

Mass Difference Matching Unfolds Hidden Molecular Structures of Dissolved Organic Matter

Carsten Simon,* Kai Dührkop, Daniel Petras, Vanessa-Nina Roth, Sebastian Böcker, Pieter C. Dorrestein, and Gerd Gleixner*



Cite This: *Environ. Sci. Technol.* 2022, 56, 11027–11040



Read Online

ACCESS |

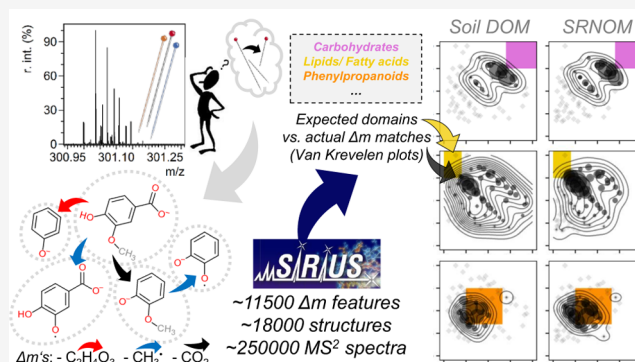
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Ultrahigh-resolution Fourier transform mass spectrometry (FTMS) has revealed unprecedented details 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 ion is potentially linked with all emerging product ions in FTMS² experiments. The resulting mass difference (Δm) matrix is deconvoluted to isolate individual precursor ion 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 11 477 Δm features with assigned structure specificities, using a dataset of $\sim 18\,000$ unique structures. We show that Δm matching is highly sensitive in predicting potential precursor ion identities in terms of molecular and structural composition. Additionally, the approach identified unresolved precursor ions and missing elements in molecular formula annotation (P, Cl, F). Our study provides first results on how Δm matching refines structural annotations in van Krevelen space but simultaneously demonstrates the wide overlap between potential structural classes. 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 both unfold the structural information encrypted in DOM and assess the quality of FTMS-derived molecular formulas of complex mixtures in general.

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 ultracomplex 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 details of molecular formulas using the exact mass (MS¹ data, m/z) even without prior separation.²³ However, FTMS techniques are selective and do not resolve all structural detail observed at the exact mass in DOM, as the presence of isobars and isomers hinders the identification of particular structures from these molecular formulas.^{19,23–25} Additionally, current structural databases cover only a small fraction of molecular formulas encountered and typically lead to annotation rates $<5\%$.^{18,26,27}

One way to obtain structure information on isomers and isobars is through collision-induced dissociation (CID; MS², or multistage MSⁿ).^{27–29} The relatively wide isolation window (~ 1 Da) of mass filters applied for precursor ion selection

Received: February 22, 2022

Revised: June 20, 2022

Accepted: June 30, 2022

Published: July 14, 2022



commonly hinders the isolation and subsequent fragmentation of single exact masses, leading to mixed “chimeric” MS² spectra of co-fragmented precursor ions.³⁰ Even though some authors achieved isolation of single masses or improved description of chimeric tandem MS data, fragmentation patterns were found to be universal across DOM samples.^{18,19,31–35} Most of these studies, however, focused on the major product ion peaks (fragments), which usually make up only 60–70% of the total product ion abundance, and thus disregarded many low-abundance signals that may be more suitable to detecting structural differences.^{19,31}

The major product ions encountered in tandem mass spectra of DOM relate to sequential neutral losses of common small building blocks, mainly CO₂, H₂O, or CO units.^{14,33} A mass difference between a precursor and a product ion in an MS² spectrum is herein called “delta mass” and referred to as Δm (plural Δm 's). Common Δm 's such as CO₂ or H₂O are deemed nonindicative for the identification of structural units.^{18,28,31,33,36} In contrast, other studies found recurring low- m/z product ions (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 structural units substituted with a set of functional groups, yet in different amounts and configurational types that would lead to highly diverse mixtures.^{37–44} From a stochastic standpoint, the occurrence of common neutral losses may not be surprising; many structures contain hydroxyl groups that could yield H₂O losses, and CO₂ could originate from ubiquitous carboxyl groups.⁴⁵ In contrast, the occurrence of two molecules sharing a larger substructure would be less probable and less easily detected as a major peak. Signatures of DOM's structural diversity could thus prevail in the high number of low-abundance fragments usually detected below m/z 200–300, as opposed to the higher abundance of fragments connected to losses of CO₂ or H₂O. Given the large number of estimated isomers and isobars underlying usual DOM data,^{18,19,31,32,39,45–48} we here build upon the hypothesis that every co-fragmented precursor ion potentially contributes to every emerging product ion signal. We interpret the resulting chimeric MS² data as a structural fingerprint that can be deconvoluted to obtain individual precursor ion Δm matching profiles. The analysis of Δm 's that link precursor and product ions is independent of the masses of the unknown precursor ions and known reference compounds in databases of annotated Δm features, and therefore does not rely on indicative product ions (fragments) alone. Although this approach sacrifices the identification of true knowns, it allows for the identification of potential structural analogues via indicative Δm 's and may be especially suited when annotation rates are as low as in the case of DOM, i.e., when most compounds are unknown.^{18,26,27}

Despite the unknown identity of most of the molecules present in DOM, its potential sources can be constrained reasonably well. Plants produce most of the organic matter that sustains food webs in natural ecosystems. Plant metabolites such as polyphenols thus represent a major source of DOM. Therefore, an early decomposition phase likely exists when the imprint of soluble/solubilized plant metabolites is still detectable by MS² experiments using current FTMS technology: Lignin-related compounds show indicative methoxyl/methyl radical losses,^{18,49,50} glycosides indicate a sugar loss,^{51,52} and hydrolyzable tannins may lose galloyl units.⁵² Mass differences related to atoms such as N, S, P, Cl, Br, I, and F could also help to identify unknown organic nutrient species

or disinfection byproducts, thereby widening the applicability of the approach.^{1,53} Finally, indicative Δm fingerprints could provide constraints to putative compound group annotations derived from molecular formula data alone (van Krevelen diagrams) or allow for a more precise annotation.^{54–56}

We hypothesized that DOM from swamps and topsoil, in close contact with plant inputs and active microbial communities, would reflect recognizable plant-related source imprints that can be revealed by tandem mass spectrometry. Specifically, we explored links between precursor ion Δm matching profiles and precursor ion characteristics such as nominal mass, mass defect, initial ion abundance, fragmentation sensitivity, oxygen-to hydrogen ratio (O/C), heteroatom content, and structure suggestions. These properties are in part predictable from the assigned molecular formula, and thus allow for an evaluation of the approach (“proof of concept”) while also revealing potential nonassigned molecules (e.g., P-, Cl-, Br-, I-, and F-containing molecular formulas). Finally, we hypothesized that indicative Δm features of plant phenols, e.g., lignin- and tannin-related losses, would match their yet unknown structural analogues in DOM and that these patterns would reflect commonly applied “structural domain” distributions.^{55,57,58}

2. EXPERIMENTAL SECTION

A detailed experimental procedure is provided in the Supporting Information of this article (Note S-1). In short, we chose 14 aromatic reference compounds as representative plant metabolites (Figure S-1 and Table S-1) and a forest topsoil pore water isolate⁵⁹ and Suwannee River Natural Organic Matter (SRNOM)⁶⁰ as exemplary DOM samples. All reference and sample solutions were directly infused into the ESI (electrospray) source of an Orbitrap Elite (Thermo Fisher Scientific, Bremen) at negative ionization mode (Table S-2) and fragmented by collision-induced dissociation (CID, MS²). We chose four nominal masses within the mass range typically observed in terrestrial DOM samples (m/z 200–500) for fragmentation (m/z 241, 301, 361, and 417, herein referred to as isolated precursor ion mixtures, “IPIMs”) to test the approach.⁶¹ Soil DOM was analyzed at three normalized collision energy (NCE) levels (15, 20, and 25%). MS³ spectra of selected key product ions (aglycons of flavonoids and demethylated dimethoxy-methyl-benzoquinone) were acquired at NCE 20 or 25. After recalibration with known (Table S-3) or predicted product ions (losses of CO₂, H₂O, etc.), all major product ions were annotated with a molecular formula in reference compounds (Figure S-2, Tables S-4, and Table S-5) and DOM. Formula annotation was conducted with a Matlab routine recently incorporated into an open FTMS data processing pipeline.⁶²

For MS² data analysis, we generated Δm matrices of every pairwise combination of precursor and product ions (“ Δm fingerprints”). Every value in this matrix is referred to as a Δm feature or Δm . We compared the unknown Δm features in DOM to three lists of known Δm features:

