Mass difference matching crystallizes hidden molecular structures of dissolved organic matter from ultrahigh-resolution tandem mass spectra

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

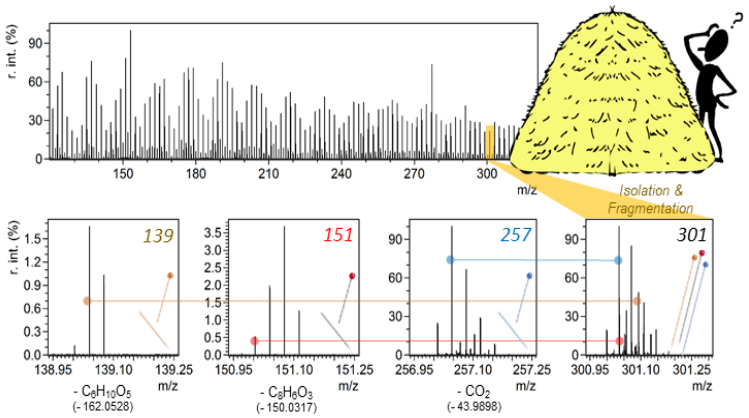
[1] Molecular Biogeochemistry, Department of Biogeochemical Processes, Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany

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

[3] Scripps Institution of Oceanography, University of California San Diego, San Diego, CA, USA

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

TOC FIGURE:



ABSTRACT: Ultrahigh-resolution Fourier transform mass spectrometry has revealed unprecedented detail of natural complex mixtures such as dissolved organic matter (DOM) on a molecular formula level. However, we lack detailed information on the underlying structural complexity which hinders full-scale molecular identification. Therefore, we applied a novel approach to decipher DOM's characteristic mixed ("chimeric") tandem mass (MS^2) spectra that represent multiple precursors of the same nominal mass. We (i) calculated mass difference (Δm) matrices for all precursor and product ions, (ii) matched them with reference Δm's from 11280 library MS^2 spectra and 14 phenolic reference compounds and (iii) linked the matched Δm's to molecular structures. Indicative Δm's revealed the presence of analogs of lignin, glycosides, hydrolyzable tannins, and flavonoids as well as unknown N- and S-containing molecules, which likely reflect remaining imprints of organic matter sources and processing. We found only weak support for postulated Van Krevelen structural domains often applied to identify DOM's molecular composition. However, we discovered multiple gradients of precursor properties significantly linked to the type

and number of matches, and structural suggestions. Additionally, the approach revealed heteroatom-containing (P, Cl) precursors not covered by our molecular formula annotation. Our paper highlights Δm matching by MS^2 as a promising tool to reveal novel structural information of complex mixtures like DOM.

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

Synopsis: We present an approach to explore the structural composition of mixtures of unknown organic molecules in environmental media to reveal their identity, source, and diversity.

INTRODUCTION

Complex mixtures are key study objects in environmental and industrial applications, but their analysis remains challenging.[1–3] One of the most complex mixtures in natural ecosystems is dissolved organic matter (DOM). [4,5] The diverse sources and molecular interactions of DOM with its abiotic and biotic environment mirror ecosystem functioning and ecosystem services[6–9] that form the basis for sustainable ecosystem management.[10–12] Despite significant advances in ultrahigh-resolution mass spectrometry (FTMS)[13,14] and nuclear magnetic resonance spectroscopy[15], scientists still struggle to decode this information on the molecular level[16–18], and novel approaches to identify distinct process markers are required.

Open and living systems are characterized by a large process diversity due to spatial organization and heterogeneity, changing boundary conditions, and community shifts, which promote the formation of an ultra-complex mixture of thousands to millions of individual constituents[19–21] that mirror these large environmental gradients.[19-26] As a consequence, most compounds found in DOM pose significant challenges in separation, isolation, and structure elucidation. Hence, direct infusion (DI) FTMS techniques have become indispensable tools for the molecular-level analysis of DOM as they reveal unprecedented detail at the nominal mass (MS^1 data) even without prior fractionation or separation.[16,21] However, FTMS techniques alone do not fully resolve all structural detail observed at the exact mass in DOM as the presence of isobars and isomers hinders the annotation of particular molecules and thus, full meta-metabolome annotation.[22–27] In addition to this, current structural databases cover only a minority

of the molecular formulas encountered, typically allowing for the annotation of less than 5% of features (i.e., ions, precursors).[19,28,29]

