Mass difference matching unfolds hidden molecular structures of dissolved organic matter

Carsten Simon[1,†], Kai Dührkop[2], Daniel Petras[3,4], Vanessa-Nina Roth[1,§], Sebastian Böcker[2], Pieter C. Dorrestein[3], and Gerd Gleixner[1,*]

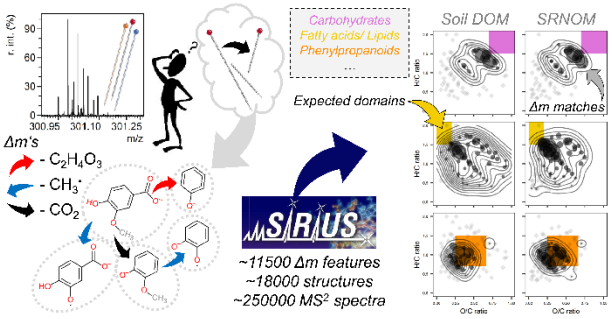
[1] Molecular Biogeochemistry, Department of Biogeochemical Processes, Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany

[2] Chair for Bioinformatics, Friedrich-Schiller-University, Jena, Germany

[3] Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, San Diego, CA, USA

[4] CMFI Cluster of Excellence, Interfaculty Institute of Microbiology and Medicine, University of Tübingen, 72076, Tübingen, Germany

TOC FIGURE:



ABSTRACT: Ultrahigh-resolution Fourier transform mass spectrometry (FTMS) has revealed unprecedented detail of natural complex mixtures such as dissolved organic matter (DOM) on a molecular formula level, but we lack approaches to access the underlying structural complexity. We here explore the hypothesis that every DOM precursor is potentially linked with all emerging product ions in $FTMS^2$ experiments. The resulting mass difference (Δm) matrix is deconvoluted to isolate individual precursor Δm profiles and matched with structural information, which was derived from 42 Δm features from 14 in-house reference compounds and a global set of 11477 Δm features with assigned structure specificities, using a dataset of ~18000 unique structures. We show that Δm matching is highly sensitive in predicting potential precursor identities in terms of molecular and structural composition. Additionally, the approach identified unresolved precursors and missing elements in molecular formula annotation (P, Cl, F). Our study provides first results how Δm matching improves structural domains in Van Krevelen space, but simultaneously demonstrates the wide overlap between the structural domains. We show that this effect is likely driven by chemodiversity and offers an explanation for the observed ubiquitous presence of molecules in the center

of the Van Krevelen space. Our promising first results suggest that Δm matching can unfold the structural information encrypted in DOM and assess the quality of FTMS-derived molecular formulas of complex mixtures in general.

Synopsis: We present an approach to deconvolute and explore the structural composition of co-fragmented mixtures of organic molecules in environmental media.

Keywords: Natural organic matter, NOM, DI-ESI-MS/MS, FTMS, Orbitrap, tandem mass spectrometry, MS/MS, deconvolution

1. INTRODUCTION

Complex mixtures are key study objects in environmental and industrial applications, but their analysis remains challenging.[1–4] One of the most complex mixtures in natural ecosystems is dissolved organic matter (DOM).[5,6] DOM is a central intermediate of ecosystem metabolism and mirrors molecular imprints of interactions with its abiotic and biotic environment[7–9], which form the basis for processes such as carbon sequestration and nutrient recycling.[10,11] Despite significant advances in ultrahigh-resolution mass spectrometry (FTMS)[2,4] and nuclear magnetic resonance spectroscopy[12], scientists still struggle to decode this information on the molecular level[13–17], and novel approaches to identify distinct structures are required to translate molecular-level information into improved process understanding.

Open and living systems promote the formation of ultra-complex mixtures of thousands to millions of individual constituents[18,19] that mirror large environmental gradients.[20–22]As a consequence, DOM poses significant challenges to separation, isolation, and structure elucidation. Direct infusion (DI) FTMS techniques have become indispensable tools for the molecular-level analysis of DOM as they reveal unprecedented detail of molecular formulas using the exact mass (MS^1 data, m/z) even without prior separation.[23] However, FTMS techniques are selective and do not resolve all structural detail observed at

51 the exact mass in DOM, as the presence of isobars and isomers hinders the identification of particular
52 structures from these molecular formulas.[19,23–25] Additionally, current structural databases cover only a small
53 fraction of molecular formulas encountered, and typically lead to annotation rates < 5%.[18,26,27]

54 One way to obtain structure information on isomers and isobars is through collision-induced dissociation
55 (CID) in fragmentation experiments (MS^2, or multistage MS^n).[27–29] The relatively wide isolation window (~
56 1 Da) of mass filters applied for precursor selection commonly hinders the isolation and subsequent
57 fragmentation of single exact masses, leading to mixed "chimeric" MS^2 spectra of co-fragmented
58 precursors.[30] Even though some authors have achieved isolation of single masses or improved description
59 of chimeric tandem MS data, most studies have pointed out that fragmentation patterns were rather universal
60 across DOM samples.[18,19,31–35] Most of these studies, however, focused on the major product ion peaks
61 (fragments), which usually make up only 60 – 70 % of the total product ion abundance, and thus disregard
62 many low-abundance signals that may be more suitable to detect structural differences.[19,31]

63 The major product ions encountered in tandem mass spectra of DOM relate to sequential neutral losses of
64 common small building blocks, mainly CO_2, H_2O, or CO units.[14,33] A mass difference between a precursor
65 and a product ion in an MS^2 spectrum is herein called "delta mass" and referred to as Δm (plural Δm's).
66 Many Δm's such as CO_2 or H_2O are commonly observed and are thus deemed non-indicative for the
67 identification of structural units.[18,28,31,33,36] In contrast, other studies found recurring low *m/z* product ions
68 (e.g., at *m/z* 95, 97, 109, 111, 123, 125, 137, 139, 151, and 153) that were interpreted as a limited set of core
69 structural units substituted with a set of functional groups, yet in different amounts and configurational types
70 that would lead to highly diverse mixtures.[37–44] From a stochastic standpoint, the occurrence of common
71 neutral losses may not be surprising; for example, many structures contain hydroxyl groups that could yield
72 H_2O losses, and CO_2 could originate from ubiquitous carboxyl groups.[45] In contrast, the occurrence of two
73 molecules sharing a larger substructure would be less probable, and thus less easily detected as a major
74 peak. Signatures of DOM's structural diversity could thus prevail in the high number of low-abundance
75 fragments usually detected below *m/z* 200-300, as opposed to the higher abundance of fragments connected

76 to losses of small substructures such as CO_2 or H_2O. Given the large number of estimated isomers and
77 isobars underlying usual DOM data[18,19,31,32,39,45–48], we here build upon the hypothesis that every co-
78 fragmented precursor potentially contributes to every emerging product ion signal. We interpret the resulting
79 chimeric MS^2 data as a structural fingerprint that can be deconvoluted to obtain individual precursor Δm
80 matching profiles. The analysis of Δm's that link precursor and product ions, in contrast to indicative product
81 ions (fragments) alone, is independent of the masses of the unknown precursors and known reference
82 compounds in databases of annotated Δm features. Although this approach will sacrifice the identification
83 of true knowns, it allows for the identification of potential structural analogs via indicative Δm's and is
84 suited best when annotation rates are as low as in the case of DOM, i.e., when most compounds are yet
85 unknown.[18,26,27]

86 Despite the unknown identity of most of the molecules present in DOM, its potential sources can be
87 constrained reasonably well. Plants produce most of the organic matter that sustains heterotrophic food webs
88 in natural ecosystems. Plant metabolites such as polyphenols and polyaromatic structures thus represent a
89 major source of DOM. Therefore, an early decomposition phase likely exists when the imprint of
90 soluble/solubilized plant metabolites is still detectable by MS^2 experiments using current FTMS technology.
91 For example, lignin-related compounds show indicative methoxyl and methyl radical losses[18,49,50];
92 glycosides indicate the loss of a sugar unit[51,52]and hydrolyzable tannins are expected to lose galloyl units.[52]
93 Even related compounds such as flavon-3-ols and flavan-3-ols could potentially be distinguished by their
94 indicative retro-cyclization products.[51,53] Mass differences related to atoms such as N, S, P, Cl, Br, I and F
95 could also help to identify unknown organic nutrient species or disinfection byproducts, thereby widening
96 the applicability of the approach.[1,54] Lastly, indicative Δm fingerprints could provide constraints to putative
97 compound group annotations derived from molecular formula data alone (Van Krevelen diagrams), or allow
98 for a more precise annotation.[55–57]

99 We hypothesized that DOM from swamps and topsoil, in close contact to plant inputs and active microbial
100 communities, would reflect recognizable plant-related source imprints that can be revealed by Orbitrap

101 tandem mass spectrometry. Specifically, we explored links between precursor Δm matching profiles and
102 precursor characteristics such as nominal mass, mass defect, initial ion abundance, fragmentation sensitivity,
103 oxygen-to hydrogen ratio (O/C), heteroatom content, and structure suggestions. These properties are in part
104 predictable from the assigned molecular formula, and thus allow for an evaluation of the approach ("proof-
105 of-concept") while also revealing potential non-assigned molecules of special interest (e.g., P-, Cl-, Br-, I-
106 and F-containing molecular formulas). Lastly, we hypothesized that indicative Δm features of plant phenols,
107 e.g., lignin- and tannin-related losses, would match their yet unknown structural analogs in DOM and that
108 these patterns would reflect commonly applied structural domain distributions.[56,58]

109 2. EXPERIMENTAL SECTION

110 A detailed experimental procedure is provided in the Supplemental Information of this article (**Note S-**
111 **1**). In short, we chose a set of 14 aromatic reference compounds as representative plant metabolites in DOM
112 (**Figure S-1**, **Table S-1**) and a forest topsoil pore water isolate[59] and Suwannee River Natural Organic
113 Matter (SRNOM)[60] as exemplary DOM samples. All reference and sample solutions were directly infused
114 into the ESI (electrospray) source of an Orbitrap Elite (Thermo Fisher Scientific, Bremen) at negative
115 ionization mode (**Table S-2**) and fragmented by collision-induced dissociation (CID, MS^2). We chose four
116 nominal masses within the average mass range typically observed in terrestrial DOM samples (m/z 200 –
117 500) for fragmentation (*m/z* 241, 301, 361, and 417, herein referred to as isolated precursor ion mixtures,
118 "IPIMs") as a first set of data to test the approach.[61] Soil DOM was analyzed at three normalized collision
119 energy (NCE) levels (15, 20, and 25%). MS^3 spectra of selected key product ions (aglycons of flavonoids
120 and demethylated dimethoxy-methyl-benzoquinone) were acquired as well at NCE 20 or 25. After
121 recalibration with known (**Table S-3**) or predicted product ions (losses of CO_2, H_2O, etc.), all major product
122 ions were annotated with a molecular formula in reference compounds (**Figure S-2**, **Table S-4**, **Table S-5**)
123 and DOM. Formula annotation was conducted with a Matlab routine recently incorporated into an open
124 FTMS data processing pipeline.[62]

125 For MS^2 data analysis, we generated *Δm matrices* of every pairwise combination of precursor and
126 product ions ("Δm fingerprints"). Every value in this matrix is referred to as a *Δm feature* or simply *Δm*. We
127 compared the unknown Δm features in DOM to three lists of known Δm features:

128 a) 54 Δm features ubiquitously found in DOM (**Table S-6**),

129 b) 55 Δm features from the set of 14 reference compounds (**Table S-7**), and

130 c) 11477 Δm features from a negative ESI MS^2 library with 249916 reference spectra of 17994 unique
131 molecular structures annotated by SIRIUS[63] (list available in the supporting datasets). Reference spectra
132 were collected from from GNPS, MassBank, MoNA, and NIST.[64,65]

133 The detection of a known Δm feature in DOM is herein called "*Δm matching*", and detected Δm features
134 are called *Δm matches*. Matching was conducted at a mass tolerance of ± 0.0002 Da (2 ppm at 200 Da). The
135 array of Δm matches of a single precursor is called the *Δm matching profile*, and all precursor profiles of an
136 IPIM form the subset of matched Δm's of the Δm matrix introduced above. The decomposition of the MS^2
137 spectrum into a Δm matrix and therefore, individual Δm matching profiles is what we define as the
138 *deconvolution* step in this study. Δm's of the literature- and reference-compound derived lists showed some
139 overlap and were largely part of the SIRIUS list as well (see details in SI). The specificity of any Δm feature
140 in the SIRIUS list was checked by their association to compound classes as defined by ClassyFire.[66] The
141 top 15 significantly associated classes were then obtained for each Δm feature in list c) and included into
142 analyses using the reference-compound derived list (list b) as well.

