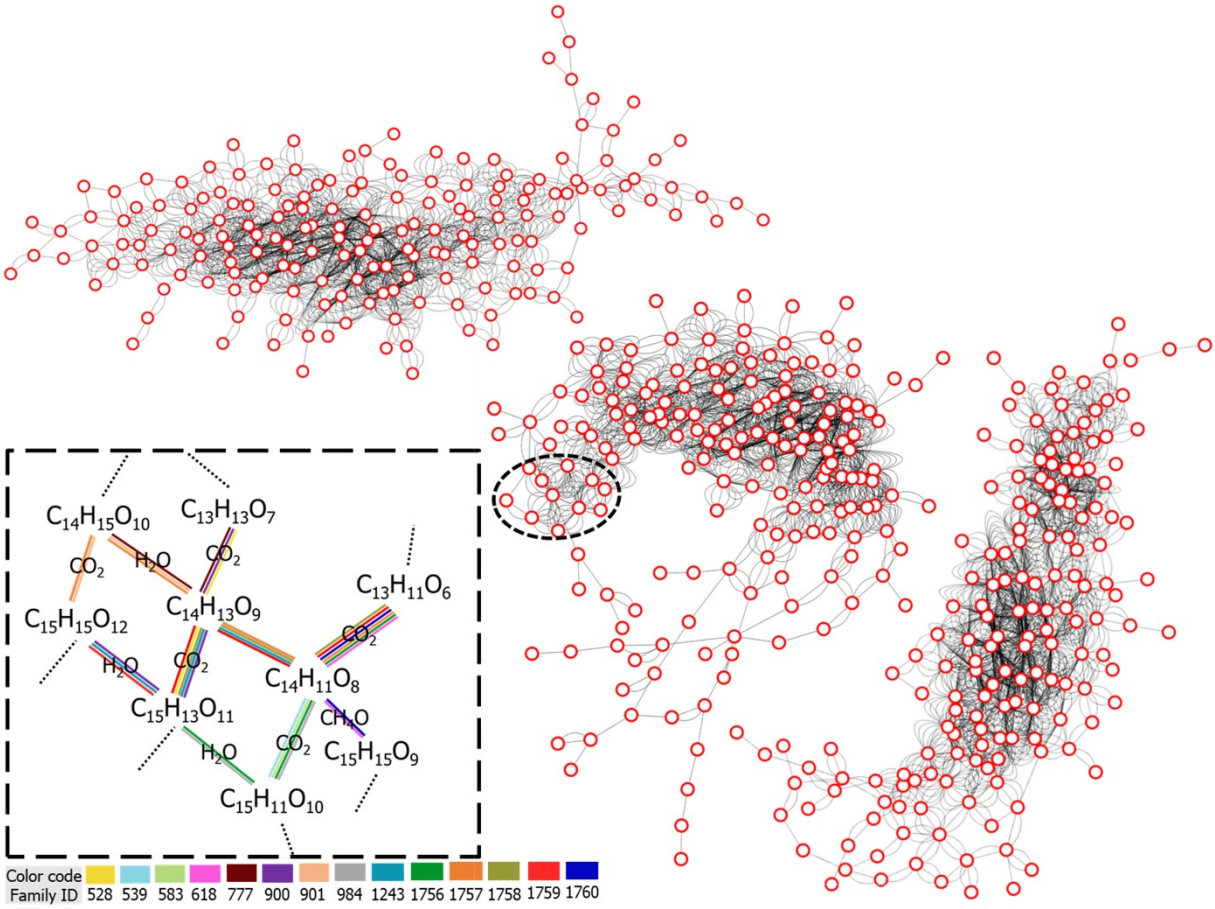
383 CO_2, -CH_4O, -CO_2, and -OH) starting from $C_{14}H_{13}O_5$ up to the family parent $C_{17}H_{19}O_{111}$ is
384 described in Figure 5A. This successive functionalization of O-depleted low molecular weight
385 compounds resulting in high molecular weight O-rich molecules, could be explained as aging
386 processes. Evidence of an increase in oxidized species observed in relatively old DOM from deep
387 ocean water[19,23] compared to younger freshwater DOM has been previously reported. Similar
388 findings of fresh (^{14}C dating) DOM exhibiting less oxygenated and lighter molecules compared to
389 older terrestrial DOM species have been also reported by Benk et.al.[72] However, the consistent
390 decrease in unsaturation of oxygenated high molecular weight DOM components observed in the
391 previous study, contrasts with our results of alternating unsaturation patterns along an ascending-
392 order structural family (e.g. DBE change 8-7-8-7-8 from $C_{14}H_{13}O_5$ to $C_{17}H_{19}O_{11}$). A notable
393 increase in O-rich molecules at the expense of the consumption of poor oxygenated species was
394 also reported in biodegradation experiments conducted on DOM from landfill leachate[73] and from
395 the surface of glaciers and ice sheets.[74] Although the impact of biodegradation on DOM structural
396 transformation was not investigated in this study, the increasing oxygenation trend reported in both
397 contributions, is in good agreement with the O-based functionalization found in our structural
398 families. Other abiotic processes such as photo or chemical oxidation have been also indicated as
399 responsible for the presence of highly oxygenated and CRAM species in DOM[19,75,76], yet
400 supporting our findings observed along a structural family in ascending order.