One way to dissect single molecular formula's chemical makeup in DOM is through gas phase fragmentation (MS^2, or multistage MS^n) experiments.[23,30] However, the relatively wide isolation windows (~ 1 Da) of mass filters applied for precursor selection often hinders the isolation and subsequent fragmentation of single exact masses, hence leading to mixed "chimeric" mass spectra.[30] Even though single authors have achieved isolation of single masses or improved description of chimeric tandem MS data, fragmentation patterns were found to be universal across DOM samples.[19,22,23,31–34] These studies, however, focused mainly on the major product ion peaks (fragments), which usually make up only 60 – 70 % of the total product ion abundance.[22,23]

The major product ions encountered in tandem mass spectra of DOM relate to sequential neutral losses of common small building blocks, mainly CO_2, H_2O, or CO units. A mass difference between a precursor and a product ion in an MS^2 spectrum is called "delta mass" and herein referred to as Δm (Δm's in the plural form). Many Δm's such as CO_2 or H_2O are commonly observed and thus are *non-indicative* for the identification of structural units (**Table S-1**).[19,22,31,35,36] In contrast, early studies found recurring low *m/z* product ions (e.g., *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, thus opening an avenue to identify their precursors.[33,35,37–42] Although many studies followed up on the core structure idea,[17,19,43,44] most recent studies mainly focus on similarities in the more abundant but non-indicative neutral losses, arguing that this reflects universal patterns of DOM diversification upon decomposition across environments.[22,23] From a stochastic standpoint, the occurrence of common neutral losses may not be surprising; for example, many structures contain hydroxyl groups that could yield H_2O losses, and CO_2 can originate from different functionalities despite carboxyl groups.[45] In contrast, the occurrence of two molecules sharing a larger substructure – a higher-order structural unit with a certain exact Δm – would be less probable. Signatures of DOM's structural diversity could thus prevail in the large number of rare higher-order structural units usually detected below *m/z* 200-300. The analysis of indicative Δm's, in contrast to indicative fragments alone, is independent of the masses of the unknown precursors and known reference compounds in databases of annotated Δm values. Although

this approach will sacrifice the identification of true knowns, it allows for the identification of potential structural analogs and is suited best when annotation rates are as low as 5% in the case of DOM, i.e., when most compounds are yet unknown.[19,29,30]

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 heterotrophic food webs in natural ecosystems. Plant metabolites such as polyphenols and polyaromatic structures thus represent a major source of DOM. Therefore, an early decomposition phase likely exists when the plant-related DOM source imprint is still detectable by MS^2 experiments using recent FTMS technology. An approach to circumvent the problem of unknown isomeric and isobaric diversity is to hypothesize about potential structural units that would be present if there was a plant-related imprint in DOM. For example, lignin-related compounds show indicative methoxyl and methyl radical losses[22,46]; glycosides indicate the loss of a sugar unit[47,48] and hydrolyzable tannins are expected to lose galloyl units.[48,49] Flavon-3-ols and flavan-3-ols show variable indicative retro-cyclization products.[47,50–52] Indicative Δm fingerprints could also provide evidence of putative compound group annotations derived solely from molecular formula data, as commonly applied for structural domains in the Van Krevelen diagram.[53–55]

We hypothesized that DOM from near-surface layers of soil in close contact to plant inputs and active microbial communities would reflect universal patterns of decomposition and recognizable plant-related source imprints that can be revealed by Orbitrap tandem mass spectrometry. We assumed that our approach allows the assignment of Δm identities in DOM based on a defined set of phenolic compounds, and that there will be clear differences among unknown precursors in Δm matching depending on precursor characteristics such as nominal mass, mass defect, initial ion abundance, fragmentation sensitivity, oxygen-to hydrogen ratio (O/C), or heteroatom content, which are predictable from the assigned molecular formula and thus allow an evaluation of the approach ("proof-of-concept"). More specifically, we hypothesized that indicative Δm features of plant phenols, e.g., lignin- and tannin-related losses, would match their yet unknown structural analogs in DOM and that these patterns reflect compound group distributions suggested by molecular formula and structural domains in the Van Krevelen plot.[55,56]