- 54 Δm features ubiquitously found in DOM (Table S-6),
- 55 Δm features from the set of 14 reference compounds (Table S-7), and
- 11477 Δm features from a negative ESI MS² library with 249 916 reference spectra of 17 994 unique molecular structures annotated by SIRIUS⁶³ (Figure S-3; list in

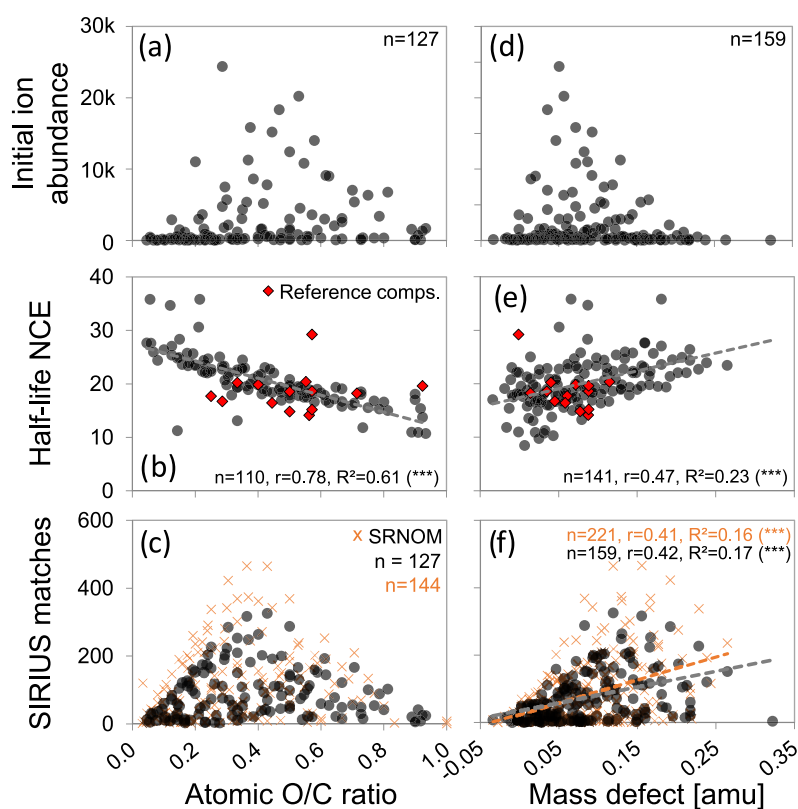


Figure 1. Links between selected DOM precursor ion properties (top panels, initial ion abundance at NCE 0; middle panels, half-life normalized collision energy (NCE) at which ion abundance has dropped by 50%; bottom panels, matches of delta masses (Δm 's) of measured precursor and product ion masses (delta masses, Δm) with a list of 11477 known Δm features from SIRIUS) and each precursor ions' O/C ratio (a–c) or mass defect (d–f). O/C ratios can only be shown for precursor ions with an annotated molecular formula. Black dots are individual soil DOM formulae. Additional data from reference compounds (red diamonds, see also Figure S-4) and SRNOM (orange crosses) is shown in middle and bottom panels, respectively. Statistical data were derived from linear fits; asterisks (***) denote p -value < 0.001.

supporting datasets). Reference spectra were collected from GNPS, MassBank, MoNA, and NIST.^{64,65}

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

We assessed the probability of false-positive matches and accounted for the number of elements in the formula, ion abundance, and measures of fragmentation sensitivity to validate our approach. The matching data were combined for each NCE level and transformed into a binary format. We classified Δm matching profiles of DOM precursor ions and reference compounds of lists (b) and (c) by two-way hierarchical clustering using Ward's method and Euclidean distance, as well as Principal Components Analysis (PCA) in

PAST (v3.10) for list b.⁶⁷ We visualized numbers of individual Δm matches and Δm cluster matches in van Krevelen space for all lists. We chose the 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-4).^{57,58,68–70} In two separate analyses, formulas were also classified with a more general and a data-based van Krevelen scheme besides the reference one.^{58,71}

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 solely 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 reference compounds (Figures S-1, S-2, and S-4) yielded 42 unique Δm features (i.e., not covered in list a, Table S-6) but also eight

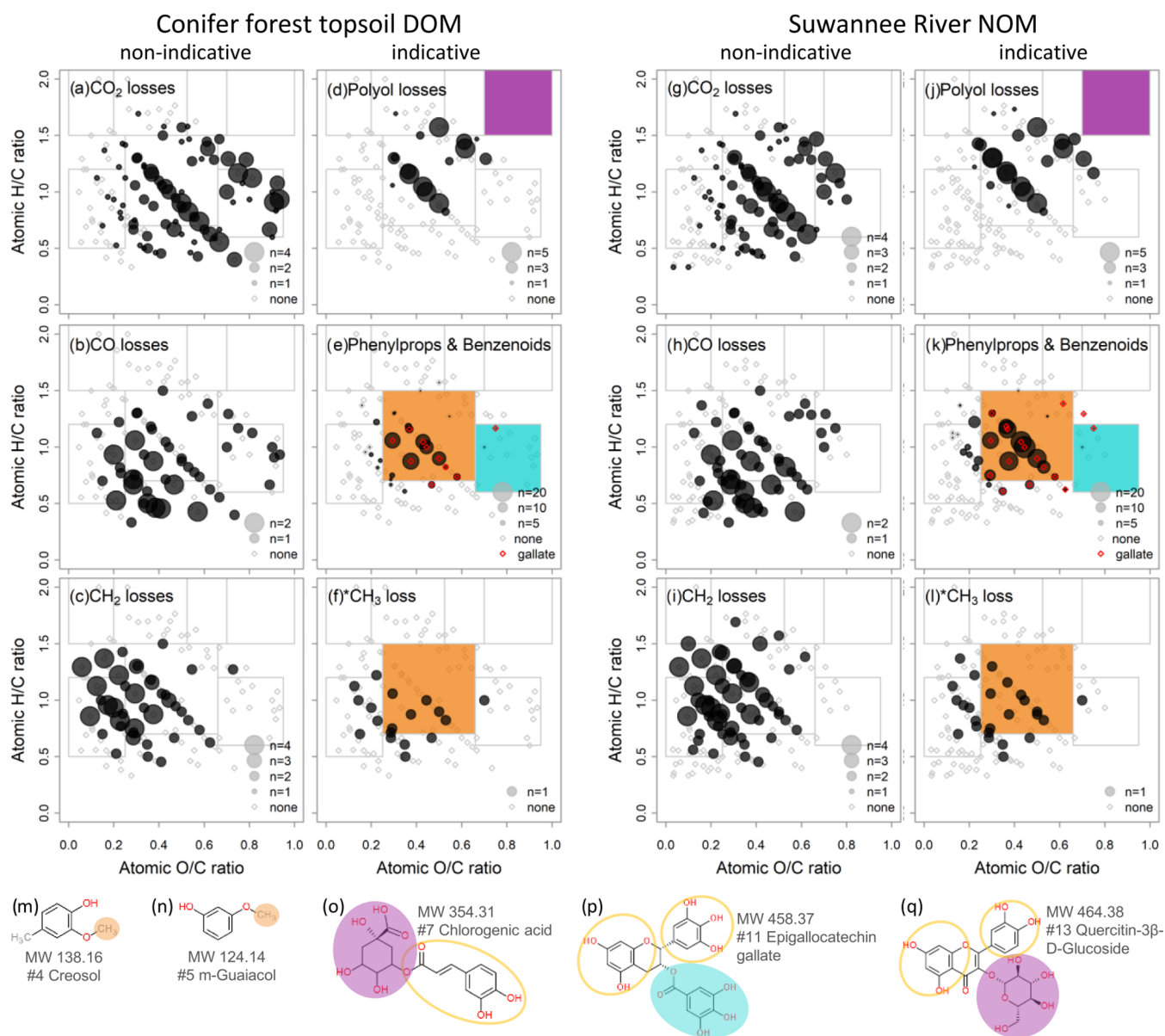


Figure 2. Δm matches visualized in chemical space for soil (porewater) DOM (a–f) and SRNOM (g–l). Exemplary reference compound structures with potential indicative Δm units are shown in bottom panels (m–q). “Phenylprops” refers to the shortened Classyfire class of “Phenylpropanoids and polyketides”. Gray outlined boxes refer to anticipated structural domains (Figure S-4).⁶⁴ (a–l) Precursor ions with an annotated molecular formula by their atomic H/C and O/C ratios (van Krevelen plot; soil DOM, $n = 127$; SRNOM, $n = 144$). Dot size encodes the number of matches to nonindicative (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 be expected to yield the respective Δm , and colors refer to the structural motifs marked in (m–q). Phenylpropanoid or benzenoid-like (sub-)structures as the ones shown in empty circles (o, p, q) may also contain methyl or methoxy groups (filled orange dots in m, n) that could produce methyl radical losses. Calculations based on Δm data are presented in more detail in Table S-13. Highlighted red open diamonds in (e) and (k) indicate loss of up to three gallic acid equivalents (size not drawn to scale).