401 The analysis of the structural families in the reverse direction (top to bottom) suggests that DOM
402 molecules could also undergo mineralization-like transformations resulting in low molecular
403 weight reduced species. For instance, the oxygen-rich family parent $C_{17}H_{19}O_{11}$ (Figure 5A)
404 experiences de-functionalization processes characterized by consecutive eliminations of H2O,
405 CO_2, CH_4O, and CO_2, resulting in the poorly oxygenated low molecular weight compound

406 $C_{14}H_{13}O_5$. Interestingly, it has been suggested that highly-oxygenated compounds[77] and aromatic
407 oxidized species[78] from terrestrial DOM, determined by broadband FT-ICR-MS, are preferentially
408 removed by biodegradation, resulting in low molecular weight components. Similarly, a
409 significant decrease in aromatic content and oxygen functionalities was observed by Ward et. al[76]
410 during photodegradation experiments of soil DOM compared to dark controls. Moreover, Hawkes
411 et.al[79] have found that hydrothermal environments such as the ones observed in ocean deep
412 hydrothermal vents, could induce potential de-functionalization processes (e.g., decarboxylation
413 and dehydration) of O-rich high molecular weight DOM species, resulting in less O-functionalized
414 low molecular weight components. The results described in Figure 5 illustrate that our model
415 provides useful information that could help to elucidate the complex DOM transformational
416 mechanisms at the structural level.

417 A representation of the DOM structural family (Figure 5A) superimposed on the van Krevelen
418 space generated for the sample's CHO class is shown in Figure 5B. The discontinuous line patterns
419 described by different directional vectors representing neutral loss-based functionalities contrast
420 with the traditional straight lines utilized in the van Krevelen plot to describe chemical
421 transformations and reaction pathways of DOM components deduced from elemental composition
422 obtained from UHRMS data (Figure 1B).[1,36] Therefore, our results suggest that DOM
423 biogeochemical transformation mechanisms are more complex than traditionally described, based
424 upon the heterogeneous nature of the structural information obtained from neutral mass loss
425 patterns observed in this study. For example, DOM molecules assigned from an MS_1 analysis
426 describing a regular addition/subtraction of H_2O chemical units are conventionally interpreted as
427 a family characterized by a hydration/condensation process. Similarly, chemical formulas differing
428 in exactly CO_2 have been also placed into a homologous series resulting from

429 carboxylation/decarboxylation pathways[49]. However, our findings indicate that CHO compounds
430 in this DOM sample form more complex families characterized by multiple heterogenous
431 combinations of neutral loss based structural moieties (e.g., H_2O, CH_4O and CO_2) such as the one
432 described in Figure 5. These results illustrate that the integration of efficient computational tools
433 with comprehensive UHRMS fragmentation workflows allows the identification of valuable
434 structural information of DOM components, that cannot be accurately predicted by traditional FT-
435 MS workflows.
436



437

438 Figure 6. View of the three main clusters observed in the network of neutral-loss based structurally
439 connected DOM precursors for the CHO class. Precursor molecules are described by nodes and

440 the family indexes are shown as edges. An expanded view of fourteen interconnected DOM
441 families is shown as inset. A comprehensive web-based network can be found at
442 https://github.com/Usman095/Graph-DOM.

443 The visualization of the computed families using Cytoscape confirms the notion that DOM forms
444 a complex assembly of interconnected molecules (Figure 6). Similar results using broadband FT-
445 ICR MS data of DOM from both surface and deep sea[26] water samples and from secondary organic
446 aerosols[80] have been reported.