EXPERIMENTAL SECTION

A detailed experimental procedure is provided in the Supplemental Information of this article (**Note S-1**). In short, we chose a set of 14 aromatic reference compounds as representative plant metabolites in soil DOM (**Figure S-1**, **Table S-2**) and forest topsoil pore water isolate[57] as an exemplary DOM sample (**Figure S-3**) and infused the reference and sample solutions directly into the ESI (electrospray) source of an Orbitrap Elite (Thermo Fisher Scientific, Bremen) in the negative ionization mode (**Table S-3**). We performed collision-induced dissociation (CID) experiments at three normalized collision energy levels (15, 20, and 25%). MS^3 spectra of selected key product ions were acquired in some cases. We chose four nominal masses spanning the range of maximum ion abundance typically observed in terrestrial DOM samples for fragmentation (*m/z* 241, 301, 361, and 417)[58], and each of these contained a potential tannic forest marker described earlier.[57] After recalibration with known (**Table S-4**) or predicted product ions (losses of CO_2, H_2O, etc.), all major product ions were annotated with a molecular formula (**Figure S-2**, **Table S-5**, **Table S-6**). Formula annotation was conducted with a Matlab routine recently incorporated into an openly available FTMS data processing pipeline.[59] For MS^2 data analysis, we generated pairwise Δm's matrices of every combination of precursor and product ions. We matched DOM features against three lists of known Δm features: a) features ubiquitously found in DOM (**Table S-1**), b) features from a set of 14 reference compounds (**Table S-7),** and c) features from 11280 reference compound MS^2 library spectra in SIRIUS[60] (based on data from GNPS, MassBank and NIST) with a mass tolerance of ± 0.0002 Da (2 ppm at 200 Da). We assessed the probability of a false positive match and accounted for molecular formula constraints to evaluate our approach's validity. To analyze patterns of matching frequency, we visualized precursor formulas in Van Krevelen space.[55] We compared individual matching profiles of reference compounds and DOM precursors to evaluate the potential identity of underlying unknown structures by two-way hierarchical clustering using Ward's method and Euclidean distance in PAST (v3.10).[61] Precursors that only matched to non-indicative Δm's were disregarded from this analysis but were considered in a separate analysis of N- and S-containing formulas identified as lignin-like (based on O/C and H/C ratios).[56] The matching data was then combined for each CID level and transformed into a binary format. To evaluate the identity of potential structures based on indicative Δm features, we compared matching profiles of individual and clustered DOM formulas with structural formula suggestions. We then assessed structure suggestions from different databases, including Dictionary of Natural Products[62], KNApSAcK[63], Metacyc[64], KEGG[65], and HMDB[66] as well as their expanded in-silico annotations based on predicted enzymatic

transformations in the MINEs database.[67] The InChi-Key of structures was used to exclude stereoisomers and classify structures into major scaffold types by ClassyFire.[68]