that were described in DOM. These eight Δm features (namely: H_2O , 18.0106; CO , 27.9949; C_2H_4 , 28.0313; $\text{C}_2\text{H}_2\text{O}$, 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 list (b) to compare DOM and reference compounds (Table S-7). Besides precursor ion formulas #2 (Hydroxy-cinnamic acid, or p-coumaric acid; $\text{C}_9\text{H}_8\text{O}_3$, 164.0473), #3 (Gallic acid; $\text{C}_7\text{H}_6\text{O}_5$, 170.0215), and #5 (m-Guaiacol; $\text{C}_7\text{H}_8\text{O}_2$, 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 list b, namely, precursor ions #1 (Vanillic acid; $\text{C}_8\text{H}_8\text{O}_4$, 168.04226), #4 (Creosol, $\text{C}_8\text{H}_{10}\text{O}_2$, 138.0681), #8

(Ellagic acid; $\text{C}_{14}\text{H}_6\text{O}_8$, 302.0063), and #10 (Catechin; $\text{C}_{15}\text{H}_{14}\text{O}_6$, 290.0790), and the neutral aglycon of compounds #12 and #13 (flavonol core of Spiraeoside and Isoquercitin; $\text{C}_{15}\text{H}_{10}\text{O}_7$, 302.0427). More details on reference compound fragmentation are given in the SI (Note S-2).

3.2. Fragmentation Behavior of Soil DOM. DOM precursor ions were isolated and fragmented to obtain Δm data via matching (Figure S-5). 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 and Table S-8). Highest numbers of product ions were found at the highest NCE (Figure S-6).

Table 1. Summary of Two-Way Clustering of DOM Precursors (Highlighted in Red) and 14 Reference Compounds (Highlighted in Green)

Cl. #	n	A	#6 Ellagic acid	#12* Spiraeoside	#13* Isoquercetin	B	#2 p-Coumaric acid	#3 Gallic acid	#7 Chlorogenic acid	#9 epDlGG	#11 EGCG	#12 Spiraeoside	#13 Isoquercetin	#14 Myricetin	C	#1 Vanillic acid	#4 Cressol	#5 m-Guaiacol	#6 diMeOMeBQ	D	E	F	#10 Catechin	G	H	Assigned cluster name
1	8	25	38	0	13	4	0	0	0	0	0	0	13	0	1	13	0	0	0	9	98	64	63	94	39	Combinations of ubiquitous losses (H ₂ O, CO)
2	5	0	0	0	0	24	0	0	20	0	0	20	40	60	0	0	0	0	0	0	93	25	0	10	77	Polyol-equivalent losses
3	12	5	0	25	25	3	0	0	8	8	8	0	0	8	2	8	0	0	0	40	68	83	67	16	15	Phenylpropanoids and Benzenoids
4	10	6	0	20	30	3	0	0	10	30	50	0	0	0	0	0	0	0	0	8	91	40	20	25	8	Phenols (i.e., Benzenoids)
5	12	1	8	0	0	3	8	0	0	17	17	8	1	17	0	0	0	0	0	1	7	3	8	2	1	none
6	5	78	60	60	100	39	40	20	20	0	20	0	0	20	35	40	20	0	0	71	100	98	80	100	90	Carboxylic acids (ubiquitous losses, CO ₂)
7	3	0	0	0	0	7	33	0	0	0	0	0	0	0	78	33	100	67	33	81	92	83	33	53	36	Benzenoids (*CH ₃ , C ₂ H ₄)
Precursors	6					22									13					16	8	7		10	14	Total = 96
Soil DOM	5					9									4					7	2	4		5	8	44
SRNOM	1					13									9					9	6	3		5	6	52

Numbers refer to Figure S1; #12 and #13* refer to MS3 spectra of flavonoid aglycons. Numbers are coverage in Δm matches compared to overall Δm 's per Δm cluster; values > 20% are highlighted in bold, values < 10% are grayed out. Δm clusters are shown in rows ("Cl. #", 1–7) and precursor clusters in columns (A–H, for details, see Table S14 and original clustering data in PANGAEA datasets). Additional columns show respective numbers of Δm matches ("n") and assigned cluster name (compare Table S14). In the lower row, numbers of precursors per precursor cluster are given for both samples combined and individually. Few reference compounds clustered with precursor clusters D–H, which were dominated by DOM precursors with higher numbers of Δm matches. Compounds #7, #12, #13, and #14 contain polyol moieties; compounds #1, #4, #5, and #6 contain -methoxy and -methyl moieties (Figure S1)

Product ion spectra did not indicate abrupt structural changes upon increasing NCE, showing no separation of isomers/isobars but a continuous increase in fragmentation across all precursor ions. Based on the above results, NCE of 25 was chosen to fragment SRNOM for comparison.

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 CO₂ loss and was interpreted as a result of intrinsic averaging.^{31,45} In contrast, Dit Foque et al. described the 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 ion molecular formula, which also increases the probability to detect meaningful links between precursor and product ions. A minor group of oxygen-poor formulas was nonresponsive (Note S-3). Matching to list c 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, and S-12, and Figure S-7). DOM precursor ions with an average O/C ratio matched more often than low-O/C, fragmentation-resistant precursor ions (Figures 1c and S-8, Note S-3)^{18,19,35} or high-O/C, easily fragmented precursor ions (Figure 1b). These observations together show that fragmentation sensitivity and

Δm matching seem to be independent DOM precursor ion properties and that Δm matching could be driven by ion abundance. SRNOM and the soil water sample shared most molecular formulas ($n = 107$; 84% of soil DOM and 74% of SRNOM formulas) and accounted for most of precursor ion abundance at NCE 25 (96.5% and 97.2%, respectively). Despite this high similarity, SRNOM precursor ions 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 (Section 3.5).

3.3. Evaluation of the Δm Matching Approach. We used the matching data of molecular formulas in DOM for a proof-of-concept evaluation of our Δm matching approach. Specifically, we aimed to test the hypothesis that all precursor ions are potentially linked to all product ions in chimeric MS² spectra of ultracomplex DOM. Our analysis was congruent with previous observations, showing losses of common Δm 's (Table S-6) while also revealing more detail (Figure S-5c, Table S-7). Details are given in the Supporting Information (Note S-4). We found expected trends in losses of CO₂, CO, and CH₂ in both samples (Figure 2a–c, g–i and Table S-13). The predicted heteroatom content (O, N, S) of assigned molecular formulas and a widened tolerance window were used for further analysis of the uncovered structural information. Random Δm matching would be expected if the calculated Δm values were affected by low resolution, low sensitivity, or artifacts such as reactions in the instrument (e.g., between the collision and Orbitrap cell⁷⁸). Instead, we found that (1) precursor ions with low ion abundance matched to less Δm features (Figure S-7), (2) nonfragmented precursor ions matched to less or no Δm 's (Figure S-8), and (3) identity of Δm matches agreed with molecular formula prediction (e.g., loss of S-containing Δm 's from S-containing precursor ions, etc.; Figures S-9 and S-10). Our evaluation also shows that Δm matching not only helps in recalibration⁷⁹ but also serves to check formula annotation, as it revealed unresolved precursor