447 A closer view at the structural network in Figure 6 revealed three main clusters of related DOM
448 components (red dots) connected by neutral loss-based structural functionalities (edges). A more
449 detailed analysis of a specific region of the network described in the inset of Figure 6, illustrates
450 that several precursors are common to multiple families. This result, not previously observed at
451 the precursor level, shows the crucial role that the structural isomers play in the interconnection of
452 DOM compounds and confirms that isomeric diversity is a fundamental component of DOM
453 molecular complexity. The level of complexity observed in this network suggests that previous
454 elemental-based composition interpretations cannot accurately describe structural patterns in
455 DOM.

456 In this model the intersection of structural families relates to the isomeric content of DOM.
457 However, it should be note that the model may overestimate the number of fragmentation pathways
458 due to the nominal mass CASI CID data collection. The analysis of the fragmentation pathways
459 determined by nominal mass and chemical formula-based MS/MS for the case of the 267.087412
460 *m/z* ion ($C_{13}H_{15}O_6$) showed that all nine fragmentation pathways determined by chemical formula-
461 based MS/MS are also observed in the nominal mass analysis (Table S1). This is an expected result
462 and speaks to the effective processing of the computational code. The nominal mass MS/MS

463 processing resulted in thirteen additional fragmentation channels. While some of the additional
464 fragmentation channels (overestimation) can be derived from differences in the fragmentation
465 mechanism (CASI CID vs SORI CID), the application of the model to nominal mass CASI CID
466 MS/MS will inherently carry potential overestimations.

467 The analytical power of this workflow is based on the fast acquisition of nominal mass CASI-
468 CID datasets from complex DOM samples. The model applied to nominal mass CASI CID MS/MS
469 effectively reports all the "real" fragmentation pathways. One alternative to reduce the workflow
470 overestimation is to utilize chemical formula-based MS/MS, but this approach is unpractical for
471 routine DOM analysis. A more viable alternative is the implementation of complementary
472 artificial intelligence and machine learning approaches trained with small subsets of chemical
473 formula-based MS/MS data from DOM samples.

474

475 ASSOCIATED CONTENT

476 **Supporting Information**

477 Figure S1 shows MS/MS data points (S/N> 3) per nominal *m/z* (top) and number of precursor
478 chemical formulas per nominal *m/z* for the assigned heteroatom classes (bottom). Figure S2 depicts
479 the distribution of number of fragmentation pathways and core fragments per assigned precursor
480 and per oxygen class of the precursor, for the CHO class. Figure S3 displays the distribution of the
481 number of structural families of CHO compounds per unique neutral loss sequence found. Table
482 S1 summarizes a comparison of the MS/MS data and fragmentation pathways obtained from ESI-
483 FT-ICR CASI-CID and ESI-FT-ICR CHEF-SORI-CID. Graph-DOM code along with the input
484 file and a web-based Cytoscape network of DOM structural families are available at

https://github.com/Usman095/Graph-DOM. The MS1 and CASI-CID raw data of the SPE-DOM sample is freely accessible at https://doi.org/10.34703/gzx1-9v95/SIXONK.

AUTHOR INFORMATION

Corresponding Author

*** Ph.: 305-348-2037. Fax: 305-348-3772.**

E-mail: fernandf@fiu.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation Division of Chemistry, under CAREER award CHE-1654274, with co-funding from the Division of Molecular and Cellular Biosciences to FFL. DL acknowledges the fellowship provided by the National Science Foundation award (HRD-1547798) to Florida International University as part of the Centers for Research Excellence in Science and Technology (CREST) Program. This is contribution number xxx from the Southeast Environmental Research Center in the Institute of Water & Environment at Florida International University The authors would like to acknowledge for their technical

support to the personnel of the Advance Mass Spectrometry Facility at Florida International University, and Dr. Jeremy Wolf.