143 We assessed the probability of false positive matches and accounted for molecular formula constraints
144 (numbers of elements in the formula), ion abundance and measures of fragmentation sensitivity to validate
145 our approach. The matching data was combined for each NCE level and transformed into a binary format.
146 We classified Δm matching profiles of DOM precursors and reference compounds by two-way hierarchical
147 clustering using Ward's method and Euclidean distance, as well as Principal Components Analysis (PCA)
148 in PAST (v3.10).[67] We visualized numbers of individual Δm matches and Δm cluster matches in Van
149 Krevelen space to analyze patterns in Δm matching frequency ("structural domains"). We chose the
150 structural domains reprinted in the 2014 review by Minor et al. for reference, because this represents the

general level of detail and type of classes distinguished in recent DOM studies (**Figure S-3**).[58,68–70] In a separate analysis, lignin-like and N- and S-containing formulas were also classified with a more general Van Krevelen scheme besides the reference one.[71]

Finally, we assessed the agreement between structures predicted by Δm matching and those suggested in natural product structural databases. We combined structure suggestions from different databases, including Dictionary of Natural Products[72], KNApSAcK[73], Metacyc[74], KEGG[75], and HMDB[76] as well as their expanded in-silico annotations based on predicted enzymatic transformations in the MINEs database.[77] Although the MINEs database covers 198 generalized chemical reaction rules it may not include all potential environmental reactions because those are not necessarily only driven by enzymes. The InChi-Key of structures was used to exclude stereoisomers and classify suggested structures into compound classes by ClassyFire.[66]

3. RESULTS AND DISCUSSION

3.1. Tandem MS fragmentation of reference compounds and construction of Δm lists. The 14 aromatic reference compounds (**Figures S-1, S-2 and S-3**) yielded 42 new Δm features (i.e., not covered in the list of common Δm's, **Table S-6**) but also eight that were described in DOM. These eight Δm features (namely: H_2O, 18.0106; CO, 27.9949; C_2H_4, 28.0313; C_2H_2O, 42.0106; CO_2, 43.9898; CH_2O_3, 62.0004; C_2O_3, 71.9847; and C_3O_5, 115.9746) were kept in the list to compare DOM and reference compounds (**Table S-7**). Besides precursor formulas **#2** (Hydroxy-cinnamic acid, or p-coumaric acid; $C_9H_8O_3$, 164.0473), **#3** (Gallic acid; $C_7H_6O_5$, 170.0215) and **#5** (m-Guaiacol; $C_7H_8O_2$, 124.0524), which were found among the 42 Δm's as potential structural equivalents, five Δm's of potential substructures likely to be found in DOM were added to the list, namely the ones of precursors **#1** (Vanillic acid; $C_8H_8O_4$, 168.04226), **#4** (Creosol, $C_8H_{10}O_2$, 138.0681), **#8** (Ellagic acid; $C_{14}H_6O_8$, 302.0063) and **#10** (Catechin; $C_{15}H_{14}O_6$, 290.0790), and the neutral aglycon of compounds **#12** and **#13** (flavonol core of Spiraeoside and Isoquercitin; $C_{15}H_{10}O_7$, 302.0427). More details on reference compound fragmentation are given in the SI (**Note S-2**).

3.2. Fragmentation behavior of soil DOM. DOM precursors were isolated and fragmented to obtain Δm data (**Figure S-4**). To find the best collision energy to fragment DOM, we analyzed soil DOM at three

NCE levels (15, 20 and 25). All IPIMs showed similar fragmentation properties (**Note S-3**, **Table S-8**). Highest numbers of product ions were found at the highest NCE (**Figure S-5**). Product ion spectra did not indicate abrupt structural changes upon increasing fragmentation energy, showing no separation of isomers/isobars but a continuous increase in fragmentation across all precursors. Based on the above results, NCE of 25 was chosen to fragment SRNOM as a second DOM sample for comparison.

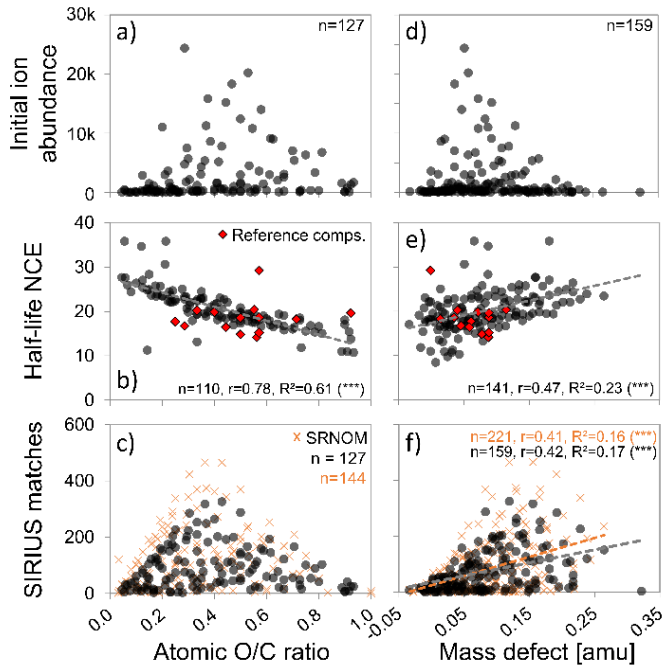


Figure 1. Links between selected DOM precursor properties (upper panels, initial ion abundance at NCE 0; mid panels, half-life normalized collision energy (NCE) at which ion abundance has dropped by 50%; lower panels, matches of delta masses (Δm's) of measured precursor and product ion masses (delta masses, Δm) with a list of 11477 known Δm features from SIRIUS) and each precursor's (**a**, **b**, **c**) O/C ratio or (**d**, **e**, **f**) mass defect. O/C ratios can only be shown for precursors with an annotated molecular formula. Additional data from reference compounds (red diamonds, see also **Figure S-3**) and SRNOM (orange crosses) is shown in mid and lower panels, respectively. Statistical data was derived from linear fits; asterisks (***) denote p-value < 0.001.

Despite common differences between precursor ion abundance and O/C ratio or mass defect (**Figure 1a**, **d**), we found a significant positive link between both metrics and fragmentation sensitivity independent of nominal mass, ranging from half-life NCE (i.e., the NCE level causing 50% decrease in ion abundance) of 10 – 35 under our instrumental settings (calculated from linear fits). Remarkably, this trend was not observed in reference compounds (**Figure 1b**, **e**). Such a discrepancy has been observed also by Zark et al. for the common CO_2 loss, and was interpreted as a result of intrinsic averaging.[31,45] In contrast, Dit Foque et al. described potential separation of less complex isomer mixtures by ramped fragmentation.[29] Bearing the

limitation in mind that we only analyzed four IPIMs here, our results support the intrinsic averaging hypothesis and indicate that fragmentation sensitivity may be an additional property shaped by DOM complexity.[18,20,45] It also supports our assumption of a high number of isomers and isobars "hidden" beneath each precursor molecular formula, which also increases the probability to detect meaningful links between precursor and product ions. A minor group of oxygen-poor formulas was non-responsive (**Note S-3**). Matching to the list of all SIRIUS Δm's showed no significant relation to O/C ratio but to mass defect (**Figure 1c**, **f**). In contrast to mass defect, initial ion abundance showed no link to fragmentation sensitivity but was significantly correlated to higher numbers of Δm matches ($r = 0.41$, $R^2 = 0.17$, $n = 157$, $p < 0.001$; see also **Tables S-9**, **S-10**, **S-11**, **S-12**, and **Figure S-6**). DOM precursors with an average O/C ratio matched more often than low O/C, fragmentation-resistant precursors (**Figure 1c**; **Figure S-7**, **Note S-3**)[18,19,35] or high O/C, easily fragmented precursors (**Figure 1b**). These observations together show that fragmentation sensitivity and Δm matching seem to be independent DOM precursor properties and that Δm matching could be driven by ion abundance. SRNOM and the soil water sample shared many molecular formulas (n=107; 84% of soil DOM and 74% of SRNOM formulas) which accounted for most of the precursor ion abundance at NCE 25 (96,5% and 97.2%, respectively). Despite this high similarity, SRNOM precursors showed higher numbers of Δm matches (**Figure 1c**, **f**) which could indicate that the same molecular formula is more chemodiverse, i.e. has more underlying structural formulas, in SRNOM compared to soil DOM (further discussion in **section 3.5**).