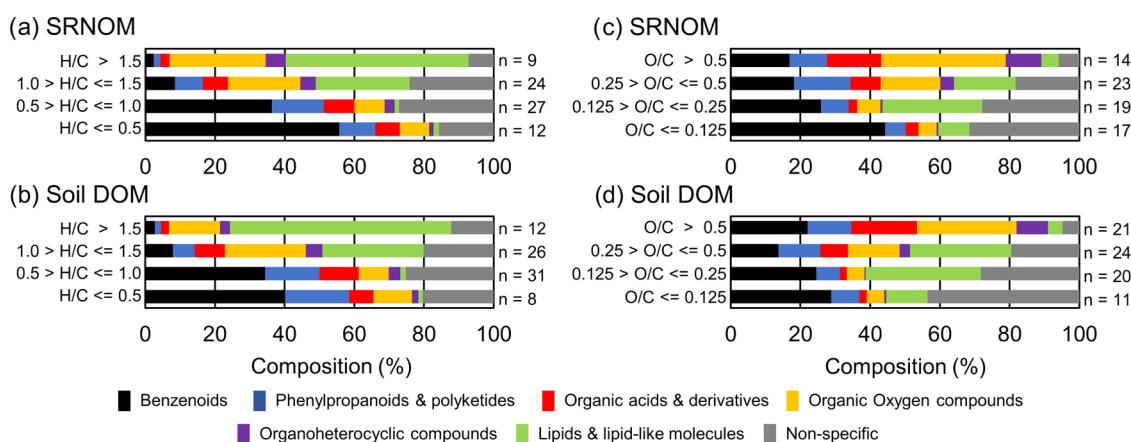


Figure 3. Potential structural composition of exemplary sections in van Krevelen space shown in the H/C (a, b) and O/C (c, d) dimensions. Data from both DOM samples and for CHO precursors are presented. Each bar represents 100% of information derived from matched Δm features (list c) per precursor ion; compositions of individual precursors are aggregated as averages here for visualization (numbers of precursors given behind bar). Examples of individual precursor compositions are shown in the Supporting Information (Figure S-12). The analysis is based on the association of Δm features (list c) with their SIRIUS-annotated molecular (host) structures (details given in Note S-1). Structural classes are shown at the most aggregated class level of the Classyfire Ontology here (compare innermost circle in Figure S-3) but would allow for finer differentiation in future applications.

ion compositions interfering especially with CHOS precursor ions (related to Cl, P, and F). This means that (1) these atoms should be included for better coverage of elemental composition (i.e., prioritization) in our specific sample context and that (2) higher resolution power may be required to resolve S-, Cl-, P-, or F-containing precursor ion compositions.¹ In summary, Δm matching revealed an inherently structured biogeochemical signal of precursor ions that seem to fragment individually and was highly sensitive in detecting precursor-product ion pairs. This suggests that chimeric DOM data can be deconvoluted to reveal differences in molecular composition not visible from MS¹ inspection.^{23,80} It should be stressed that these results will need further evaluation due to the small number of DOM precursor ions, m/z values, and samples analyzed here (159 in soil DOM, 221 in SRNOM), and that deconvolution should be further tested with better-characterized mixtures, including, e.g., structural analogues, artificial mixtures, or standard additions (spiking).^{14,19,27,42,81}

3.4. Clustering with Reference Compound Δm 's Reflects Structural Trends. DOM precursor ions from both samples were compared based on Δm matching (Table S-7, see Section 3.1). We grouped DOM precursor ions, reference compounds, and Δm features (list b) by two-way hierarchical clustering (Table S-14), i.e., matching of precursor ions across Δm features and vice versa. In the following, precursor ion clusters will be referred to by letters (A–H) and Δm clusters by numbers (1–7; Table 1). Based on the specificity of SIRIUS Δm features (Table S-14), we defined five Δm clusters found herein as structure-specific (Table 1, Figure 2d,e,j,k; and Table S-13).

Δm features $C_4H_8O_4$ (120.0423 Da, tetrose equivalent) and $C_6H_{10}O_5$ (162.0528 Da, hexose equivalent), both members of cluster 2, were annotated to alcohols and polyols, carbohydrates, carbohydrate conjugates, and ether structures via SIRIUS (Table S-14). Reference compounds containing a polyol (quinic acid, #7) or a sugar (glucose, #12 and #13; mannose, #14) contributed Δm 's to this cluster (Table 1).^{51,52} Cluster 2 Δm 's matched to 18 soil DOM and 24 SRNOM precursor ions in the central van Krevelen plot despite the absence of “carbohydrate-like” precursor ions (lilac square,

Figure 2d, j and o, q). The anticipated shift toward higher O/C and H/C ratios was nonetheless apparent (Figure 2e,f and k, l).

Cluster 3 and 4 Δm features, partly specific to phenylpropanoid and benzenoid structures, were contributed by flavan-3-ols (#10, #11), flavon-3-ol aglycons (#12 and #13), and compounds containing cinnamic, coumaric, or gallic acid (#7, #9, #11).^{28,33,52} Precursor ions that matched to clusters 3 and 4 (soil DOM: $n = 27$ and $n = 12$; SRNOM: $n = 29$, $n = 21$) were found in the “lignin-like domain” (orange square in 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 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 requires further testing with model mixtures. Aliphatic chains could prevail as O-poor substructures in substituted cyclic core structures.^{82,83} Similar to the detection of polyol-equivalent Δm matches outside the expected “carbohydrate domain”, gallate-equivalent losses were not matched to precursor ions in the anticipated “tannic domain” but to precursor ions outside of that box (red diamonds, turquoise square, Figure 2e, k; turquoise circle, panel p).

Among the most prominent features was the methyl radical loss,^{35,49,50} which matched to oxygen-poor DOM precursor ions and was one of three cluster 7 Δm features (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 precursor ions 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 ($r = 0.55$, $R^2 = 0.30$, $n = 127$, $p < 0.001$; $r = 0.58$, $R^2 = 0.34$, $n = 144$, $p < 0.001$), implying structural similarities (Figure 2f,l), e.g., condensed structures with aliphatic, lactone, or quinone moieties.³⁴ CO and CH_3^\bullet were also annotated to benzenoid structures via SIRIUS (Table S-14). The methyl radical loss is an expected diagnostic Δm of methoxylated aromatic rings as in lignin (orange square in Figure 2f, l; orange circles in panels m, n; see Note S-5), but was also matched to DOM precursor ions 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 which describe regular 0.0364 Da increments in DOM that are formally annotated by

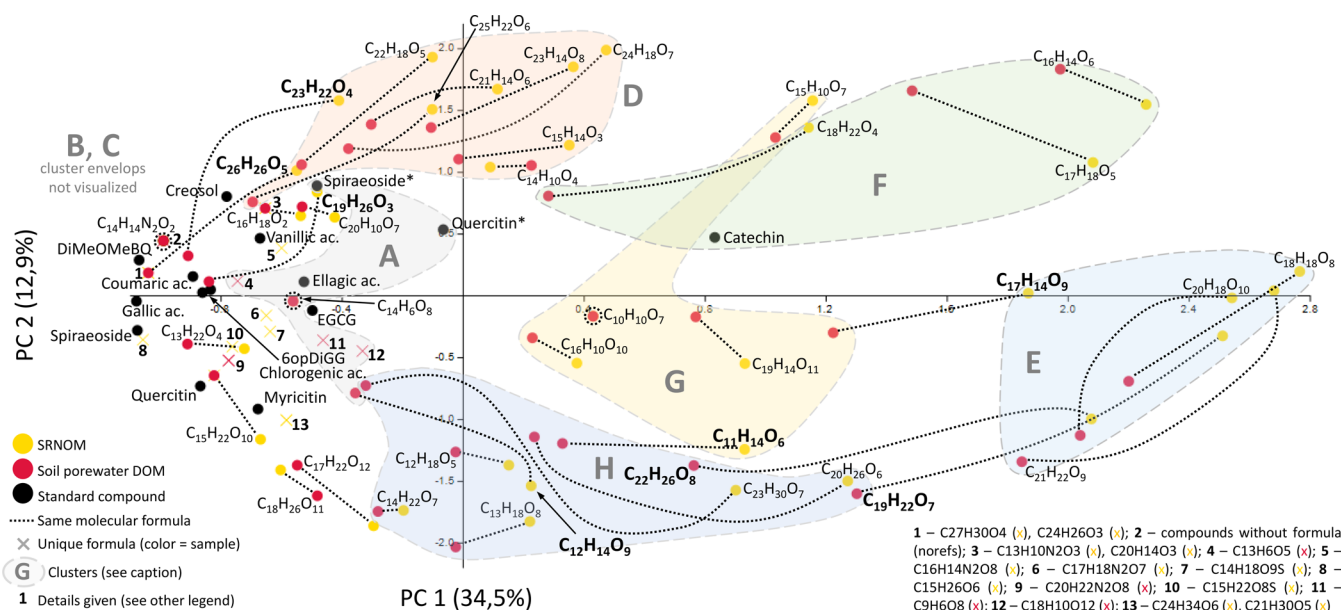


Figure 4. Separation of DOM precursor ions based on Δm matching with list b. Principal component analysis of all precursor ions with more than one match to indicative Δm features of the 14 reference compounds (i.e., Δm features shown in Table S-7 that are not part of Table S-6, see Section 3.1). Colors of dots distinguish precursor ions from both samples and reference compounds (see legend). Precursor ions detected in both samples are connected by dotted black lines. Precursor ion clusters (A–H) are marked by envelopes and letters (compare Tables 1 and S-14). Eight shared precursor ions that switched precursor ion clusters are highlighted by bold molecular formula (C₁₂H₁₄O₉, A in soil DOM → H in SRNOM; C₁₉H₂₆O₃, B → C; C₂₆H₂₆O₅ and C₂₃H₂₂O₄, B → D; C₁₇H₁₄O₉, G → E; C₁₉H₂₂O₇ and C₂₂H₂₆O₈, H → E; C₁₁H₁₄O₆, H → G).

an exchange of CH₄ by O (Figure S-11).^{37,38} Concurrent losses of CO and C₂H₄ explained the presence of these increments on the product ion level and were paralleled by losses of CH₃[•]. This finding could explain the ubiquitous presence of CH₄ vs O series in nonfragmented DOM; for example, concurrent β -oxidation and de-carboxylation could be enzymatic MS¹ analogues of the patterns seen in MS².²⁶ Alternatively, both methoxy-phenols (#4, #5) indicated an insertion of O for CH₄ upon fragmentation (Note S-2, first section). Both observations may be relevant for the explanation of a key DOM property and require further exploration.