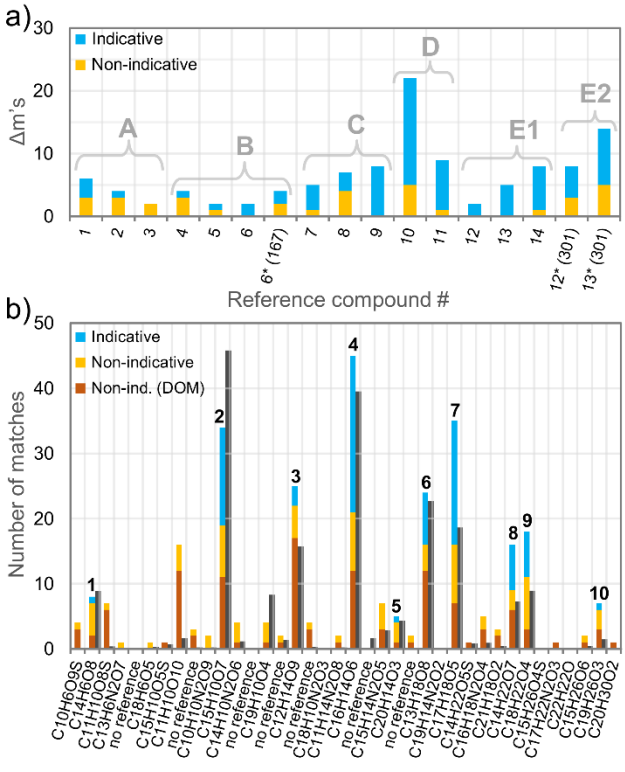


Figure 1. Tandem MS data from a) reference compounds and b) soil DOM. a) The total number of Δm's matched. Colors denote the number of Δm's that were specific to reference compounds or non-indicative (such as CO_2). Groups are A) small carboxy-phenols, B) small methoxy-phenols and methoxy-quinones, C) linked carboxy-phenols, D) flavanol-related structures, and E) flavonol glycosides and aglycones. b) The total number of Δm matches of precursors at *m/z* 301 (at CID25, n = 38, *m/z* increases to the right) with lists of indicative (blue), non-indicative (yellow) and ubiquitous DOM Δm's (orange). Black bars show (dimensionless) initial ion abundance of precursors. Only ten precursors (numbers) contributed to indicative matches.

RESULTS AND DISCUSSION

Tandem MS fragmentation behavior of reference compounds. The 14 phenolic reference compounds (**Figure S-1**, **Figure S-4**) yielded non-indicative as well as indicative Δm features. We observed CO_2 losses in nine reference compounds but this was not limited to the presence of carboxyl functionalities (as in substances **#1-3**).[32] Despite some common Δm's such as CO_2, the reference compounds also showed distinct fragmentation patterns (**Table S-5**). A dominant CO_2 loss characterized the three small carboxy-phenols (**Figure 1a**, **Figure S-1**, group A, Vanillic acid, Hydroxy-cinnamic acid, Gallic acid). Vanillic acid (**#1**) shared with members of group **B** (methoxy-phenols and methoxy-quinones) the presence of a methoxy group, which gave rise to the loss of a methyl radical

($CH_3^{\bullet}$). This loss was the main Δm in group B (Creosol, m-Guaiacol, 2,3-Dimethoxy-5-methyl-1,4-benzoquinone). Both methoxy-phenols indicated a formal O vs. CH_4 insertion. Ion abundance of the oxidized product was below 1% at CID0 and increased to 2% (#5, m-Guaiacol) and 17% at CID 15 (#4, Creosol). The benzoquinone did not expel a CO unit.[69] Group **C** (linked carboxy-phenols, **#7** – **#9**) was mainly characterized by cleavage of ester bonds (e.g., loss of quinoyl or caffeoyl moieties from #**7**). The intramolecular lactone bonds in ellagic acid (#**8**) were, in contrast, exceptionally stable upon fragmentation and yielded rich product spectra only at higher relative CID energies (> 25), featuring indicative CO losses[69], but also losses of CO_2. Compounds **#10** and **#11** (group **D**) shared a $C_6H_6O_3$ loss (unmodified A ring in #**10**, abstraction of trihydroxy-benzene from gallate unit in #**11**). [51,70] Catechin (#**10**) had the most diverse product spectrum among all compounds investigated, including some indicative Δm's of retro-cyclization reactions (fragments at *m/z* 205, 203, 179, 151, 125, and 109, **Table S-5**).[50,52,71] Compound #**11**, containing a flavan-3-ol subunit, resembled especially #**9** through the presence of a gallate subunit that produced similar Δm's: An incomplete galloyl loss with retention of H_2O ($C_7H_4O_4$), a galloyl loss ($C_7H_6O_5$), or a combined galloyl and H_2O loss ($C_7H_8O_6$). The flavonoids (group **E**, compounds #**12** – #**14**, containing flavon-3-ol cores, Spiraeoside, Isoquercetin, Myricitrin) showed a clear loss of the attached glycosidic sugar as the main Δm (**Table S-7**). They differed in the type of sugar (**12** and **13**, glucose, **14**, rhamnose), and also in the charge state (**12**, ion form of aglycon dominated; **13**, equal; **14**, radical anion form dominated).[48] This effect also influenced the further fragmentation of the aglycon, which proceeded in **14** (less so in **13**) but not in **12**. MS^3 spectra of the aglycon ions (*m/z* 301; **#12***, **#13***) showed indicative retro-cyclization products (at *m/z* 179, 151, 121 and 107).[47] More details on reference compound fragmentation are given in the supplementary material (**Note S-2**).