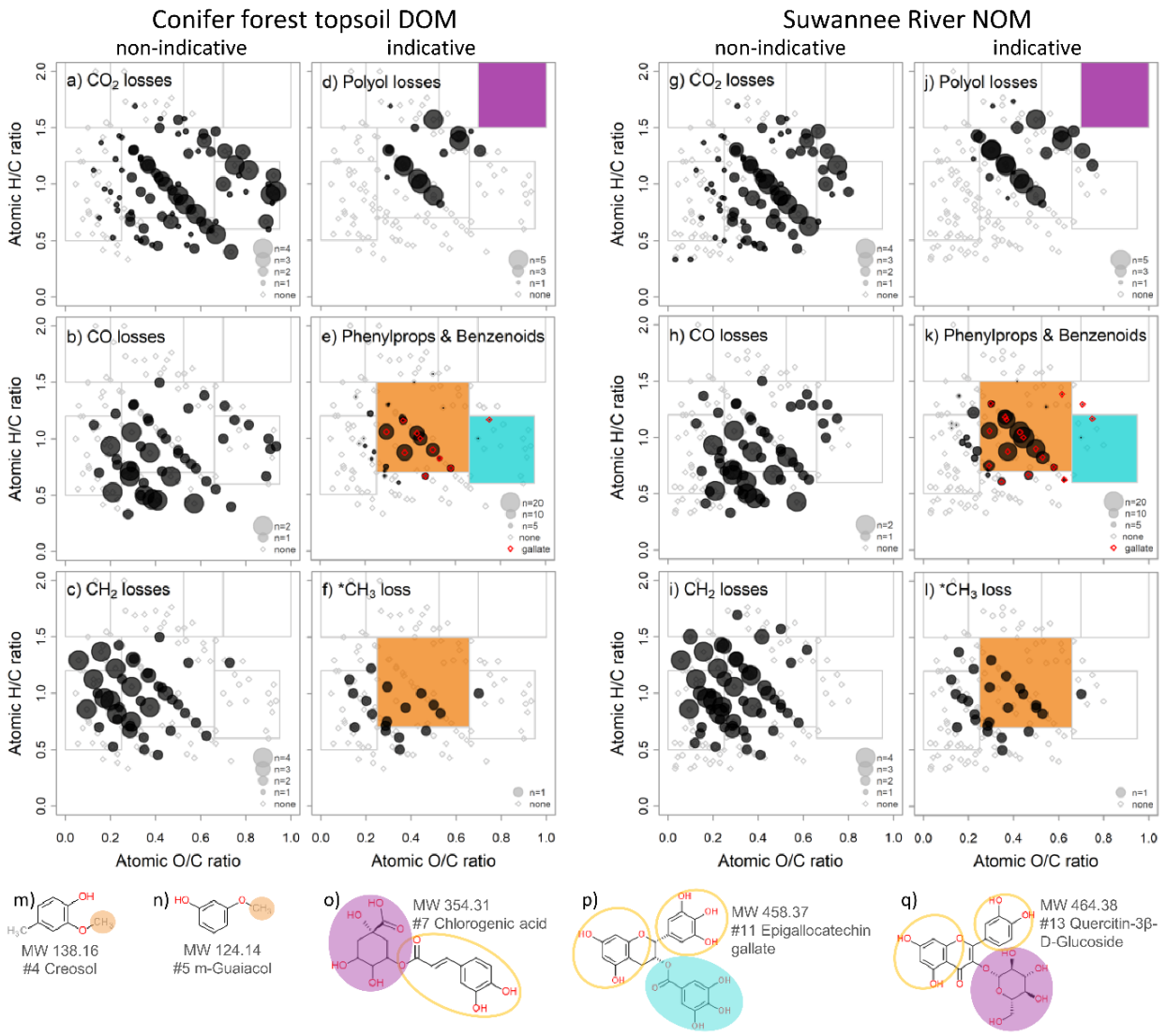


Figure 2. Δm matching in chemical space for soil (porewater) DOM (panels **a** – **f**) and SRNOM (panels **g** – **l**). Exemplary reference compound structures with marked indicative Δm units are shown in lower panels (**m** – **q**). Grey boxes refer to anticipated structural domains (**Figure S-3**).[64] Panels **a** – **l** show precursors with an annotated molecular formula by their atomic H/C and O/C ratios (Van Krevelen plot; soil DOM, n = 127; SRNOM, n = 144); grey boxes indicate representative structural domains that are commonly used (see **Figure S-3** for details). Dot size encodes the number of matches to non-indicative (**a** – **c**, **g** – **i**) vs. indicative Δm's (**d** – **f**, **j** - **l**); see legends in every plot. Colored boxes in indicative VK plots mark the expected structural region of formulas that would yield the respective Δm, and colors refer to the structural motifs marked in panels **m** - **q**. Calculations based on Δm data are presented in more detail

225 in **Table S-13**. Highlighted red open diamonds in panels **e** and **k** indicate loss of up to three gallic acid equivalents
226 (size not drawn to scale).

227 **3.3. Evaluation of the Δm matching approach.** We used the matching data of molecular formulas in
228 DOM for a proof-of-concept evaluation of our Δm matching approach. Specifically, we aimed to test the
229 hypothesis that all precursors are potentially linked to all product ions in chimeric MS^2 spectra of
230 ultracomplex DOM. Our analysis was congruent with previous observations, showing ubiquitous losses of
231 non-indicative oxygen-containing functionalities (**Table S-6**) while also revealing more detail (**Figure S-**
232 **4c**, **Table S-7**). Details are given in the Supporting Information (**Note S-4**); in short, we found expected
233 trends in losses of CO_2, CO, and CH_2 in both samples (**Figure 2a – c, g – i**, **Table S-13)**. The predicted
234 heteroatom content (O, N, S) of assigned molecular formulas and a widened tolerance window were used
235 for further analysis of the uncovered structural information. Random Δm matching would be expected if the
236 calculated Δm values were affected by low resolution, low sensitivity, or artifacts such as reactions in the
237 instrument (e.g., between the collision and Orbitrap cell[78]). Instead, we found that 1) precursors with low
238 ion abundance matched to less Δm features (**Figure S-6**), 2) non-fragmented precursors matched to less or
239 no Δm's (**Figure S-7**), and 3) identity of Δm matches agreed with molecular formula prediction (e.g., loss
240 of S-containing Δm's from S-containing precursors; ≤ 3 CO_2 losses from precursors containing seven O,
241 etc.; **Figures S-8** and **S-9**). Our evaluation also shows that Δm matching not only helps in recalibration[79]
242 but also serves to check formula annotation routines, as it revealed unresolved precursor compositions
243 interfering especially with CHOS precursors (related to Cl, P and F). This means 1) that these atoms should
244 be included for better coverage of elemental composition (i.e., prioritization) in our specific sample context
245 and that 2) higher resolution power may be required to resolve S-, Cl-, P-, or F-containing precursor
246 compositions.[1] In summary, Δm matching revealed an inherently structured biogeochemical signal of
247 precursors that seem to fragment individually and was highly sensitive in detecting precursor-product ion
248 links. This suggests that chimeric DOM data can be deconvoluted to reveal differences in molecular
249 composition not visible from MS^1 inspection.[23,80] It should be stressed that these results will need further
250 evaluation due to the small number of DOM precursors, m/z values and samples analyzed here (159 in soil

251 DOM, 221 in SRNOM), and that deconvolution should be further tested with better-characterized mixtures,
252 including, e.g., structural analogs, artificial mixtures or standard additions (spiking).[14,19,27,42,81]

253 **3.4. Structural domains emerge from clustering with reference compound Δm's.** DOM precursors
254 from both samples were compared based on Δm matching as an indicator of structural information (**Table**
255 **S-7**, see section **3.1**). We grouped DOM precursors, reference compounds and Δm features by two-way
256 hierarchical clustering (**Table S-14**). In the following, precursor clusters will be referred to by letter (A - H)
257 and Δm clusters by number (1 – 7; **Table S-15**). Based on the specificity of SIRIUS Δm features (**Table S-**
258 **14**) and clustering with 14 reference compounds we defined five of the Δm clusters found herein as being
259 structure-specific (**Table S-15**, some shown in **Figure 3d, e, j** and **k**; for details, see also **Table S-13**).

260 Δm features $C_4H_8O_4$ (120.0423 Da, equivalent to tetrose loss) and $C_6H_{10}O_5$ (162.0528 Da, equivalent to
261 hexose loss), both members of cluster 2, were found to be specific for alcohols and polyols, carbohydrates,
262 and carbohydrate conjugates, as well as ethers (**Table S-14**). Reference compounds containing a polyol
263 (quinic acid in #7) or a sugar (glucose in #12 and #13, mannose in #14) contributed Δm's to this cluster
264 (**Table S-15**).[51,52] Δm equivalents of these losses matched to 18 soil DOM and 24 SRNOM precursors in the
265 central Van Krevelen plot despite the absence of "carbohydrate-like" precursors (lilac square in **Figure 2d**,
266 **j** and **o**, **q**). The anticipated shift towards higher O/C and H/C ratios was nonetheless apparent in both
267 samples, especially compared to precursors associated with clusters 3, 4 and 7 (**Figure 2e**, **f** and **k**, **l**).

268 Δm features of clusters 3 and 4, partly specific to phenylpropanoid and benzenoid structures, were
269 contributed by flavan-3-ols (#10, #11) and aglycones of flavon-3-ols (#12 and #13) and those containing
270 cinnamic, coumaric or gallic acid units (#7, #9, #11).[28,33,52] Precursors that matched to clusters 3 and 4 (soil
271 DOM: n = 27 and n=12; SRNOM: n = 29, n = 21) were found in the "lignin-like" domain (orange square in
272 **Figure 2e**, **k**; orange circles in panels **o**, **p**, **q**). These C- or H-rich Δm's (e.g., $C_8H_{10}O_2$ or $C_7H_4O_4$) are likely
273 no combinations of common O-rich losses (CO, H_2O, or CO_2) due to their low O/C and O/H ratios, but this
274 requires further testing with model mixtures. Aliphatic chains prevail as O-poor substructures in substituted
275 cyclic core structures and could contribute.[82,83] Similar to detection of polyol-equivalent Δm matches outside
276 the expected carbohydrate domain, gallate-equivalent losses were not matched to precursors in the

anticipated "tannic" domain (red diamonds and turquoise square in **Figure 2e**, **k**; turquoise circle in panel **p**).

Among the most prominent features was the methyl radical loss[35,49,50] which matched to oxygen-poor DOM precursors and was one of three Δm features in cluster 7 (soil DOM: n = 18, average O/C = 0.33, SRNOM: n = 25, average O/C = 0.32, **Figure 2f**, **l**). The distribution of $CH_3^{\bullet}$-yielding precursors was paralleled by CH_2 (soil DOM: r = 0.60, R^2 = 0.35, n = 127, $p < 0.001$; SRNOM: r = 0.63, R^2 = 0.39, n = 144, $p < 0.001$) and CO losses (r = 0.55, R^2 = 0.30, n = 127, $p < 0.001$; r = 0.58, R^2 = 0.34, n = 144, $p < 0.001$) and implied structural similarities (**Figure 2f, l**), e.g., condensed structures with aliphatic, lactone, or quinone moieties.[34] CO and $CH_3^{\bullet}$ were both indicative of benzenoid structures in the SIRIUS-annotated Δm data (**Table S-14**). The methyl radical loss is an expected diagnostic Δm of methoxylated aromatic rings such as present in lignin (orange square in **Figure 2f, l**; orange circles in panels **m**, **n**; see **Note S-5**), but was also matched to DOM precursors not classified as "lignin-like".[18,31,35,49] The Δm features $CH_3^{\bullet}$, CO and C_2H_4 were also linked to CH_4 vs. O series. These describe a regularity in DOM data explained by increments of 0.0364 Da, and are formally annotated as an exchange of CH_4 for O (**Figure S-10**).[37,38] Concurrent losses of CO and C_2H_4 explained the presence of CH_4 vs. O increments on the product ion level and were paralleled by losses of $CH_3^{\bullet}$. This finding could also explain the ubiquitous presence of CH_4 vs O series in non-fragmented DOM; for example, concurrent β-oxidation and de-carbonylation could be enzymatic analogues of the patterns seen in MS^2 data.[26]

Matching to Δm features derived from a small set of reference compounds revealed emerging clusters of precursor and Δm feature families that may prove more indicative if constrained with further DOM and reference compound data.[14] Anticipated structural domains were apparent but showed clear overlap, which means that the same precursor was part of more than one Δm-predicted structural domain. An extended analysis using the set of compound class-associated SIRIUS Δm features showed similar trends (**Figure S-11**, compare **Figure 2**). These findings must however be taken with caution for four reasons:

1) SIRIUS Δm features were not obtained on the same instrument and thus may include features that, although correlated with certain compound classes, may not appear in DOM under the same instrumental settings.

2) SIRIUS Δm features may be biased towards certain classes of compounds, as is our reference set of 14 aromatic compounds. Here, we only considered negative ESI mode data which is commonly employed for DOM analysis, and thus, adding positive ESI or other ionizations would extend the range of Δm features and structural classes covered and likely decrease bias.[14,16,23,84] The same applies to other fragmentation techniques than CID.