REFERENCES

1. Dittmar, T.; Stubbins, A. 12.6 - Dissolved Organic Matter in Aquatic Systems. In *Treatise on Geochemistry (Second Edition)*, Holland, H. D.; Turekian, K. K., Eds.; Elsevier: Oxford, 2014; pp 125-156.
2. Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. *Biogeosciences* **2016,** *13* (8), 2257-2277.
3. Lu, K.; Liu, Z. Molecular Level Analysis Reveals Changes in Chemical Composition of Dissolved Organic Matter From South Texas Rivers After High Flow Events. *Frontiers in Marine Science* **2019,** *6* (673), 1-18.
4. Kurek, M. R.; Poulin, B. A.; McKenna, A. M.; Spencer, R. G. M. Deciphering Dissolved Organic Matter: Ionization, Dopant, and Fragmentation Insights via Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Environ. Sci. Technol.* **2020,** *54* (24), 16249-16259.
5. Xu, W.; Gao, Q.; He, C.; Shi, Q.; Hou, Z.-Q.; Zhao, H.-Z. Using ESI FT-ICR MS to Characterize Dissolved Organic Matter in Salt Lakes with Different Salinity. *Environ. Sci. Technol.* **2020,** *54* (20), 12929-12937.
6. Pan, Q.; Zhuo, X.; He, C.; Zhang, Y.; Shi, Q. Validation and Evaluation of High-Resolution Orbitrap Mass Spectrometry on Molecular Characterization of Dissolved Organic Matter. *ACS Omega* **2020,** *5* (10), 5372-5379.
7. Cooper, W. T.; Chanton, J. C.; D'Andrilli, J.; Hodgkins, S. B.; Podgorski, D. C.; Stenson, A. C.; Tfaily, M. M.; Wilson, R. M. A History of Molecular Level Analysis of Natural Organic Matter by FTICR Mass Spectrometry and The Paradigm Shift in Organic Geochemistry. *Mass Spectrom. Rev.* **2020,** *n/a* (n/a).
8. Zark, M.; Christoffers, J.; Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar. Chem.* **2017,** *191*, 9-15.
9. Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science* **2017,** *4* (405), 1-14.
10. Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J. R.; Tranvik, L. J.; Bergquist, J. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnol. Oceanogr. Lett.* **2018,** *3* (2), 21-30.
11. Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the structural complexity of dissolved organic matter: isomeric diversity. *Faraday Discuss.* **2019,** *218* (0), 431-440.
12. Leyva, D.; Jaffe, R.; Fernandez-Lima, F. Structural Characterization of Dissolved Organic Matter at the Chemical Formula Level Using TIMS-FT-ICR MS/MS. *Anal. Chem.* **2020,** *92* (17), 11960-11966.

13. Lam, B.; Baer, A.; Alaee, M.; Lefebvre, B.; Moser, A.; Williams, A.; Simpson, A. J. Major Structural Components in Freshwater Dissolved Organic Matter. *Environ. Sci. Technol.* **2007,** *41* (24), 8240-8247.
14. Woods, G. C.; Simpson, M. J.; Koerner, P. J.; Napoli, A.; Simpson, A. J. HILIC-NMR: Toward the Identification of Individual Molecular Components in Dissolved Organic Matter. *Environ. Sci. Technol.* **2011,** *45* (9), 3880-3886.
15. Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **2013,** *10* (3), 1583-1624.
16. Zhang, F.; Harir, M.; Moritz, F.; Zhang, J.; Witting, M.; Wu, Y.; Schmitt-Kopplin, P.; Fekete, A.; Gaspar, A.; Hertkorn, N. Molecular and structural characterization of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry. *Water Res.* **2014,** *57*, 280-294.
17. Hertkorn, N.; Harir, M.; Schmitt-Kopplin, P. Nontarget analysis of Murchison soluble organic matter by high-field NMR spectroscopy and FTICR mass spectrometry. *Magn. Reson. Chem.* **2015,** *53* (9), 754-768.
18. Simpson, A. J.; Kingery, W. L.; Hatcher, P. G. The Identification of Plant Derived Structures in Humic Materials Using Three-Dimensional NMR Spectroscopy. *Environ. Sci. Technol.* **2003,** *37* (2), 337-342.
19. Reemtsma, T.; These, A.; Linscheid, M.; Leenheer, J.; Spitzy, A. Molecular and Structural Characterization of Dissolved Organic Matter from the Deep Ocean by FTICR-MS, Including Hydrophilic Nitrogenous Organic Molecules. *Environ. Sci. Technol.* **2008,** *42* (5), 1430-1437.
20. Cortés-Francisco, N.; Caixach, J. Fragmentation studies for the structural characterization of marine dissolved organic matter. *Anal. Bioanal. Chem.* **2015,** *407* (9), 2455-2462.
21. Lu, K.; Gardner, W. S.; Liu, Z. Molecular Structure Characterization of Riverine and Coastal Dissolved Organic Matter with Ion Mobility Quadrupole Time-of-Flight LCMS (IM Q-TOF LCMS). *Environ. Sci. Technol.* **2018,** *52* (13), 7182-7191.
22. Zark, M.; Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **2018,** *9* (1), 3178.
23. Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **2006,** *70* (12), 2990-3010.
24. Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. Pushing the analytical limits: new insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. *Chemical Science* **2019,** *10* (29), 6966-6978.
25. Le Maître, J.; Hubert-Roux, M.; Paupy, B.; Marceau, S.; Rüger, C. P.; Afonso, C.; Giusti, P. Structural analysis of heavy oil fractions after hydrodenitrogenation by high-resolution tandem mass spectrometry and ion mobility spectrometry. *Faraday Discuss.* **2019,** *218* (0), 417-430.
26. Longnecker, K.; Kujawinski, E. B. Using network analysis to discern compositional patterns in ultrahigh-resolution mass spectrometry data of dissolved organic matter. *Rapid Commun. Mass Spectrom.* **2016,** *30* (22), 2388-2394.
27. Herzsprung, P.; Wentzky, V.; Kamjunke, N.; von Tümpling, W.; Wilske, C.; Friese, K.; Boehrer, B.; Reemtsma, T.; Rinke, K.; Lechtenfeld, O. J. Improved Understanding of Dissolved Organic Matter Processing in Freshwater Using Complementary Experimental and Machine Learning Approaches. *Environ. Sci. Technol.* **2020,** *54* (21), 13556-13565.