Fragmentation behavior of unknown DOM precursor mixtures. The nominal masses of the isolated precursor ion mixtures (IPIMs) *m/z* 241, 301, 361 and 417 cover the mass range that is typically observed in soil DOM and were thus chosen for MS^2 experiments. Each IPIM yielded a mixture of up to 44 isolated precursor ions and showed universal continua of fragmentation properties. The molecular weight of the four IPIMs was significantly (Pearson, $p<0.05$) related to lower numbers of double bonds (DBE) and aromaticity (AImod), higher nominal oxidation state of carbons (NOSC) and higher numbers of precursor ions and product ions (up to 44 and 491, respectively; **Table S-8**). Independent of *m/z*, we always detected the highest numbers of product ions at the highest CID of 25 (**Figure S-5**). The product ion spectra did not indicate abrupt structural changes upon increasing

fragmentation energy, showing no clear separation of groups of isomers or scaffold types but rather a continuous increase in fragmentation across all precursors. Fragmentation was selective in terms of mass defect across all IPIMs. With increasing collision energy, the remaining mixture of precursors significantly increased in average DBE, DBE-O, and AImod, and decreased in O/C and NOSC (ion abundance-weighted averages; **Table S-8**). IPIMs also became more similar in molecular composition upon fragmentation (i.e., average H/C, O/C, etc.; not shown), suggesting common properties among precursors resisting fragmentation. This finding supports the view that DOM's structure is based on a limited set of regular backbone structures with similar properties.[33,37,38,41]

Precursors more sensitive to fragmentation showed a significantly (Pearson, $p < 0.05$) lower mass defect, lower numbers of C and H atoms per formula, and a higher O/C and NOSC (but not related to N/C or S/C). As a result, we found a continuum of fragmentation sensitivities across the mass defect scale at each of the four IPIMs, ranging from half-abundance energies (i.e., the collision energy causing 50% decrease in ion abundance) of CID 10 – 35 under our instrumental settings (calculated from linear fits, **Figure S-6**). A minor group of oxygen-poor formulas was non-responsive (**Note S-3**). These findings indicate that intrinsic averaging prevailed in the property of fragmentation sensitivity in our study, similar to other continua reported in DOM.[19,21,72] In contrast, initial ion abundance was not linked to fragmentation sensitivity but showed a significant correlation to higher numbers of non-indicative (**Table S-9**, **Table S-10**, **Table S-11**, and **Table S-12**) and indicative Δm matches (**Figure S-7**). Abundant and relatively oxygen-rich precursors matched more often to both non-indicative and indicative Δm's (**Figure 1b**, **Figure S-7, Figure S-8**).[19,23,33] These observations show that fragmentation sensitivity and Δm matching are independent DOM precursor properties, except in the case of fragmentation-resistant precursors that showed no Δm matches (**Figure S-9**).

Evaluation of the Δm matching approach. We used the matching data of unknown DOM precursors annotated with a molecular formula for a proof-of-concept evaluation of our Δm matching approach. Our analysis of Δm's in DOM was congruent with previous observations, showing ubiquitous losses of small non-indicative oxygen-containing functionalities (**Table S-1**) while also revealing more detail (**Figure S-3c**, **Table S-7**). In line with continua reported in the previous section, we found distinct trends in the Van Krevelen distributions of unknown precursors, indicating regular shifts in dominance of serial losses of CO_2, CO, and CH_2 units (**Figure 2a – c**). Highly oxidized precursors (high O/C) tended to expel CO_2 rather than CH_2 units (as noted above, they were also more

sensitive to fragmentation and matched to more Δm's). In contrast, precursors with low O/C ratios were generally more resistant to fragmentation and subsequently showed a tendency to match with Δm's related to subsequent losses of CH_2 units, and precursors with low H/C ratios tended to expel CO units.