3) Product ion abundance was disregarded in our analysis, but could be used to weigh probabilities of potential precursor-product ion links in future, potentially in combination with fragmentation energy gradients (fragmentation tree analysis)[85], moving m/z isolation windows, or variations in ion accumulation times that influence MS^1 ion abundance.[86]

4) Despite a seemingly improved separation of extreme classes (high H/C ratios in fatty acids, high O/C ratios in carbohydrates, etc.), potential overlap in domain boundaries remained considerable (**Figure S-11**).

Categorization of precursors into multiple Δm-defined structural domains was also reflected by large differences in Δm matching between members of the same a-priori defined structural domains or classes (i.e., only based on molecular formula). Twenty-seven precursors shown in **Figure 3** were classified as "lignin-like" formulas and were part of seven precursor clusters (B – H; **Table S-16**), thereby showing clear differences in potential structural composition. Likewise, CHNO and CHOS precursors matched with many of the S- and N-containing SIRIUS Δm features (spanning 3 – 78 S- and 4 – 251 N-containing Δm's in soil DOM and 0 – 154/ 0 – 350 in SRNOM; **Tables S-17**, **S-18**, **S-19** and **S-20**). These represented on average 79 ± 19% (63 ± 31% in SRNOM) of all Δm matches per CHOS precursor or 91 ± 7% (79 ± 28%) of all CHNO precursor matches (detailed analysis, see **Note S-6**). Many Δm features were also associated to compound classes, revealing potential structural detail (**Table S-21**). For example, CHNO precursor matching indicated the absence of nitrate esters, but indicated the presence of reduced forms of N partly

327 explained in the literature[87,88], including specific Δm's related to aralkylamines, amino acids,
328 carboximidamides, and dicarboximides/ urea-containing compounds. S-containing Δm matches indicated
329 the potential presence of sulfonic, thiol, thioether or aromatic S precursors.[84] Taken together, our results
330 show a wide potential diversity of N and S compounds in DOM that contradicts with earlier reports of
331 mainly aromatic N and sulfonic S.[34,89,90] As most of these studies analyzed marine DOM, the detection of
332 potentially more diverse sets of CHOS and CHNO precursors could relate to the terrestrial, less degraded
333 DOM analyzed here.[16,91–93] Further tests with N- and S-containing reference compounds and DOM samples
334 are warranted to reveal the hidden diversity and identity of dissolved organic nitrogen and sulfur and confirm
335 potential structures, e.g., by NMR.

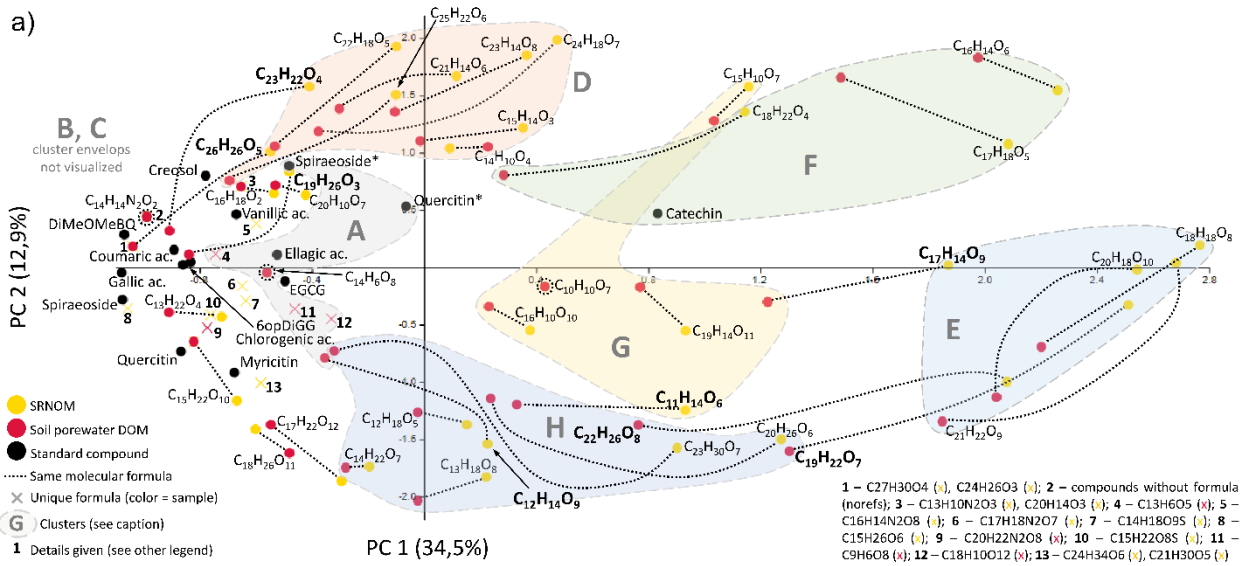
336 All in all, our results show that it may be possible to refine Van Krevelen domains by deconvoluted MS^2
337 data, and that complementary precursor information could be used to assess false or biased Δm-based class
338 assignments (e.g., elemental composition, DBE, ionization, fragmentation sensitivity, ion mobility, polarity
339 index, etc.).[13,56] Fluorescence or NMR spectroscopy could add valuable information if DOM would be
340 fractionated before MS^2 data acquisition.[21,94]

341 Data-dependent and data-independent acquisition (DDA, DIA) techniques could be used to cover the
342 whole mass range of precursors in DOM mass spectra in future, and are widely employed in LC-MS of
343 complex mixtures.[16,27,95,96] For example, Ludwig et al. presented a DIA scheme (SWATH-MS) that employs
344 one precursor scan and 32 isolation windows of 25 Da width, covering 800 Da within 3.3 seconds; similar
345 schemes are likely transferable to acquire full mass range data of directly-injected DOM.[97] Kurek et al.
346 recently presented such data, covering product ions generated from similar isolation window (*m/z* 392 –
347 408).[16] Smaller isolation windows as used herein were also employed by Leyva et al. to discern
348 fragmentation pathways and structural families in DOM (mass range *m/z* 261 – 477)[14]; this approach could
349 be extended to include the diversity of Δm features shown here. Together, this shows that practicable tandem
350 MS acquisition strategies are in reach and will enable deeper analyses of Δm features in DOM soon.

351 **3.5. Drivers of differences in Δm matching between soil DOM and SRNOM.** Although matching
352 among the two samples was largely consistent, slight differences were apparent from Van Krevelen

distributions (**Figure 2**). We therefore tested the separation of precursor clusters by ordination (**Figure 3**). Precursor clusters were clearly separated on PC1 and PC2 which together held about 47% of variation. Most considered precursors were shared among samples (64%, 38 out of 59), only a small number was sample-specific (SRNOM = 14, Soil DOM = 7). Sample-specific precursors were found in clusters A (linked to carboxylic acids), B (phenols, polyols) and C (benzenoids, **Table S-15**), the remaining clusters D – H were dominated by the shared precursors. Out of the 38 shared precursors, 30 (79%) grouped in the same precursor cluster despite a general trend to higher numbers of matches in SRNOM, but eight grouped differently (bold precursors in **Figure 3a**). These differences in matching could be related to different chemistries, i.e., different isomeric/ isobaric composition.[84] For example, based on the correlation of precursor properties (**Figure 3b**), the cluster "switch" in $C_{11}H_{14}O_6$ was largely explained by higher ion abundance and Δm matches in SRNOM, while in $C_{23}H_{22}O_4$, the effect was partly linked to higher fragmentation resistance in SRNOM. Unfortunately, we only have data on initial ion abundance and fragmentation sensitivity from the soil DOM isolate; other precursor properties, however, showed very similar trends in both samples (**Figure 3b**).

a)



b)

PC	Var. expl. [%]	Set	IPIM	Mass defect	H/C	O/C	DBE	AI_{MOD}	NOSC	Matches	Structures	Half-life NCE	Ion abund. (NCE0)
#1	34,5	All (n=94)	0.11	-0.11	-0.12	0.08	0.08	0.11	0.09	0.56***	0.56***	-	-
		SRNOM (n=50)	0.13	-0.18	-0.18	0.17	0.11	0.15	0.16	0.57***	0.54***	-	-
		Soil porewater DOM (n=44)	0.10	-0.06	-0.08	0.02	0.05	0.08	0.04	0.53***	0.59***	-0.20	0.70***
#2	12,9	All (n=94)	-0.11	-0.18	-0.57***	-0.52***	0.61***	0.67***	-0.02	-0.18	0.11	-	-
		SRNOM (n=50)	-0.07	-0.20	-0.63***	-0.55***	0.66***	0.72***	0.02	-0.19	0.07	-	-
		Soil porewater DOM (n=44)	-0.17	-0.19	-0.54***	-0.48**	0.55***	0.64***	-0.02	-0.27	0.17	0.53***	0.06
#3	8,8	All (n=94)	0.31**	0.75***	0.42***	-0.52***	0.02	-0.28**	-0.67***	0.59**	0.19	-	-
		SRNOM (n=50)	0.35*	0.72***	0.35*	-0.51***	0.07	-0.22	-0.65***	0.57***	0.26	-	-
		Soil porewater DOM (n=44)	0.27	0.77***	0.47**	-0.52***	-0.03	-0.34*	-0.68***	0.64***	0.22	0.47**	-0.41**
#4	6,2	All (n=94)	0.41***	-0.06	-0.17	0.19	0.25*	0.09	0.26*	0.21*	-0.29**	-	-
		SRNOM (n=50)	0.47***	-0.03	-0.16	0.18	0.27	0.08	0.24	0.27	-0.24	-	-
		Soil porewater DOM (n=44)	0.35*	-0.11	-0.18	0.22	0.23	0.10	0.30*	0.12	-0.35*	-0.16	-0.16

Figure 3. Separation of DOM precursors based on Δm matching. **a)** Principal Components analysis of all precursors with more than one match to indicative Δm features of the 14 reference compounds (i.e., Δm features shown in **Table S-7** that are not part of **Table S-6**, see section 3.1). Colors of dots distinguish precursors from both samples and reference compounds (see legend). Precursors detected in both samples are connected by dotted black lines. Precursor clusters (A – H) are marked by envelops and letters (compare **Tables S-14** and **S-15**). Eight shared precursors that switched precursor clusters are highlighted by bold molecular formula ($C_{12}H_{14}O_9$, A in soil DOM→H in SRNOM; $C_{19}H_{26}O_3$, B→C; $C_{26}H_{26}O_5$ and $C_{23}H_{22}O_4$, B→D; $C_{17}H_{14}O_9$, G→E; $C_{19}H_{22}O_7$ and $C_{22}H_{26}O_8$, H→E; $C_{11}H_{14}O_6$, H→G). **b)** Correlations of selected precursor properties with scores of PC axes (only DOM precursors with assigned molecular formula included in the correlation). PC axes 3 and 4 are shown in addition. Correlations are indicated for all precursors (n=94) and those detected in each sample (Column "Sets"). For each combination (PC = x, property = y), Pearson's r and significance are given ($0.05 \geq p > 0.01$, "*"; $0.01 \geq p > 0.001$, "**"; $p \leq 0.001$, "***"). Negative/ positive correlation is indicated also by color (blue, red); non-significant correlations are shown in lighter color or no color if no direction dominated. Matches, matches against the global list of Δm features; Structures, number of hits in natural product and in-silico databases.