28. Vialykh, E. A.; McKay, G.; Rosario-Ortiz, F. L. Computational Assessment of the Three-Dimensional Configuration of Dissolved Organic Matter Chromophores and Influence on Absorption Spectra. *Environ. Sci. Technol.* **2020,** *54* (24), 15904-15913.
29. Devarajan, D.; Liang, L.; Gu, B.; Brooks, S. C.; Parks, J. M.; Smith, J. C. Molecular Dynamics Simulation of the Structures, Dynamics, and Aggregation of Dissolved Organic Matter. *Environ. Sci. Technol.* **2020,** *54* (21), 13527-13537.
30. Kujawinski, E. B. Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS): Characterization of Complex Environmental Mixtures. *Environmental Forensics* **2002,** *3* (3), 207-216.
31. Kujawinski, E. B.; Freitas, M. A.; Zang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. *Org. Geochem.* **2002,** *33* (3), 171-180.
32. Sleighter, R. L.; Hatcher, P. G. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* **2007,** *42* (5), 559-574.
33. Sleighter, R. L.; Hatcher, P. In *Fourier Transform Mass Spectrometry for the Molecular Level Characterization of Natural Organic Matter: Instrument Capabilities, Applications, and Limitations*, 2011.
34. Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Anal. Bioanal. Chem.* **2007,** *389* (5), 1311-1327.
35. Smith, D. F.; Podgorski, D. C.; Rodgers, R. P.; Blakney, G. T.; Hendrickson, C. L. 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures. *Anal. Chem.* **2018,** *90* (3), 2041-2047.
36. Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal. Chem.* **2003,** *75* (20), 5336-5344.
37. Wu, Z.; Rodgers, R. P.; Marshall, A. G. Two- and Three-Dimensional van Krevelen Diagrams: A Graphical Analysis Complementary to the Kendrick Mass Plot for Sorting Elemental Compositions of Complex Organic Mixtures Based on Ultrahigh-Resolution Broadband Fourier Transform Ion Cyclotron Resonance Mass Measurements. *Anal. Chem.* **2004,** *76* (9), 2511-2516.
38. Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol. Oceanogr.* **2010,** *55* (4), 1467-1477.
39. D'Andrilli, J.; Cooper, W. T.; Foreman, C. M.; Marshall, A. G. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Commun. Mass Spectrom.* **2015,** *29* (24), 2385-2401.
40. Nebbioso, A.; Piccolo, A. Molecular characterization of dissolved organic matter (DOM): a critical review. *Anal. Bioanal. Chem.* **2013,** *405* (1), 109-124.
41. Kendrick, E. A Mass Scale Based on CH2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **1963,** *35* (13), 2146-2154.
42. Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001,** *73* (19), 4676-4681.