Taken together, matching to Δm features derived from a small set of reference compounds revealed emerging clusters of precursor ion 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 indicated clear overlap, which means that the same precursor ion was part of more than one Δm -predicted structural class. For example, 27 classic lignin-like precursor ions were part of seven precursor ion clusters (B–H; Table S-15) with clear differences in potential structural composition. An extended analysis using >700 compound class-associated SIRIUS Δm features (list c) substantiated these findings (Figures 3, S-12 and Note S-1).

CHNO and CHOS precursor ions matched with many of the S- and N-containing SIRIUS Δm features (list c, spanning 3–78 S- and 4–251 N-containing Δm 's in soil DOM and 0–154/0–350 in SRNOM; Tables S-16, S-17, S-18, and S-19). These represented on average 79 ± 19% (63 ± 31% in SRNOM) of all Δm matches per CHOS precursor ion or 91 ± 7% (79 ± 28%) of all CHNO precursor ion matches (Note S-6). CHNO precursor ions were annotated with reduced forms of N (including aralkylamines, amino acids, carboximidamides, and dicarboximides/urea-containing compounds, Table S-20) but not to nitrate esters.^{34,84,85} S-containing Δm matches

indicated the potential presence of sulfonic, thiol, thioether, or aromatic CHOS compounds.⁸⁶ These results show a wide potential diversity of N and S compounds in DOM that differs from earlier reports of mainly aromatic N and sulfonic S.^{34,87,88}

As most of these studies analyzed marine DOM, the detection of more diverse sets of CHOS and CHNO precursor ions could relate to the terrestrial, less degraded DOM analyzed here.^{16,89–91} Further tests with N- and S-containing reference compounds and DOM samples are warranted to reveal the hidden diversity of CHNO/CHOS compounds and confirm potential structures, e.g., by NMR.

All in all, our results show that it may be possible to refine molecular structure representations in van Krevelen plots by deconvoluted MS² data and that complementary precursor ion information could be used to assess false or biased Δm -based class assignments (e.g., elemental composition, DBE, ionization, fragmentation sensitivity, ion mobility, polarity index, etc.).^{13,55,58} Fluorescence or NMR spectroscopy could add valuable information if DOM would be fractionated before MS² data acquisition, i.e., to assess indirect (statistical) links of MS² features with complementary forms of structural insight.^{21,92–94} Our findings must however be taken with caution for four reasons:

- (1) SIRIUS Δm features (list c) 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 toward certain classes of compounds (Figure S-3), as our set of 14 aromatic compounds. Here, we only considered negative ESI mode data which is commonly employed for DOM analysis. Adding positive ESI or other ionizations would extend the range of Δm features and structural classes

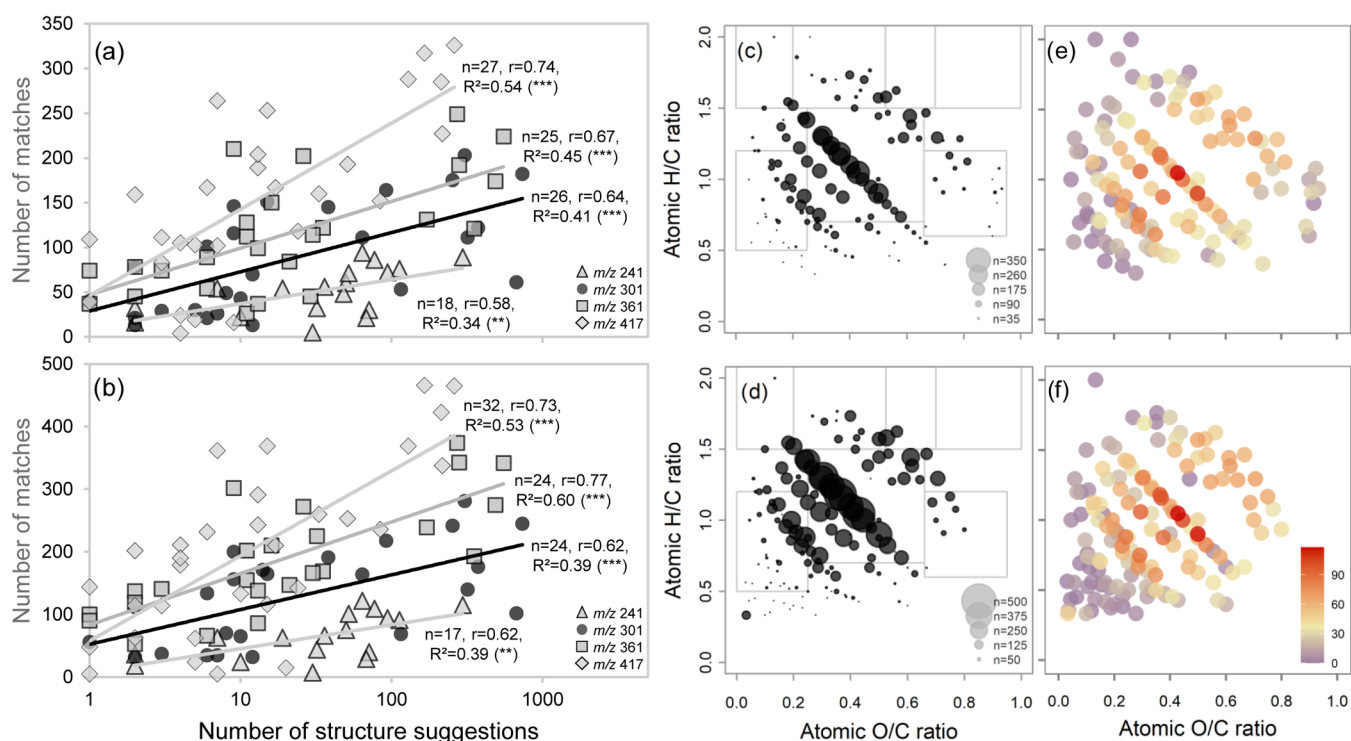


Figure 5. Agreement between chemodiversity estimates based on molecular formula (structure suggestions) and precursor-product ion links (Δm matches). (a, b) Correlations between numbers of SIRIUS Δm matches (list c) vs structure suggestions (note log scale, incl. in silico hits): (a) soil DOM, (b) SRNOM. (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; gray boxes refer to structural domains defined in Figure S-4. (e, f) Number of predicted classes per precursor ion based on SIRIUS Δm matches (color scale similar in both panels). Structural classes are associated with SIRIUS-annotated Δm features through correlation analysis of host structures and their Δm features (classification based on Classyfire): (e) soil DOM, (f) SRNOM.

covered and likely decrease bias.^{14,16,23,86} 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 pairs in future, potentially in combination with fragmentation energy gradients (fragmentation trees),⁹⁵ moving m/z isolation windows, or ion accumulation time variation.⁹⁶
- (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 structural class boundaries remained considerable (Figures 3 and S-12).

Data-dependent and data-independent acquisition (DDA, DIA) techniques could be used to cover the whole mass range of precursor ions in DOM mass spectra in future, and are widely employed in LC-MS of complex mixtures.^{16,27,97,98} For example, Ludwig et al. presented a DIA scheme (SWATH-MS) that employs one precursor ion scan and 32 isolation windows of 25 Da width, covering 800 Da within 3.3 s; similar schemes are likely transferable to acquire full mass range data of directly injected DOM.⁹⁹ Kurek et al. recently presented such data (m/z 392–408),¹⁶ Leyva et al. discerned fragmentation pathways and structural families (mass range m/z 261–477).¹⁴ The latter approach could be extended to include the diversity of structure-associated Δm features presented here. Together, this shows that practicable tandem MS acquisition strategies are in reach and will enable deeper analyses of Δm features in DOM soon.