Similar clustering and Δm-predicted structural classes (**Figure S-11**) in shared precursors could indicate a conserved structural composition. Likewise, Kurek et al. observed high similarity in APPI-ionized and IMPRD-fragmented DOM samples, but observed clear differences in CHOS fragmentation.[16] High

386 similarities between DOM samples would be in line with stoichiometric principles (i.e., due to a large share
387 in precursors between DOM samples) and could suggest that DOM processing diversifies, but also
388 "randomizes" the molecular composition of each precursor ("universal" signal).[31,98,99] High congruence of
389 fragmentation patterns (and thus, Δm matching) among DOM precursors has also been interpreted as a sign
390 of similarly substituted but slightly differing core structures.[35,37] The clusters devised here were small due
391 to the relatively small number of precursors and m/z values analyzed, and thus may not detect significant
392 differences between samples yet. However, even with our small set of precursors, the clustering by Δm
393 matching showed conserved differences in fragmentation between precursor clusters, and in part, even the
394 same precursor in different samples. The fact that this could relate to differences in ion abundance (and
395 therefore, possibly also ionization efficiency) or fragmentation sensitivity is intriguing and should be
396 investigated across a wider range of DOM chemotypes using improved classification approaches as applied
397 here (see also **section 3.4**).[14]

398 **3.6. Ion abundance is linked to Δm matching frequency and structural diversity.** Ion abundance was
399 the most important driver for Δm matching in both samples and highest in the structural domain usually
400 assigned to ubiquitous lignin structures or carboxyl-rich aromatic molecules.[59,83] This domain also parallels
401 with a maximum in potential underlying chemodiversity[30,100], which could explain why these signals are
402 ubiquitously found and especially dominant in reworked DOM.[92,101] Δm matching showed potential to reveal
403 this underlying chemodiversity effect and was therefore compared to numbers of structure suggestions and
404 Δm-predicted compound classes per precursor (**Figure 4**). Numbers of Δm matches were significantly and
405 positively related to the number of structure suggestions in absolute terms and for specific compound classes
406 (**Table S-22**). The correlation between Δm-predicted and suggested compound classes was surprisingly
407 similar in both samples and significant for almost all benzenoid-type (benzopyrans, methoxybenzenes,
408 anisoles, phenols, etc.) and most phenylpropanoid-type structures (flavonoids, linear 1,3-diarylpropanoids).
409 Among the organic acids, only vinylogous acids stood out (i.e., containing carboxylic acid groups with
410 insertions of C=C bond(s)). Significant correlations were also found for pyrans, acryloyl compounds,
411 carbohydrates, aryl ketones and alkyl aryl ethers (fatty acids and analogues only in SRNOM).

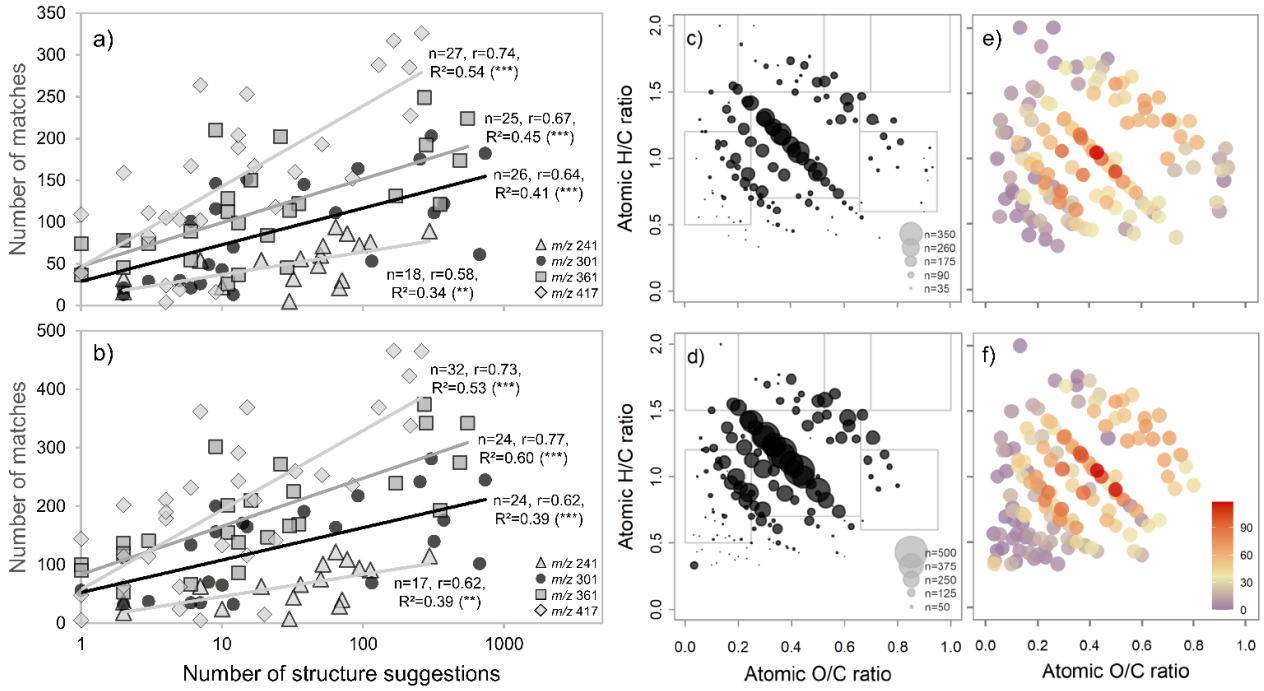


Figure 4. Agreement between chemodiversity estimates based on molecular formula (structure suggestions) and precursor-product ion links (Δm matches). Panels **a**, **b**) Correlations between numbers of SIRIUS Δm matches vs. structure suggestions (note log scale, incl. in-silico hits); a) soil DOM, b) SRNOM. Panels **c**, **d**) Number of SIRIUS Δm matches in Van Krevelen space (scales are similar but legends show different dot sizes); c) soil DOM, d) SRNOM; grey boxes refer to domains defined in **Figure S-3**. Panels **e**, **f**) Number of predicted classes per precursor based on SIRIUS Δm matches (color scale similar in both panels). Structural classes are associated to SIRIUS-annotated Δm features through correlation analysis of host structures and their Δm features (classification based on Classyfire); e) soil DOM, f) SRNOM.

The positive link between ion abundance and numbers of Δm matches on the one hand and predicted and suggested structures on the other indicates that ion abundance may be linked to the number of structural isomers and isobars per molecular formula in FTMS spectra of DOM and explains why Δm-defined structural domains showed strong overlap in this study. It also provides additional support to our assumption that all precursors potentially contribute to all product ions in DOM: The patterns revealed through Δm matching were largely congruent with the independent estimate of structural composition by natural product databases. The fact that only some classes of compounds (mainly benzenoids and phenylpropanoids) showed significant correlations could point to bias towards plant natural products in the databases employed here; this means that the inclusion of other structure databases and the additional assignment of Δm's not

430 only to their host structures but also to host organisms (e.g., in GNPS[65]) could reveal further clues about the
431 potential sources of molecular formulas in DOM.

432 We propose that the number of Δm matches could be interpreted as a novel, relatively easily accessible
433 measure to account for a precursor's underlying potential structural diversity. Such information could help
434 to better understand mechanisms of DOM formation and persistence in the environment. Our results
435 encourage further studies on the Δm matching behavior of synthetic mixtures of known structures and across
436 DOM chemotypes, and the improved bioinformatic exploitation of chimeric (LC-) $FTMS^n$ data of complex
437 organic mixtures.[14,102–104] We acknowledge that natural product and in-silico databases are far from being
438 complete, same as the database of annotated Δm matches we used here, despite its large coverage of ~18000
439 unique structures and ~11500 Δm's. For example, precursors with low mass defects showed exceptionally
440 few structural hits, indicating bias in natural product databases (**Figure S-12**).[18] These structures were easily
441 fragmented and yielded few Δm matches in our analysis; N- and S-containing precursors were double as
442 likely to show no suggestion compared to CHO precursors. This shows that DOM contains unique molecular
443 structures to be identified in future.

444 4. IMPLICATIONS

445 Tandem MS data of complex samples such as dissolved organic matter (DOM) is impeded by the co-
446 fragmentation of precursors with similar nominal mass, and further complicated by the contribution of
447 potential isomers and isobars of a precursor. We employed an approach that analyzes the pairwise mass
448 differences between all precursor and product ions as a whole (Δm matrix). Using a very limited set of
449 precursor features from two samples, we found potential signs of structural imprints related to benzenoids,
450 phenylpropanoids, carbohydrates, sulfonic acids, thiols, thioethers and amino acids, amongst others. The
451 successful matching of indicative Δm features and precursor clustering suggests a remaining – and
452 recognizable – source imprint of primary or recycled plant remains in DOM. Tests with more DOM samples
453 and artificial/ treated mixtures (e.g., DOM with spiked known compounds, or DOM degraded by specific
454 enzymes) are required to test the assumptions employed here and to improve classifications of DOM
455 precursors by Δm clusters. Our first results indicate that $FTMS^2$ data may be useful to differentiate molecular

456 composition on the molecular formula level, and that ion abundance and fragmentation sensitivity are two
457 key variables that explain differences in MS^2 data within and among samples. This is intriguing because a
458 shared molecular formula could harbor a completely different set of structures but must be assessed with
459 larger sets of DOM data which would improve detection of such differences. Generally, our findings support
460 the view that Van Krevelen domains are associated with indicative mass losses that relate to stoichiometric
461 differences between compound classes. The most abundant precursors however showed a mixed MS^2 signal
462 that caused boundary overlap of these Δm-defined domains (**Figure 4e, f**). While this finding is in line with
463 known patterns of structural diversity and partly explains the ubiquitous presence of abundant DOM signals,
464 it introduces a new paradigm to the interpretation of DOM FTMS data by assigning unknown precursors to
465 multiple structural categories instead of just one. Further evaluation of both natural and spiked/ treated
466 complex mixtures, constantly growing MS databases, and comprehensive decomplexation methods (LC-
467 MS, IMS) will together provide fundamental insights into the deconvolution of chimeric spectra from
468 complex samples, and ultimately show the potential to unfold the hidden molecular diversity and identity
469 of DOM.

470 **ASSOCIATED CONTENT**

471 **Data and Software Code Accessibility**

472 All tandem MS data can be found in on the Mass Spectrometry Interactive Virtual Environment (MassIVE) under the
473 following links: ftp://massive.ucsd.edu/MSV000087117/ (soil DOM data), ftp://massive.ucsd.edu/MSV000088869/
474 (SRNOM data) ftp://massive.ucsd.edu/MSV000087133/ (reference compound data) (**Data Set S-1**, raw peak data,
475 *.mzML files). All other data associated to this manuscript (extensive tables, Δm feature lists, Δm specificity, two-
476 way clustering table, processed data used to create figures, etc.) is available as online free of charge from the
477 PANGAEA Data Publisher under the following link: https://doi.pangaea.de/10.1594/PANGAEA.932592 (**Data Set S-**
478 **2**, processed data as *.xlsx files).