43. Roth, V.-N.; Dittmar, T.; Gaupp, R.; Gleixner, G. The Molecular Composition of Dissolved Organic Matter in Forest Soils as a Function of pH and Temperature. *PLOS ONE* **2015,** *10* (3), e0119188.
44. Koch, B. P.; Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **2006,** *20* (5), 926-932.
45. Stenson, A. C.; Landing, W. M.; Marshall, A. G.; Cooper, W. T. Ionization and Fragmentation of Humic Substances in Electrospray Ionization Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2002,** *74* (17), 4397-4409.
46. Witt, M.; Fuchser, J.; Koch, B. P. Fragmentation Studies of Fulvic Acids Using Collision Induced Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2009,** *81* (7), 2688-2694.
47. LeClair, J. P.; Collett, J. L.; Mazzoleni, L. R. Fragmentation Analysis of Water-Soluble Atmospheric Organic Matter Using Ultrahigh-Resolution FT-ICR Mass Spectrometry. *Environ. Sci. Technol.* **2012,** *46* (8), 4312-4322.
48. These, A.; Winkler, M.; Thomas, C.; Reemtsma, T. Determination of molecular formulas and structural regularities of low molecular weight fulvic acids by size-exclusion chromatography with electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004,** *18* (16), 1777-1786.
49. Stenson, A. C.; Marshall, A. G.; Cooper, W. T. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Anal. Chem.* **2003,** *75* (6), 1275-1284.
50. Doerr, A. DIA mass spectrometry. *Nat. Methods* **2015,** *12* (1), 35-35.
51. Pegg, C. L.; Phung, T. K.; Caboche, C. H.; Niamsuphap, S.; Bern, M.; Howell, K.; Schulz, B. L. Quantitative data independent acquisition glycoproteomics of sparkling wine. *Molecular & Cellular Proteomics* **2020**, mcp.RA120.002181.
52. Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020,** *20* (17-18), 1900276.
53. Ross, C. W.; Simonsick, W. J.; Aaserud, D. J. Application of Stored Waveform Ion Modulation 2D-FTICR MS/MS to the Analysis of Complex Mixtures. *Anal. Chem.* **2002,** *74* (18), 4625-4633.
54. van Agthoven, M. A.; Delsuc, M.-A.; Bodenhausen, G.; Rolando, C. Towards analytically useful two-dimensional Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Bioanal. Chem.* **2013,** *405* (1), 51-61.
55. Bray, F.; Bouclon, J.; Chiron, L.; Witt, M.; Delsuc, M.-A.; Rolando, C. Nonuniform Sampling Acquisition of Two-Dimensional Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Increased Mass Resolution of Tandem Mass Spectrometry Precursor Ions. *Anal. Chem.* **2017,** *89* (17), 8589-8593.
56. Pfändler, P.; Bodenhausen, G.; Rapin, J.; Houriet, R.; Gäumann, T. Two-dimensional fourier transform ion cyclotron resonance mass spectrometry. *Chem. Phys. Lett.* **1987,** *138* (2), 195-200.
57. Ross, C. W.; Guan, S.; Grosshans, P. B.; Ricca, T. L.; Marshall, A. G. Two-dimensional Fourier transform ion cyclotron resonance mass spectrometry/mass spectrometry with stored-waveform ion radius modulation. *J. Am. Chem. Soc.* **1993,** *115* (17), 7854-7861.