3.5. Drivers of Differences in Δm Matching between Soil DOM and SRNOM.

Although matching among the two samples was largely consistent, slight differences were apparent in van Krevelen distributions (list b: Figure 2, list c: Figures 3 and S-12). We therefore tested the separation of precursor ion clusters by ordination (principal component analysis, Figure 4) using list b. Precursor ion clusters were clearly separated on Principal Components 1 and 2 which together held about 47% of variation. Most considered precursor ions were shared among samples (64%, 38 out of 59), only a small number was sample-specific (SRNOM = 14, Soil DOM = 7). Sample-specific precursor ions were found in clusters A (linked to carboxylic acids), B (phenols, polyols) and C (benzenoids, Table 1), the remaining clusters D–H were dominated by the shared precursor ions. Out of the 38 shared precursor ions, 30 (79%) grouped in the same precursor ion cluster despite a general trend to higher numbers of matches in SRNOM, but eight grouped differently (bold precursor ions in Figure 4). These differences in matching could be related to different chemistries, i.e., different isomeric/isobaric composition.⁸⁴ For example, 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 could be partly linked to higher fragmentation resistance in SRNOM (Table S-21). Unfortunately, we only have data on initial ion abundance and fragmentation sensitivity from the soil DOM isolate; other precursor ion properties, however, showed very similar trends in both samples (Table S-21).

Similar clustering and Δm -predicted structural classes (Figure S-12) in shared precursor ions could indicate a conserved structural composition. Likewise, Kurek et al.

observed high similarity in photoionized (APPI) and IMPRD-fragmented DOM samples but observed clear differences in CHOS fragmentation.¹⁶ High similarities between DOM samples would be in line with stoichiometric principles (i.e., due to a large share in precursor ions between DOM samples) and could suggest that DOM processing diversifies, but also “randomizes” the molecular composition of each precursor ion (“universal” signal).^{31,100,101} High congruence of fragmentation patterns (and thus, Δm matching) among DOM precursor ions has also been interpreted as a sign of similarly substituted but slightly differing core structures.^{35,37} The clusters devised here were small due to the relatively small number of precursor ions and m/z values analyzed, and thus may not detect significant differences between samples yet. However, even with our small set of precursor ions, the clustering by Δm matching showed conserved differences in fragmentation between precursor ion clusters, and in part, even the same precursor ion in different samples. The fact that this could relate to differences in ion abundance (and therefore, possibly also ionization efficiency) or fragmentation sensitivity is intriguing and should be investigated across a wider range of DOM chemotypes using improved classification approaches as applied here (see also Section 3.4).¹⁴ In line with this, potential compositional differences between DOM samples became more apparent when more Δm features were used for the clustering (list c instead of list b; Figures 3 and S-12).

3.6. Ion Abundance Is Linked to Δm Matching Frequency and Structural Diversity. Ion abundance was the most important driver for Δm matching in both samples and highest in the van Krevelen plot “region” usually assigned to ubiquitous lignin structures or carboxyl-rich aliphatic molecules.^{59,83} This region also parallels with a maximum in potential underlying chemodiversity,^{30,102} which could explain why these signals are ubiquitously found and especially dominant in recycled DOM.^{90,103} Δm matching showed potential to reveal this underlying chemodiversity effect and was therefore compared to numbers of structure suggestions and Δm -predicted compound classes per precursor ion (Figure 5). Numbers of Δm matches were significantly and positively related to the number of structure suggestions in absolute terms and for specific compound classes (Table S-21). The correlation between Δm -predicted and suggested compound classes was surprisingly similar in both samples and significant for almost all benzenoid-type (benzopyrans, methoxybenzenes, anisoles, phenols, etc.) and most phenylpropanoid-type structures (flavonoids, linear 1,3-diarylpropanoids). Among the organic acids, only vinylogous acids stood out (i.e., containing carboxylic acid groups with insertions of C=C bond(s)). Significant correlations were also found for pyrans, acryloyl compounds, carbohydrates, aryl ketones, and alkyl aryl ethers (fatty acids and analogues only in 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 classes showed strong overlap in this study. It also provides additional support to our assumption that all precursor ions 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 phenyl-

propanoids) showed significant correlations could point to bias toward plant natural products in the databases employed here; in turn, this means that the inclusion of other structure databases and the additional assignment of Δm 's not only to their host structures but also to host organisms (e.g., in GNPS⁶⁵) could reveal further clues about the potential sources of molecular formulas in DOM.

We propose that the number of Δm matches could be interpreted as a novel, relatively easily accessible measure to account for a precursor ions' underlying potential structural diversity. Such information could help to better understand the mechanisms of DOM formation and persistence in the environment. Our results encourage further studies on the Δm matching behavior of synthetic mixtures of known structures and across DOM chemotypes, and the improved bioinformatic exploitation of chimeric (LC-) FTMSⁿ data of complex organic mixtures.^{14,104–106} We acknowledge that natural product and in silico databases are far from being complete, same as the database of annotated Δm matches we used here, despite its large coverage of $\sim 18\,000$ unique structures and $\sim 11\,500$ Δm 's (Figure S-3). For example, precursor ions with low mass defects showed exceptionally few structural hits, indicating bias in natural product databases (Figure S-13).¹⁸ These structures were easily fragmented and yielded few Δm matches in our analysis. CHO precursor ions were double as likely to yield a suggestion than N- and S-containing precursor ions. These observations show that DOM contains unique molecular structures to be identified in future, potentially through the application of a wider range of ionization and fragmentation techniques that reduce structural bias.^{14,16,23}

4. IMPLICATIONS

Tandem MS data of complex samples such as dissolved organic matter (DOM) is impeded by the co-fragmentation of precursor ions with similar nominal mass, and further complicated by the contribution of potential isomers and isobars. We employed an approach that analyzes the pairwise Δm 's between all precursor and product ions (Δm matrix). Using a very limited set of precursor ion features from two samples, we found potential signs of structural imprints related to e.g., benzenoids, phenylpropanoids, carbohydrates, sulfonic acids, thiols, thioethers, and amino acids. The successful matching of indicative Δm features and precursor ion clustering suggests a recognizable source imprint of primary or recycled plant remains in DOM. Tests with more DOM samples and artificial/treated mixtures (e.g., spiked DOM, or enzyme-degraded DOM) are required to test the assumptions employed here and to improve classifications by Δm clustering. Our first results indicate that FTMS² data may be useful to differentiate molecular composition on the molecular formula level and that ion abundance and fragmentation sensitivity are two key variables that explain differences in MS² data within and among samples. This is intriguing because a shared molecular formula could harbor a completely different set of structures, and larger sets of DOM data would improve the detection of these differences. Generally, our findings support the view that regions of the van Krevelen plot are associated with indicative Δm 's that relate to stoichiometric differences between compound classes. The most abundant precursor ions however showed a mixed MS² signal that caused boundary overlap of these “ Δm -defined regions” (Figures 3 and 5e, f). While this finding is in line with known patterns of structural

diversity and partly explains the ubiquitous presence of abundant DOM signals, it introduces a new paradigm to the interpretation of DOM FTMS data by assigning unknown precursor ions to *multiple* structural categories instead of just one (Figure S-12). Further evaluation of both natural and spiked/treated complex mixtures, constantly growing MS databases, and comprehensive decomplexation methods (LC-MS, IMS) will together provide fundamental insights into the deconvolution of chimeric spectra from complex samples, and ultimately show the potential to unfold the hidden molecular diversity and identity of DOM.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.2c01332>.