479 **Supporting Information**

480 The supporting information contains 22 tables and twelve figures, six additional notes, and 69 references.

481 Table S-1: Information on reference compounds and solutions used in this study. Table S-2: Instrument settings for
482 fragmentation experiments. Table S-3: Recalibration peaks used for reference compound in FTMS measurements.
483 Table S-4: Precursor and major product ions of the 14 reference compounds. Table S-5: Results of reference

compound's tandem MS data analysis with CSI:FingerID. Table S-6: List of reported DOM Δm features from MS1 and MS2 studies. Table S-7: List of all 50+5 Δm features extracted from the reference compound dataset. Table S-8: Properties of four isolated nominal masses (IPIMs) at different NCE levels. Table S-9: Overview of correlations between key properties of the IPIM 241. Table S-10: Overview of correlations between key properties of the IPIM 301. Table S-11: Overview of correlations between key properties of the IPIM 361. Table S-12: Overview of correlations between key properties of the IPIM 417. Table S-13: Lists of Δm values used for analysing matching patterns in Van Krevelen space. Table S-14: Matching behavior of precursor clusters against Δm features (Table S-7). Table S-15: Summary of two-way clustering of DOM precursors and reference compounds. Table S-16: Lignin-like precursor formulas and their molecular properties and clustering. Table S-17: S-containing precursor formulas in soil porewater DOM. Table S-18: N-containing precursor formulas in soil porewater DOM. Table S-19: S-containing precursor formulas in SRNOM. Table S-20: N-containing precursor formulas in SRNOM . Table S-21: Structural class-correlated Δm features matched to CHOS or CHNO precursors. Table S-22: Correlations between structure hits and specific Δm features in CHO precursors. Figure S-1: Overview of reference compounds used in the study. Figure S-2: Mass accuracy assessment based on reference compound Δm's. Figure S-3: Distribution of exemplary known structures in chemical space. Figure S-4: Orbitrap tandem MS of soil porewater DOM. Figure S-5: Comparison of matches to the two short Δm lists in relation to m/z and NCE. Figure S-6: Δm matches in relation to precursor ion abundance in soil DOM. Figure S-7: Δm matches in relation to precursor fragmentation sensitivity in soil DOM. Figure S-8: Matching assessment with SIRIUS Δm's (Molecular formula check). Figure S-9: Changes in Δm matching frequency upon widening of tolerance window. Figure S-10: Link between matches to $CH_3^{\bullet}$, CO and C_2H_4 and CH_4 vs. O exchange series. Figure S-11: Structural VK domains based on class-correlated SIRIUS Δm features. Figure S-12: Effect of mass defect on the number of structure suggestions. Note S-1: Supplementary experimental details. Note S-2: Detailed description of reference compound fragmentation behavior. Note S-3: Behavior of non-responsive DOM precursor ions. Note S-4: Δm matching: Proof-of-concept data and key findings. Note S-5: Potential esterification of DOM by methanol during SPE and storage. Note S-6: Structural insight into N- and S-containing DOM precursors.

The Supporting Information is available free of charge on the ACS Publications website.

Supporting information (PDF)

AUTHOR INFORMATION

Corresponding Author

*Email: gerd.gleixner@bgc-jena.mpg.de

Present Addresses

† C.S.: Institute for Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zurich, Switzerland & Swiss Federal Institute of Aquatic Science and Technology (Eawag), Department Water Resources and Drinking Water, Duebendorf, Switzerland

§ V.-N. R.: Thüringer Landesamt für Umwelt, Bergbau und Naturschutz (TLUBN), Jena, Germany

Author Contributions

CS performed the measurements. DP, VNR, PD and GG were involved in planning and supervised the work. KD and SB compiled global Δm feature data, analyzed its specificity, and performed structural classifications of Δm host structures as well as structure suggestions of DOM precursors. CS processed the experimental data, performed the downstream analyses, drafted the manuscript, and designed the figures. The manuscript was revised through the contributions of all authors. All authors have approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank all members of the Dorrestein lab for helpful discussions and insights. We also like to acknowledge Vivian Stefanow for initial structure database surveys. We acknowledge the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC) for sponsoring CS' research stay in the Dorrestein lab at UCSD, CA, USA. The authors acknowledge the financial support from the Max-Planck-Gesellschaft (MPG) and the German Research Foundation (DFG, Deutsche Forschungs-Gemeinschaft) as part of the CRC 1076 "Aqua Diva" and a Postdoctoral Research Fellowship PE 2600/1, and support through the CMFI Cluster of Excellence (EXC 2124) to D.P. V.-N.R. received additional funding by the Zwillenberg-Tietz Stiftung. C.S. received a Ph.D. stipend from the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC). The authors are grateful for critical comments and helpful suggestions received from the three anonymous reviewers.

REFERENCES

(1) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51*, 11505–11512. https://doi.org/10.1021/acs.est.7b02184.

(2) D'Andrilli, J.; Fischer, S. J.; Rosario-Ortiz, F. L. Advancing Critical Applications of High Resolution Mass Spectrometry for DOM Assessments: Re-Engaging with Mass Spectral Principles, Limitations, and Data Analysis. *Environ. Sci. Technol.* **2020**, *54*, 11654−11656. https://doi.org/10.1021/acs.est.0c04557.

(3) Kruve, A. Strategies for Drawing Quantitative Conclusions from Nontargeted Liquid Chromatography–High-Resolution Mass Spectrometry Analysis. *Anal. Chem.* **2020**, *92*, 4691–4699. https://doi.org/10.1021/acs.analchem.9b03481.

(4) Bahureksa, W.; Tfaily, M. M.; Boiteau, R. M.; Young, R. B.; Logan, M. N.; McKenna, A. M.; Borch, T. Soil Organic Matter Characterization by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR MS): A Critical Review of Sample Preparation, Analysis, and Data Interpretation. *Environ. Sci. Technol.* **2021**, *55*, 9637–9656. https://doi.org/10.1021/acs.est.1c01135.

(5) Zsolnay, Á. Dissolved Organic Matter: Artefacts, Definitions, and Functions. *Geoderma* **2003**, *113*, 187–209. https://doi.org/10.1016/S0016-7061(02)00361-0.

(6) Wells, M. J. M.; Stretz, H. A. Supramolecular Architectures of Natural Organic Matter. *Sci. Total Environ.* **2019**, *671*, 1125–1133. https://doi.org/10.1016/j.scitotenv.2019.03.406.

(7) Prescott, C. E.; Grayston, S. J.; Helmisaari, H. S.; Kaštovská, E.; Körner, C.; Lambers, H.; Meier, I. C.; Millard, P.; Ostonen, I. Surplus Carbon Drives Allocation and Plant–Soil Interactions. *Trends Ecol. Evol.* **2020**, *35*, 1110–1118. https://doi.org/10.1016/j.tree.2020.08.007.

(8) Leinemann, T.; Preusser, S.; Mikutta, R.; Kalbitz, K.; Cerli, C.; Höschen, C.; Mueller, C. W.; Kandeler, E.; Guggenberger, G. Multiple Exchange Processes on Mineral Surfaces Control the Transport of Dissolved Organic Matter through Soil Profiles. *Soil Biol. Biochem.* **2018**, *118*, 79–90. https://doi.org/10.1016/j.soilbio.2017.12.006.

(9) Lange, M.; Roth, V. N.; Eisenhauer, N.; Roscher, C.; Dittmar, T.; Fischer-Bedtke, C.; González Macé, O.; Hildebrandt, A.; Milcu, A.; Mommer, L.; et al. Plant Diversity Enhances Production and Downward Transport of Biodegradable Dissolved Organic Matter. *J. Ecol.* **2020**, No. February, 1–14. https://doi.org/10.1111/1365-2745.13556.

(10) Lehmann, J.; Hansel, C. M.; Kaiser, C.; Kleber, M.; Maher, K.; Manzoni, S.; Nunan, N.; Reichstein, M.; Schimel, J. P.; Torn, M. S.; et al. Persistence of Soil Organic Carbon Caused by Functional Complexity. *Nat. Geosci.* **2020**, *13*, 529–534. https://doi.org/10.1038/s41561-020-0612-3.

(11) Bünemann, E. K.; Bongiorno, G.; Bai, Z.; Creamer, R. E.; De Deyn, G.; de Goede, R.; Fleskens, L.; Geissen, V.; Kuyper, T. W.; Mäder, P.; et al. Soil Quality – A Critical Review. *Soil Biol. Biochem.* **2018**, *120*, 105–125. https://doi.org/10.1016/j.soilbio.2018.01.030.

(12) Simpson, A. J.; Simpson, M. J.; Soong, R. Environmental Nuclear Magnetic Resonance Spectroscopy: An Overview and a Primer. *Anal. Chem.* **2018**, *90*, 628–639. https://doi.org/10.1021/acs.analchem.7b03241.

(13) Lu, K.; Li, X.; Chen, H.; Liu, Z. Constraints on Isomers of Dissolved Organic Matter in Aquatic Environments: Insights from Ion Mobility Mass Spectrometry. *Geochim. Cosmochim. Acta* **2021**, *308*, 353–372. https://doi.org/10.1016/j.gca.2021.05.007.

(14) Leyva, D.; Tariq, M. U.; Jaffé, R.; Saeed, F.; Lima, F. F. Unsupervised Structural Classification of Dissolved Organic Matter Based on Fragmentation Pathways. *Environ. Sci.*

Technol. **2022**, *56*, 1458–1468. https://doi.org/10.1021/acs.est.1c04726.

(15) Wegley, L.; Nelson, C. E.; Petras, D.; Koester, I.; Quinlan, Z. A.; Arts, M. G. I.; Nothias, L.; Comstock, J.; White, B. M.; Hopmans, E. C. Distinguishing the Molecular Diversity , Nutrient Content , and Energetic Potential of Exometabolomes Produced by Macroalgae and Reef-Building Corals. *PNAS* **2022**, *119*, e2110283119. https://doi.org/10.1073/pnas.2110283119/-/DCSupplemental.Published.

(16) 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*, 16249–16259. https://doi.org/10.1021/acs.est.0c05206.

(17) Arakawa, N.; Aluwihare, L. I.; Simpson, A. J.; Soong, R.; Stephens, B. M.; Lane-Coplen, D. Carotenoids Are the Likely Precursor of a Significant Fraction of Marine Dissolved Organic Matter. *Sci. Adv.* **2017**, *3* (9), e1602976. https://doi.org/10.1126/sciadv.1602976.

(18) Brown, T. A.; Jackson, B. A.; Bythell, B. J.; Stenson, A. C. Benefits of Multidimensional Fractionation for the Study and Characterization of Natural Organic Matter. *J. Chromatogr. A* **2016**, *1470*, 84–96. https://doi.org/10.1016/j.chroma.2016.10.005.

(19) 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. https://doi.org/10.1002/lol2.10064.

(20) Mostovaya, A.; Hawkes, J. A.; Koehler, B.; Dittmar, T.; Tranvik, L. J. Emergence of the Reactivity Continuum of Organic Matter from Kinetics of a Multitude of Individual Molecular Constituents. *Environ. Sci. Technol.* **2017**, *51*, 11571–11579. https://doi.org/10.1021/acs.est.7b02876.

(21) Murphy, K. R.; Timko, S. A.; Gonsior, M.; Powers, L. C.; Wünsch, U. J.; Stedmon, C. A. Photochemistry Illuminates Ubiquitous Organic Matter Fluorescence Spectra. *Environ. Sci. Technol.* **2018**, *52*, 11243–11250. https://doi.org/10.1021/acs.est.8b02648.

(22) Benner, R.; Amon, R. M. W. The Size-Reactivity Continuum of Major Bioelements in the Ocean. *Ann. Rev. Mar. Sci.* **2014**, No. July 2014, 1–21. https://doi.org/10.1146/annurev-marine-010213-135126.

(23) Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E. M. Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Anal. Chem.* **2008**, *80*, 8908–8919.

(24) van Agthoven, M. A.; Lam, Y. P. Y.; O'Connor, P. B.; Rolando, C.; Delsuc, M. A. Two-Dimensional Mass Spectrometry: New Perspectives for Tandem Mass Spectrometry. *Eur. Biophys. J.* **2019**, No. 48, 213–229. https://doi.org/10.1007/s00249-019-01348-5.

(25) 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*, 11960–11966. https://doi.org/10.1021/acs.analchem.0c02347.