58. van Agthoven, M. A.; Coutouly, M.-A.; Rolando, C.; Delsuc, M.-A. Two-dimensional Fourier transform ion cyclotron resonance mass spectrometry: reduction of scintillation noise using Cadzow data processing. *Rapid Commun. Mass Spectrom.* **2011,** *25* (11), 1609-1616.
59. Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D. H.; Marshall, A. G. External accumulation of ions for enhanced electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **1997,** *8* (9), 970-976.
60. Lakshmanan, R.; Wolff, J. J.; Alvarado, R.; Loo, J. A. Top-down protein identification of proteasome proteins with nanoLC-FT-ICR-MS employing data-independent fragmentation methods. *Proteomics* **2014,** *14* (10), 1271-1282.
61. Wenke, J. L.; Rose, K. L.; Spraggins, J. M.; Schey, K. L. MALDI Imaging Mass Spectrometry Spatially Maps Age-Related Deamidation and Truncation of Human Lens Aquaporin-0. *Investigative Ophthalmology & Visual Science* **2015,** *56* (12), 7398-7405.
62. Sleighter, R. L.; McKee, G. A.; Hatcher, P. G. Direct Fourier transform mass spectral analysis of natural waters with low dissolved organic matter. *Org. Geochem.* **2009,** *40* (1), 119-125.
63. Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **2008,** *6* (6), 230-235.
64. Heck, A. J. R.; de Koning, L. J.; Pinkse, F. A.; Nibbering, N. M. M. Mass-specific selection of ions in Fourier-transform ion cyclotron resonance mass spectrometry. Unintentional off-resonance cyclotron excitation of selected ions. *Rapid Commun. Mass Spectrom.* **1991,** *5* (9), 406-414.
65. de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H. Mass selection of ions in a Fourier transform ion cyclotron resonance trap using correlated harmonic excitation fields (CHEF). *Int. J. Mass Spectrom. Ion Processes* **1997,** *165-166*, 209-219.
66. Herzsprung, P.; Hertkorn, N.; von Tümpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Anal. Bioanal. Chem.* **2014,** *406* (30), 7977-7987.
67. Simon C, P. D., Roth V-N, Dührkop K, Böcker S, Dorrestein P, et al. Mass difference matching crystallizes hidden molecular structures of dissolved organic matter from ultrahigh-resolution tandem mass spectra. *ChemRxiv. Cambridge: Cambridge Open Engage* **2021**, 1-26.
68. Osterholz, H.; Niggemann, J.; Giebel, H.-A.; Simon, M.; Dittmar, T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nat. Commun.* **2015,** *6* (1), 7422.
69. Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P.-L.; Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **2010,** *27* (3), 431-432.
70. Minor, E. C.; Swenson, M. M.; Mattson, B. M.; Oyler, A. R. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environmental Science: Processes & Impacts* **2014,** *16* (9), 2064-2079.
71. Capley, E. N.; Tipton, J. D.; Marshall, A. G.; Stenson, A. C. Chromatographic Reduction of Isobaric and Isomeric Complexity of Fulvic Acids To Enable Multistage Tandem Mass Spectral Characterization. *Anal. Chem.* **2010,** *82* (19), 8194-8202.
72. Benk, S. A.; Li, Y.; Roth, V.-N.; Gleixner, G. Lignin Dimers as Potential Markers for 14C-young Terrestrial Dissolved Organic Matter in the Critical Zone. *Frontiers in Earth Science* **2018,** *6* (168), 1-9.

73. Yuan, Z.; He, C.; Shi, Q.; Xu, C.; Li, Z.; Wang, C.; Zhao, H.; Ni, J. Molecular Insights into the Transformation of Dissolved Organic Matter in Landfill Leachate Concentrate during Biodegradation and Coagulation Processes Using ESI FT-ICR MS. *Environ. Sci. Technol.* **2017,** *51* (14), 8110-8118.
74. Antony, R.; Willoughby, A. S.; Grannas, A. M.; Catanzano, V.; Sleighter, R. L.; Thamban, M.; Hatcher, P. G.; Nair, S. Molecular Insights on Dissolved Organic Matter Transformation by Supraglacial Microbial Communities. *Environ. Sci. Technol.* **2017,** *51* (8), 4328-4337.
75. Ward, C. P.; Cory, R. M. Assessing the prevalence, products, and pathways of dissolved organic matter partial photo-oxidation in arctic surface waters. *Environmental Science: Processes & Impacts* **2020,** *22* (5), 1214-1223.
76. Ward, C. P.; Cory, R. M. Complete and Partial Photo-oxidation of Dissolved Organic Matter Draining Permafrost Soils. *Environ. Sci. Technol.* **2016,** *50* (7), 3545-3553.
77. Kim, S.; Kaplan, L. A.; Hatcher, P. G. Biodegradable dissolved organic matter in a temperate and a tropical stream determined from ultra-high resolution mass spectrometry. *Limnol. Oceanogr.* **2006,** *51* (2), 1054-1063.
78. Mostovaya, A.; Hawkes, J. A.; Dittmar, T.; Tranvik, L. J. Molecular Determinants of Dissolved Organic Matter Reactivity in Lake Water. *Frontiers in Earth Science* **2017,** *5* (106), 1-13.
79. Hawkes, J. A.; Hansen, C. T.; Goldhammer, T.; Bach, W.; Dittmar, T. Molecular alteration of marine dissolved organic matter under experimental hydrothermal conditions. *Geochim. Cosmochim. Acta* **2016,** *175*, 68-85.
80. Tziotis, D.; Hertkorn, N.; Schmitt-Kopplin, P. 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* (4), 415-421.

TOC Abstract