Table S-1: Information on reference compounds and solutions used in this study; Table S-2: Instrument settings for fragmentation experiments; Table S-3: Recalibration peaks used for reference compound in FTMS measurements; 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 analyzing matching patterns in van Krevelen space; Table S-14: Matching behavior of precursor ion clusters against Δm features (Table S-7); Table S-15: Lignin-like precursor ion formulas and their molecular properties and clustering; Table S-16: S-containing precursor ion formulas in soil porewater DOM; Table S-17: N-containing precursor ion formulas in soil porewater DOM; Table S-18: S-containing precursor ion formulas in SRNOM; Table S-19: N-containing precursor ion formulas in SRNOM; Table S-20: Structural class-correlated Δm features matched to CHOS or CHNO precursor ions; Table S-21: Correlations of selected precursor ion properties with PC scores (Figure 4); Table S-22: Correlations between structure hits and specific Δm features in CHO precursor ions; 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 compound classes in SIRIUS-annotated molecular structures; Figure S-4: Distribution of exemplary known structures in chemical space; Figure S-5: Orbitrap tandem MS of soil porewater DOM; Figure S-6: Comparison of matches to the two short Δm lists in relation to m/z and NCE; Figure S-7: Δm matches in relation to precursor ion abundance in soil DOM; Figure S-8: Δm matches in relation to precursor ion fragmentation sensitivity in soil DOM; Figure S-9:

Matching assessment with SIRIUS Δm 's (molecular formula check); Figure S-10: Changes in Δm matching frequency upon widening of tolerance window; Figure S-11: Link between matches to CH_3^+ , CO and C_2H_4 and CH_4 vs O exchange series; Figure S-12: Compositional variation of individual precursor ions in van Krevelen space; Figure S-13: Effect of mass defect on the number of structure suggestions; Note S-1: Supporting experimental details; Note S-2: Detailed description of reference compound fragmentation behavior; Note S-3: Behavior of nonresponsive 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; and Note S-6: Structural insight into N- and S-containing DOM precursor ions (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Carsten Simon – Molecular Biogeochemistry, Department of Biogeochemical Processes, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany; Present Address: Institute for Biogeochemistry and Pollutant Dynamics, ETH Zürich, Universitätsstrasse 16, 8092 Zürich, Switzerland; Present Address: Swiss Federal Institute of Aquatic Science and Technology (Eawag), Department Water Resources and Drinking Water, Überlandstrasse 133, 8600 Dübendorf, Switzerland; orcid.org/0000-0003-3882-3013; Email: carsten.simon@usys.ethz.ch

Gerd Gleixner – Molecular Biogeochemistry, Department of Biogeochemical Processes, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany; orcid.org/0000-0002-4616-0953; Email: gerd.gleixner@bgc-jena.mpg.de

Authors

Kai Dührkop – Chair for Bioinformatics, Faculty of Mathematics and Computer Science, Friedrich Schiller University Jena, 07743 Jena, Germany; orcid.org/0000-0002-9056-0540

Daniel Petras – Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093-0657, United States of America; CMFI Cluster of Excellence, Interfaculty Institute of Microbiology and Medicine, University of Tübingen, 72076 Tübingen, Germany; orcid.org/0000-0002-6561-3022

Vanessa-Nina Roth – Molecular Biogeochemistry, Department of Biogeochemical Processes, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany; Present Address: Thüringer Landesamt für Umwelt, Bergbau und Naturschutz (TLUBN), Göschwitzer Straße 41, 07745 Jena, Germany

Sebastian Böcker – Chair for Bioinformatics, Faculty of Mathematics and Computer Science, Friedrich Schiller University Jena, 07743 Jena, Germany; orcid.org/0000-0002-9304-8091

Pieter C. Dorrestein – Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093-0657, United States of America; orcid.org/0000-0002-3003-1030

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.est.2c01332>

Author Contributions

C.S. performed the measurements. D.P., V.N.R., P.D., and G.G. were involved in planning and supervising the work. K.D. and S.B. compiled global Δm feature data, analyzed its specificity, and performed structural classifications of Δm host structures as well as structure suggestions of DOM precursor ions. C.S. 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.

Funding

Open access funded by Max Planck Society.

Notes

The authors declare no competing financial interest. Supplementary data is available online (<https://doi.org/10.1594/PANGAEA.944673>). All tandem MS data can be found on the Mass Spectrometry Interactive Virtual Environment (MassIVE) under the following links: [ftp://massive.ucsd.edu/MSV000087117/\(soil DOM data\)](ftp://massive.ucsd.edu/MSV000087117/(soil%20DOM%20data)), [ftp://massive.ucsd.edu/MSV000088869/\(SRNOM data\)](ftp://massive.ucsd.edu/MSV000088869/(SRNOM%20data)), and [ftp://massive.ucsd.edu/MSV000087133/\(reference compound data\)](ftp://massive.ucsd.edu/MSV000087133/(reference%20compound%20data)) (Data Set S-1, raw peak data, *.mzML files). All other data associated with this manuscript (extensive tables, Δm feature lists, Δm specificity, two-way clustering table, processed data used to create figures, etc.) are available online free of charge from the PANGAEA Data Publisher under the following citation: Simon, Carsten; Dührkop, Kai; Petras, Daniel; Roth, Vanessa-Nina; Böcker, Sebastian; Dorrestein, Pieter C; Gleixner, Gerd (2022): Structural insight into conifer forest topsoil and blackwater dissolved organic matter by Orbitrap tandem MS mass difference matching. PANGAEA, <https://doi.pangaea.de/10.1594/PANGAEA.944673> (Data Set S-2, processed data as *.xlsx files).

ACKNOWLEDGMENTS

The authors thank all members of the Dorrestein lab for helpful discussions and insights. They also acknowledge Vivian Stefanow for initial structure database surveys, 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, and the financial support from the Max-Planck-Gesellschaft (MPG) and the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) 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., K.D. and S.B. are supported by Deutsche Forschungsgemeinschaft (BO 1910/20 and BO 1910/23). 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.
- (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.
- (3) Krueve, A. Strategies for Drawing Quantitative Conclusions from Nontargeted Liquid Chromatography–High-Resolution Mass Spectrometry Analysis. *Anal. Chem.* **2020**, *92*, 4691–4699.
- (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.
- (5) Zsolnay, Á. Dissolved Organic Matter: Artefacts, Definitions, and Functions. *Geoderma* **2003**, *113*, 187–209.
- (6) Wells, M. J. M.; Stretz, H. A. Supramolecular Architectures of Natural Organic Matter. *Sci. Total Environ.* **2019**, *671*, 1125–1133.
- (7) Prescott, C. E.; Grayston, S. J.; Helmisari, 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.
- (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.
- (9) Lange, M.; Roth, V. N.; Eisenhauer, N.; Roscher, C.; Dittmar, T.; Fischer-Bedtke, C.; González Macé, O.; Hildebrandt, A.; Milcu, A.; Mommer, L.; Oram, N. J.; Ravenek, J.; Scheu, S.; Schmid, B.; Strecker, T.; Wagg, C.; Weigelt, A.; Gleixner, G. Plant Diversity Enhances Production and Downward Transport of Biodegradable Dissolved Organic Matter. *J. Ecol.* **2020**, *109*, 1284–1297.
- (10) Lehmann, J.; Hansel, C. M.; Kaiser, C.; Kleber, M.; Maher, K.; Manzoni, S.; Nunan, N.; Reichstein, M.; Schimel, J. P.; Torn, M. S.; Wieder, W. R.; Kögel-Knabner, I. Persistence of Soil Organic Carbon Caused by Functional Complexity. *Nat. Geosci.* **2020**, *13*, 529–534.
- (11) Binemann, E. K.; Bongiorno, G.; Bai, Z.; Creamer, R. E.; De Deyn, G.; de Goede, R.; Fleskens, L.; Geissen, V.; Kuyper, T. W.; Mäder, P.; Pulleman, M.; Sukkel, W.; van Groenigen, J. W.; Brussaard, L. Soil Quality – A Critical Review. *Soil Biol. Biochem.* **2018**, *120*, 105–125.
- (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.
- (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.
- (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.
- (15) Wegley Kelly, 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.; et al. Distinguishing the Molecular Diversity, Nutrient Content, and Energetic Potential of Exometabolomes Produced by Macroalgae and Reef-Building Corals. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2110283119.
- (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.