(26) 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**, *57C*, 280–294. https://doi.org/10.1016/j.watres.2014.02.051.

(27) Petras, D.; Minich, J. J.; Cancelada, L. C.; Torres, R. E.; Kunselman, E.; Wang, M.; White, M. E.; Allen, E. E.; Prather, K. A.; Aluwihare, L. I.; et al. Non-Targeted Tandem Mass Spectrometry Enables the Visualization of Organic Matter Chemotype Shifts in Coastal Seawater. *Chemosphere* **2021**, *271*, 129450. https://doi.org/10.1016/j.chemosphere.2020.129450.

(28) Leenheer, J. A.; Rostad, C. E.; Gates, P. M.; Furlong, E. T.; Ferrer, I. Molecular Resolution and Fragmentation of Fulvic Acid by Electrospray Ionization/ Multistage Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73* (7), 1461–1471. https://doi.org/10.1021/ac0012593.

(29) Dit Fouque, D. J.; Maroto, A.; Memboeuf, A. Purification and Quantification of an Isomeric Compound in a Mixture by Collisional Excitation in Multistage Mass Spectrometry Experiments. *Anal. Chem.* **2016**, *88* (22), 10821–10825. https://doi.org/10.1021/acs.analchem.6b03490.

(30) 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. *Front. Mar. Sci.* **2017**, *4* (December), 406. https://doi.org/10.3389/fmars.2017.00405.

(31) Zark, M.; Dittmar, T. Universal Molecular Structures in Natural Dissolved Organic Matter. *Nat. Commun.* **2018**, *9* (1), 3178. https://doi.org/10.1038/s41467-018-05665-9.

(32) 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. https://doi.org/10.1021/acs.est.8b00999.

(33) 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. https://doi.org/10.1021/ac802624s.

(34) 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*, 1430–1437. https://doi.org/10.1021/es7021413.

(35) 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. https://doi.org/10.1021/ac1016216.

(36) Cortés-Francisco, N.; Caixach, J. Fragmentation Studies for the Structural Characterization of Marine Dissolved Organic Matter. *Anal. Bioanal. Chem.* **2015**, *407*, 2455–2462.

https://doi.org/10.1007/s00216-015-8499-3.

(37) 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. https://doi.org/10.1002/rcm.1550.

(38) 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 Molecular Formulas Have Been Assigned for 4626 Indi- Mass Measurements Fr. *Anal. Chem.* **2003**, *75*, 1275–1284. https://doi.org/10.1021/ac026106p.

(39) Nimmagadda, R. D.; McRae, C. Characterisation of the Backbone Structures of Several Fulvic Acids Using a Novel Selective Chemical Reduction Method. *Org. Geochem.* **2007**, *38* (7), 1061–1072. https://doi.org/10.1016/j.orggeochem.2007.02.016.

(40) Perdue, E. M.; Hertkorn, N.; Kettrup, A. Substitution Patterns in Aromatic Rings by Increment Analysis. Model Development and Application to Natural Organic Matter. *Anal. Chem.* **2007**, *79* (3), 1010–1021. https://doi.org/10.1021/ac061611y.

(41) Kunenkov, E. V.; Kononikhin, A. S.; Perminova, I. V.; Hertkorn, N.; Gaspar, A.; Schmitt-kopplin, P.; Popov, I. A.; Garmash, A. V.; Nikolaev, E. N. Total Mass Difference Statistics Algorithm : A New Approach to Identification of High-Mass Building Blocks in Electrospray Ionization Fourier Transform Ion Cyclotron Mass Spectrometry Data of Natural Organic Matter. *Anal. Chem.* **2009**, *81* (24), 10106–10115. https://doi.org/10.1021/ac901476u.

(42) Zherebker, A. Y.; Airapetyan, D.; Konstantinov, A. I.; Kostyukevich, Y. I.; Kononikhin, A. S.; Popov, I. A.; Zaitsev, K. V.; Nikolaev, E. N.; Perminova, I. V. Synthesis of Model Humic Substances: A Mechanistic Study Using Controllable H/D Exchange and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analyst* **2015**, *140* (13), 4708–4719. https://doi.org/10.1039/c5an00602c.

(43) Bell, N. G. A.; Michalchuk, A. A. L.; Blackburn, J. W. T.; Graham, M. C.; Uhrín, D. Isotope-Filtered 4D NMR Spectroscopy for Structure Determination of Humic Substances. *Angew. Chemie - Int. Ed.* **2015**, *54* (29), 8382–8385. https://doi.org/10.1002/anie.201503321.

(44) McIntyre, C.; McRae, C.; Jardine, D.; Batts, B. D. Identification of Compound Classes in Soil and Peat Fulvic Acids as Observed by Electrospray Ionization Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1604–1609. https://doi.org/10.1002/rcm.761.

(45) 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. https://doi.org/10.1016/j.marchem.2017.02.005.

(46) Stenson, A. C.; Ruddy, B. M.; Bythell, B. J. Ion Molecule Reaction H/D Exchange as a Probe for Isomeric Fractionation in Chromatographically Separated Natural Organic Matter.

Int. J. Mass Spectrom. **2014**, *360*, 45–53. https://doi.org/10.1016/j.ijms.2013.12.026.

(47) Kostyukevich, Y.; Kononikhin, A.; Zherebker, A.; Popov, I.; Perminova, I.; Nikolaev, E. Enumeration of Non-Labile Oxygen Atoms in Dissolved Organic Matter by Use of 16O/18O Exchange and Fourier Transform Ion-Cyclotron Resonance Mass Spectrometry. *Anal. Bioanal. Chem.* **2014**, *406* (26), 6655–6664. https://doi.org/10.1007/s00216-014-8097-9.

(48) Zherebker, A.; Kostyukevich, Y.; Kononikhin, A.; Kharybin, O.; Konstantinov, A. I.; Zaitsev, K. V.; Nikolaev, E.; Perminova, I. V. Enumeration of Carboxyl Groups Carried on Individual Components of Humic Systems Using Deuteromethylation and Fourier Transform Mass Spectrometry. *Anal. Bioanal. Chem.* **2017**, *409*, 2477–2488. https://doi.org/10.1007/s00216-017-0197-x.

(49) Liu, Z.; Sleighter, R. L.; Zhong, J.; Hatcher, P. G. The Chemical Changes of DOM from Black Waters to Coastal Marine Waters by HPLC Combined with Ultrahigh Resolution Mass Spectrometry. *Estuar. Coast. Shelf Sci.* **2011**, *92*, 205–216. https://doi.org/10.1016/j.ecss.2010.12.030.

(50) Dier, T. K. F.; Egele, K.; Fossog, V.; Hempelmann, R.; Volmer, D. A. Enhanced Mass Defect Filtering to Simplify and Classify Complex Mixtures of Lignin Degradation Products. *Anal. Chem.* **2016**, *88*, 1328–1335. https://doi.org/10.1021/acs.analchem.5b03790.

(51) Fabre, N.; Rustan, I.; De Hoffmann, E.; Quetin-Leclercq, J. Determination of Flavone, Flavonol, and Flavanone Aglycones by Negative Ion Liquid Chromatography Electrospray Ion Trap Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2001**, *12* (6), 707–715. https://doi.org/10.1016/S1044-0305(01)00226-4.

(52) Engström, M. T.; Pälijärvi, M.; Salminen, J. P. Rapid Fingerprint Analysis of Plant Extracts for Ellagitannins, Gallic Acid, and Quinic Acid Derivatives and Quercetin-, Kaempferol- and Myricetin-Based Flavonol Glycosides by UPLC-QqQ-MS/MS. *J. Agric. Food Chem.* **2015**, *63* (16), 4068–4079. https://doi.org/10.1021/acs.jafc.5b00595.

(53) Miketova, P.; Schram, K. H.; Whitney, J.; Li, M.; Huang, R.; Kerns, E.; Valcic, S.; Timmermann, B. N.; Rourick, R.; Klohr, S. Tandem Mass Spectrometry Studies of Green Tea Catechins. Identification of Three Minor Components in the Polyphenolic Extract of Green Tea. *J. Mass Spectrom.* **2000**, *35* (7), 860–869. https://doi.org/10.1002/1096-9888(200007)35:7<860::AID-JMS10>3.0.CO;2-J.

(54) Luek, J. L.; Schmitt-kopplin, P.; Mouser, P. J.; Petty, W. T.; Richardson, S. D.; Gonsior, M. Halogenated Organic Compounds Identified in Hydraulic Fracturing Wastewaters Using Ultrahigh Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2017**, *51*, 5377–5385. https://doi.org/10.1021/acs.est.6b06213.

(55) Reemtsma, T. The Carbon versus Mass Diagram to Visualize and Exploit FTICR-MS Data of Natural Organic Matter. *J. Mass Spectrom.* **2010**, *45* (4), 382–390. https://doi.org/10.1002/jms.1722.

(56) Rivas-Ubach, A.; Liu, Y.; Bianchi, T. S.; Tolić, N.; Jansson, C.; Paša-Tolić, L. Moving beyond the van Krevelen Diagram: A New Stoichiometric Approach for Compound

Classification in Organisms. *Anal. Chem.* **2018**, *90*, 6152–6160. https://doi.org/10.1021/acs.analchem.8b00529.

(57) Davies, N. W.; Sandron, S.; Nesterenko, P.; Paull, B.; Wilson, R.; Haddad, P.; Shellie, R.; Rojas, A. Comment on "Structural Characterization of Dissolved Organic Matter: A Review of Current Techniques for Isolation and Analysis" by E. C. Minor, M. M. Swenson, B. M. Mattson, and A. R. Oyler, Environ. Sci.: Processes Impacts, 2014, 16 , 2064. *Environ. Sci. Process. Impacts* **2015**, *17* (2), 495. https://doi.org/10.1039/C4EM00631C.

(58) 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. *Environ. Sci. Process. Impacts* **2014**, *16*, 2064–2079. https://doi.org/10.1039/C4EM00062E.

(59) Roth, V.-N.; Dittmar, T.; Gaupp, R.; Gleixner, G. Ecosystem-Specific Composition of Dissolved Organic Matter. *Vadose Zo. J.* **2014**, *13*. https://doi.org/http://dx.doi.org/10.2136/vzj2013.09.0162.

(60) Green, N. W.; Mcinnis, D.; Hertkorn, N.; Maurice, P. A.; Perdue, M. E. Suwannee River Natural Organic Matter : Isolation of the 2R101N Reference Sample by Reverse Osmosis. *Environ. Eng. Sci.* **2014**, *32*, 38–44. https://doi.org/10.1089/ees.2014.0284.

(61) Simon, C.; Roth, V.-N.; Dittmar, T.; Gleixner, G. Molecular Signals of Heterogeneous Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. *Front. Earth Sci.* **2018**, *6*, 1–16. https://doi.org/10.3389/feart.2018.00138.

(62) Merder, J.; Freund, J. A.; Feudel, U.; Hansen, C. T.; Hawkes, J. A.; Jacob, B.; Klaproth, K.; Niggemann, J.; Noriega-Ortega, B. E.; Osterholz, H.; et al. ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Anal. Chem.* **2020**, *92*, 6832–6838. https://doi.org/10.1021/acs.analchem.9b05659.

(63) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16*, 299–302. https://doi.org/10.1038/s41592-019-0344-8.

(64) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714. https://doi.org/10.1002/jms.1777.

(65) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. T. D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. https://doi.org/10.1038/nbt.3597.

(66) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminform.* **2016**, *8*, 61.

https://doi.org/10.1186/s13321-016-0174-y.

(67) Hammer, Ø.; Harper, D. A.; Ryan, P. D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2001**, *4*, 9.

(68) Milstead, R. P.; Remucal, C. K. Molecular-Level Insights into the Formation of Traditional and Novel Halogenated Disinfection Byproducts. *ACS ES&T Water* **2021**, *1*, 1966–1974. https://doi.org/10.1021/acsestwater.1c00161.

(69) Wilson, R. M.; Tfaily, M. M.; Kolton, M.; Johnston, E. R.; Petro, C.; Zalman, C. A.; Hanson, P. J.; Heyman, H. M.; Kyle, J. E.; Hoyt, D. W.; et al. Soil Metabolome Response to Whole-Ecosystem Warming at the Spruce and Peatland Responses under Changing Environments Experiment. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, 1–11. https://doi.org/10.1073/pnas.2004192118.

(70) Wu, S.; You, F.; Boughton, B.; Liu, Y.; Nguyen, T. A. H.; Wykes, J.; Southam, G.; Robertson, L. M.; Chan, T. S.; Lu, Y. R.; et al. Chemodiversity of Dissolved Organic Matter and Its Molecular Changes Driven by Rhizosphere Activities in Fe Ore Tailings Undergoing Eco-Engineered Pedogenesis. *Environ. Sci. Technol.* **2021**, *55*, 13045–13060. https://doi.org/10.1021/acs.est.1c04527.

(71) Hawkes, J. A.; D'Andrilli, J.; Agar, J. N.; Barrow, M. P.; Berg, S. M.; Catalán, N.; Chen, H.; Chu, R. K.; Cole, R. B.; Dittmar, T.; et al. An International Laboratory Comparison of Dissolved Organic Matter Composition by High Resolution Mass Spectrometry: Are We Getting the Same Answer? *Limnol. Oceanogr. Methods* **2020**, *18*, 235–258.

(72) Chassagne, F.; Cabanac, G.; Hubert, G.; David, B.; Marti, G. The Landscape of Natural Product Diversity and Their Pharmacological Relevance from a Focus on the Dictionary of Natural Products®. *Phytochem. Rev.* **2019**, 1–22. https://doi.org/10.1007/s11101-019-09606-2.

(73) Nakamura, Y.; Mochamad Afendi, F.; Kawsar Parvin, A.; Ono, N.; Tanaka, K.; Hirai Morita, A.; Sato, T.; Sugiura, T.; Altaf-Ul-Amin, M.; Kanaya, S. KNApSAcK Metabolite Activity Database for Retrieving the Relationships between Metabolites and Biological Activities. *Plant Cell Physiol.* **2014**, *55*, e7. https://doi.org/10.1093/pcp/pct176.

(74) Caspi, R.; Billington, R.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Midford, P. E.; Ong, W. K.; Paley, S.; Subhraveti, P.; Karp, P. D. The MetaCyc Database of Metabolic Pathways and Enzymes - a 2019 Update. *Nucleic Acids Res.* **2019**, *48*, D455–D453. https://doi.org/10.1093/nar/gkz862.

(75) Okuda, S.; Yamada, T.; Hamajima, M.; Itoh, M.; Katayama, T.; Bork, P.; Goto, S.; Kanehisa, M. KEGG Atlas Mapping for Global Analysis of Metabolic Pathways. *Nucleic Acids Res.* **2008**, *36*, 423–426. https://doi.org/10.1093/nar/gkn282.

(76) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; et al. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35*, 521–526. https://doi.org/10.1093/nar/gkl923.

(77) Jeffryes, J. G.; Colastani, R. L.; Elbadawi-Sidhu, M.; Kind, T.; Niehaus, T. D.; Broadbelt,

L. J.; Hanson, A. D.; Fiehn, O.; Tyo, K. E. J.; Henry, C. S. MINEs: Open Access Databases of Computationally Predicted Enzyme Promiscuity Products for Untargeted Metabolomics. *J. Cheminform.* **2015**, *7*, 44. https://doi.org/10.1186/s13321-015-0087-1.

(78) Baumeister, T. U. H.; Ueberschaar, N.; Pohnert, G. Gas-Phase Chemistry in the GC Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2018**. https://doi.org/10.1007/s13361-018-2117-5.

(79) Smirnov, K. S.; Forcisi, S.; Moritz, F.; Lucio, M.; Schmitt-Kopplin, P. Mass Difference Maps and Their Application for the Re-Calibration of Mass Spectrometric Data in Non-Targeted Metabolomics. *Anal. Chem.* **2019**. https://doi.org/10.1021/acs.analchem.8b04555.

(80) Adair, E.; Afonso, C.; Bell, N. G. A.; Davies, A. N.; Delsuc, M.-A.; Godfrey, R.; Goodacre, R.; Hawkes, J. A.; Hertkorn, N.; Jones, D.; et al. High Resolution Techniques: General Discussion. *Faraday Discuss.* **2019**, *218*, 247–267. https://doi.org/10.1039/c9fd90045d.

(81) Novotny, N. R.; Capley, E. N.; Stenson, A. C. Fact or Artifact: The Representativeness of ESI-MS for Complex Natural Organic Mixtures. *J. Mass Spectrom.* **2014**, *49* (4), 316–326. https://doi.org/10.1002/jms.3345.

(82) 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*, 8240–8247. https://doi.org/10.1021/es0713072.

(83) 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*, 2990–3010. https://doi.org/10.1016/j.gca.2006.03.021.

(84) Liu, L.; Song, C.; Tian, S.; Zhang, Q.; Cai, X.; Liu, Y.; Liu, Z.; Wang, W. Structural Characterization of Sulfur-Containing Aromatic Compounds in Heavy Oils by FT-ICR Mass Spectrometry with a Narrow Isolation Window. *Fuel* **2019**, *240*, 40–48. https://doi.org/10.1016/j.fuel.2018.11.130.

(85) Böcker, S.; Dührkop, K. Fragmentation Trees Reloaded. *J. Cheminform.* **2016**, *8*, 5. https://doi.org/10.1007/978-3-319-16706-0_10.

(86) Cao, D.; Lv, J.; Geng, F.; Rao, Z.; Niu, H.; Shi, Y.; Cai, Y.; Kang, Y. Ion Accumulation Time Dependent Molecular Characterization of Natural Organic Matter Using Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 12210–12218. https://doi.org/10.1021/acs.analchem.6b03198.

(87) Lemr, K.; Holčapek, M.; Jandera, P.; Lyka, A. Analysis of Metal Complex Azo Dyes by High-Performance Liquid Chromatography/Electrospray Ionization Mass Spectrometry and Multistage Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1881–1888.

(88) Piraud, M.; Vianey-Saban, C.; Petritis, K.; Elfakir, C.; Steghens, J. P.; Morla, A.; Bouchu, D. ESI-MS/MS Analysis of Underivatised Amino Acids: A New Tool for the Diagnosis of Inherited Disorders of Amino Acid Metabolism. Fragmentation Study of 79 Molecules of Biological Interest in Positive and Negative Ionisation Mode. *Rapid Commun. Mass*

Spectrom. **2003**, *17*, 1297–1311. https://doi.org/10.1002/rcm.1054.

(89) Pohlabeln, A. M.; Dittmar, T. Novel Insights into the Molecular Structure of Non-Volatile Marine Dissolved Organic Sulfur. *Mar. Chem.* **2015**, *168*, 86–94. https://doi.org/10.1016/j.marchem.2014.10.018.

(90) Wagner, S.; Dittmar, T.; Jaffé, R. Molecular Characterization of Dissolved Black Nitrogen via Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Org. Geochem.* **2015**, *79*, 21–30. https://doi.org/10.1016/j.orggeochem.2014.12.002.

(91) Poulin, B. A.; Ryan, J. N.; Nagy, K. L.; Stubbins, A.; Dittmar, T.; Orem, W.; Krabbenhoft, D. P.; Aiken, G. R. Spatial Dependence of Reduced Sulfur in Everglades Dissolved Organic Matter Controlled by Sulfate Enrichment. *Environ. Sci. Technol.* **2017**, *51*, 3630–3639. https://doi.org/10.1021/acs.est.6b04142.

(92) Roth, V.-N.; Lange, M.; Simon, C.; Hertkorn, N.; Bucher, S.; Goodall, T.; Griffiths, R. I.; Mellado-Vázquez, P. G.; Mommer, L.; Oram, N. J.; et al. Persistence of Dissolved Organic Matter Explained by Molecular Changes during Its Passage through Soil. *Nat. Geosci.* **2019**, *12*, 755–761. https://doi.org/10.1038/s41561-019-0417-4.

(93) Warren, C. R. High Diversity of Small Organic N Observed in Soil Water. *Soil Biol. Biochem.* **2013**, *57*, 444–450. https://doi.org/10.1016/j.soilbio.2012.09.025.

(94) Woods, G.; Simpson, M.; Koerner, P. J.; Napoli, A.; Simpson, A. HILIC-NMR: Toward the Identification of Individual Molecular Components in Dissolved Organic Matter. *Environ. Sci. Technol.* **2011**, *45* (13), 5910. https://doi.org/10.1021/es201716u.

(95) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-Ion Fragmentation. *Mol. Cell. Proteomics* **2010**, *9*, 2252–2261. https://doi.org/10.1074/mcp.M110.001537.

(96) Naz, S.; Gallart-Ayala, H.; Reinke, S. N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C. E. Development of a Liquid Chromatography-High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal. Chem.* **2017**, *89*, 7933–7942. https://doi.org/10.1021/acs.analchem.7b00925.

(97) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-independent Acquisition-based SWATH - MS for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2018**, *14*, 1–23. https://doi.org/10.15252/msb.20178126.

(98) Lechtenfeld, O. J.; Hertkorn, N.; Shen, Y.; Witt, M.; Benner, R. Marine Sequestration of Carbon in Bacterial Metabolites. *Nat. Commun.* **2015**, *6*, 6711. https://doi.org/10.1038/ncomms7711.

(99) Mentges, A.; Feenders, C.; Seibt, M.; Blasius, B.; Dittmar, T. Functional Molecular Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. *Front. Mar. Sci.* **2017**, *4*, 194. https://doi.org/10.3389/fmars.2017.00194.

(100) 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*, 1311–1327. https://doi.org/10.1007/s00216-007-1577-4.

(101) Lechtenfeld, O. J.; Kattner, G.; Flerus, R.; McCallister, S. L.; Schmitt-Kopplin, P.; Koch, B. P. Molecular Transformation and Degradation of Refractory Dissolved Organic Matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **2014**, *126*, 321–337. https://doi.org/10.1016/j.gca.2013.11.009.

(102) Dührkop, K.; Nothias, L. F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M. A.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; et al. Systematic Classification of Unknown Metabolites Using High-Resolution Fragmentation Mass Spectra. *Nat. Biotechnol.* **2020**. https://doi.org/10.1038/s41587-020-0740-8.

(103) Rogers, S.; Wei Ong, C.; Wandy, J.; Ernst, M.; Ridder, L.; van der Hooft, J. J. J. Deciphering Complex Metabolite Mixtures by Unsupervised and Supervised Substructure Discovery and Semi-Automated Annotation from MS/MS Spectra. *Faraday Discuss.* **2019**, *218*, 284–302. https://doi.org/10.1039/c8fd00235e.

(104) Wolfender, J.-L.; Nuzillard, J.-M.; Van Der Hooft, J. J. J.; Renault, J.-H.; Bertrand, S. Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics. *Anal. Chem.* **2019**, *91*, 704–742. https://doi.org/10.1021/acs.analchem.8b05112.