- (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*, No. e1602976.
- (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.
- (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*, 21–30.
- (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.
- (21) Murphy, K. R.; Timko, S. A.; Gonsior, M.; Powers, L. C.; Wunsch, U. J.; Stedmon, C. A. Photochemistry Illuminates Ubiquitous Organic Matter Fluorescence Spectra. *Environ. Sci. Technol.* **2018**, *52*, 11243–11250.
- (22) Benner, R.; Amon, R. M. W. The Size-Reactivity Continuum of Major Bioelements in the Ocean. *Ann. Rev. Mar. Sci.* **2015**, *7*, 185–205.
- (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**, *48*, 213–229.
- (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.
- (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**, *57*, 280–294.
- (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.; Dorrestein, P. C. Non-Targeted Tandem Mass Spectrometry Enables the Visualization of Organic Matter Chemotype Shifts in Coastal Seawater. *Chemosphere* **2021**, *271*, No. 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*, 1461–1471.
- (29) Jeanne Dit Fouque, D.; 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*, 10821–10825.
- (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*, No. 405.
- (31) Zark, M.; Dittmar, T. Universal Molecular Structures in Natural Dissolved Organic Matter. *Nat. Commun.* **2018**, *9*, No. 3178.
- (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*, 7182–7191.
- (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*, 2688–2694.
- (34) Reemtsma, T.; These, A.; Linscheid, M.; Leenheer, J.; Spitzzy, 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.
- (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*, 8194–8202.
- (36) Cortés-Franisco, N.; Caixach, J. Fragmentation Studies for the Structural Characterization of Marine Dissolved Organic Matter. *Anal. Bioanal. Chem.* **2015**, *407*, 2455–2462.
- (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*, 1777–1786.
- (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 Individual Mass Measurements. *Fr. Anal. Chem.* **2003**, *75*, 1275–1284.
- (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*, 1061–1072.
- (40) Perdue, E. M.; Hertkorn, N.; Ketrup, A. Substitution Patterns in Aromatic Rings by Increment Analysis. Model Development and Application to Natural Organic Matter. *Anal. Chem.* **2007**, *79*, 1010–1021.
- (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*, 10106–10115.
- (42) Zhrebker, 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*, 4708–4719.
- (43) Bell, N. G. A.; Michalchuk, A. A. L.; Blackburn, J. W. T.; Graham, M. C.; Uhrin, D. Isotope-Filtered 4D NMR Spectroscopy for Structure Determination of Humic Substances. *Angew. Chem., Int. Ed.* **2015**, *54*, 8382–8385.
- (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.
- (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.
- (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.
- (47) Kostyukevich, Y.; Kononikhin, A.; Zhrebker, A.; Popov, I.; Perminova, I.; Nikolaev, E. Enumeration of Non-Labile Oxygen Atoms in Dissolved Organic Matter by Use of $^{16}\text{O}/^{18}\text{O}$ Exchange and Fourier Transform Ion-Cyclotron Resonance Mass Spectrometry. *Anal. Bioanal. Chem.* **2014**, *406*, 6655–6664.
- (48) Zhrebker, 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.

- (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.
- (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.
- (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*, 707–715.
- (52) Engström, M. T.; Päljälä, 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*, 4068–4079.
- (53) 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.
- (54) Reemtsma, T. The Carbon versus Mass Diagram to Visualize and Exploit FTICR-MS Data of Natural Organic Matter. *J. Mass Spectrom.* **2010**, *45*, 382–390.
- (55) 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.
- (56) 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. Process. Impacts*, **2014**, *16*, 2064. *Environ. Sci.: Processes Impacts* **2015**, *17*, 495–496.
- (57) 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.
- (58) Laszakovits, J. R.; MacKay, A. A. Data-Based Chemical Class Regions for Van Krevelen Diagrams. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 198–202.
- (59) Roth, V.-N.; Dittmar, T.; Gaupp, R.; Gleixner, G. Ecosystem-Specific Composition of Dissolved Organic Matter. *Vadose Zone J.* **2014**, *13* DOI: 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.* **2015**, *32*, 38–44.
- (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.
- (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.; Rossel, P. E.; Seidel, M.; Singer, G.; Stubbins, A.; Waska, H.; Dittmar, T. ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Anal. Chem.* **2020**, *92*, 6832–6838.
- (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.
- (64) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714.
- (65) Gauglitz, J.M., West, K.A., Bittremieux, W. et al. Enhancing untargeted metabolomics using metadata-based source annotation. *Nat Biotechnol* **2022**, DOI: 10.1038/s41587-022-01368-1
- (66) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminform.* **2016**, *8*, No. 61.
- (67) Hammer, Ø.; Harper, D. A.; Ryan, P. D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2001**, *4*, 1–9.
- (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.
- (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.; Eder, E. K.; Purvine, S. O.; Kolka, R. K.; Sebestyen, S. D.; Griffiths, N. A.; Schadt, C. W.; Keller, J. K.; Bridgman, S. D.; Chanton, J. P.; Kostka, J. E. 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.
- (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.; Lutz, A.; Yu, D.; Yi, Q.; Saha, N.; Huang, L. 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.
- (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.; Gavard, R.; Gleixner, G.; Hatcher, P. G.; He, C.; Hess, N. J.; Hutchins, R. H. S.; Ijaz, A.; Jones, H. E.; Kew, W.; Khaksari, M.; Lozano, D. C. P.; Lv, J.; Mazzoleni, L.; Noriega-Ortega, B.; Osterholz, H.; Radoman, N.; Remucal, C. K.; Schmitt, N. D.; Schum, S.; Shi, Q.; Simon, C.; Singer, G.; Sleighter, R. S.; Stubbins, A.; Thomas, M. J.; Tolic, N.; Zhang, S.; Zito, P.; Podgorski, D. C. 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**, 601–622.
- (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*, No. e7.
- (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.* **2020**, *48*, D445–D453.
- (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*, W423–W426.
- (76) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatbadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie,

- T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolite Database. *Nucleic Acids Res.* **2007**, *35*, D521–D526.
- (77) Jeffries, 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*, No. 44.
- (78) Baumeister, T. U. H.; Ueberschar, N.; Pohnert, G. Gas-Phase Chemistry in the GC Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2018**, DOI: 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**, *91*, 3350–3358.
- (80) Adair, E.; Alfonso, C.; Bell, N. G. A.; Davies, A. N.; Delsuc, M.-A.; Godfrey, R.; Goodacre, R.; Hawkes, J. A.; Hertkorn, N.; Jones, D.; Lameiras, P.; Le Guennec, A.; Lubben, A.; Nilsson, M.; Paša-Tolić, L.; Richards, J.; Rodgers, R. P.; Rüger, C. P.; Schmitt-Kopplin, P.; Schoenmakers, P. J.; Sidebottom, P.; Staerk, D.; Summerfield, S.; Uhrin, D.; van Delft, P.; van der Hooft, J. J. J.; van Zelst, F. H. M.; Zherebker, A. High Resolution Techniques: General Discussion. *Faraday Discuss.* **2019**, *218*, 247–267.
- (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*, 316–326.
- (82) Lam, B.; Baer, A.; Alae, 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.
- (83) Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Ketttrup, A.; Hedges, J. I. Characterization of a Major Refractory Component of Marine Dissolved Organic Matter. *Geochim. Cosmochim. Acta* **2006**, *70*, 2990–3010.
- (84) Lemr, K.; Holčápek, 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.
- (85) Piraud, M.; Vianey-Saban, C.; Petritis, K.; Elfakir, C.; Steghens, J. P.; Morla, A.; Bouchu, D. ESI-MS/MS Analysis of Underivatized 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.
- (86) 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.
- (87) Pohlbeln, A. M.; Dittmar, T. Novel Insights into the Molecular Structure of Non-Volatile Marine Dissolved Organic Sulfur. *Mar. Chem.* **2015**, *168*, 86–94.
- (88) 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.
- (89) 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.
- (90) 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.; Weigelt, A.; Dittmar, T.; Gleixner, G. Persistence of Dissolved Organic Matter Explained by Molecular Changes during Its Passage through Soil. *Nat. Geosci.* **2019**, *12*, 755–761.
- (91) Warren, C. R. High Diversity of Small Organic N Observed in Soil Water. *Soil Biol. Biochem.* **2013**, *57*, 444–450.
- (92) 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*, 3880–3886.
- (93) 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*, 1583–1624.
- (94) Wünsch, U. J.; Hawkes, J. A. Mathematical Chromatography Deciphers the Molecular Fingerprints of Dissolved Organic Matter. *Analyst* **2020**, *145*, 1789–1800.
- (95) Böcker, S.; Dührkop, K. Fragmentation Trees Reloaded. *J. Cheminform.* **2016**, *8*, No. 5.
- (96) 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.
- (97) Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-Ion Fragmentation. *Mol. Cell. Proteomics* **2010**, *9*, 2252–2261.
- (98) 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.
- (99) 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.
- (100) Lechtenfeld, O. J.; Hertkorn, N.; Shen, Y.; Witt, M.; Benner, R. Marine Sequestration of Carbon in Bacterial Metabolites. *Nat. Commun.* **2015**, *6*, No. 6711.
- (101) Mentges, A.; Feenders, C.; Blasius, B.; Dittmar, T. Functional Molecular Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. *Front. Mar. Sci.* **2017**, *4*, No. 194.
- (102) 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.
- (103) 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.
- (104) Dührkop, K.; Nothias, L. F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M. A.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; Böcker, S. Systematic Classification of Unknown Metabolites Using High-Resolution Fragmentation Mass Spectra. *Nat. Biotechnol.* **2021**, *39*, 462–471.
- (105) 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.
- (106) 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.

NOTE ADDED AFTER ASAP PUBLICATION

Due to a production error, this paper was published ASAP on July 14, 2022, without Table 1. The corrected version was reposted on July 18, 2022.