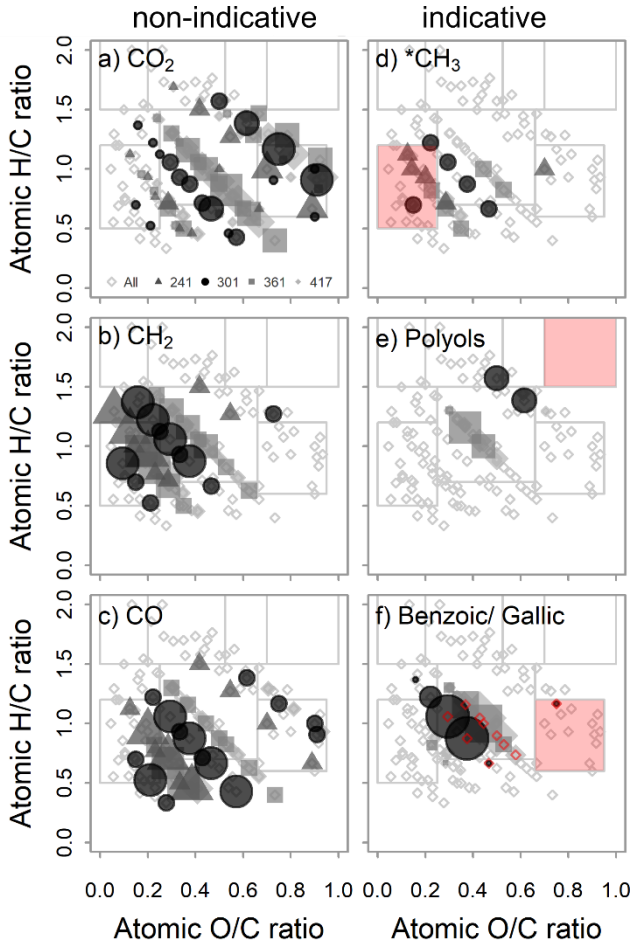


Figure 2. Δm matching in Van Krevelen space. Small open diamonds show all precursors with an assigned molecular formula (n=127) at their atomic ratios of O/C and H/C. Grey boxes indicate representative structural domains that are commonly used (taken from Minor et al. (2014), see also **Figure S-4**).[56] Symbol and color show the IPIM of matching formulas (see legend in panel a). Symbol size encodes the number of matches to non-indicative (**a-c**, left column) and indicative Δm's (**d-f**, right column). Red boxes in indicative VK plots mark the expected structural region of formulas that would yield the respective Δm. Symbol size adjusted for each VK plot to visualize broad trends. Δm's are a) CO_2 (max=4, size reduced by factor 0.5), b) CH_2 (max=4, size factor 0.5; CH_2 losses can occur in sequence or as C_nH_{2n} units, with n being 2, 3 or 4), c) CO (max=2), d) Methyl radical (max=1) e) Δm's equivalent to polyol losses (max=4, size factor 0.5), and f) Δm's equivalent to benzoic acid or phenol loss (max=10, size factor 0.33). Red open diamonds in f) indicate loss of up to three gallic acid equivalents (size not drawn to scale).

Our approach's revealing of inherent structural information was also supported by other key observations, such as the predicted heteroatom content (O, N, S) of assigned molecular formulas. As expected, matching Δm's were

constrained by precursor formula and vice versa. Precursors rich in oxygen were predicted to expel more oxygen-containing Δm's than oxygen-poor precursors that tended to lose CH_2 or $CH_3^{\bullet}$ (and CO) units instead. Most notably, no precursors matched to a Δm that would have exceeded the number of atoms present in their assigned molecular formula, a condition that has not always been met in earlier studies.[33] Sulfur- and Nitrogen-containing precursors – and only those – indicated the release of previously described element-specific Δm's.[29,73,74] A second matching exercise against the complete library of Δm's available from 11280 tandem mass spectra combined in the SIRIUS database substantiated this finding (**Figure S-10**). We report reoccurring Δm's from this database for each CHO, CHNO, and CHOS formulas, many for the first time in DOM (Supplementary Material, **Table S-13**, **Table S-14**, and **Table S-15**, and further discussion below). We furthermore did not observe an increase in the number of false-positive matches upon widening of the tolerance window applied during the Δm matching process (**Figure S-11**, increase from 2 to 10 ppm, at 200 *m/z*). Lastly, precursors resisting fragmentation did not match any Δm, whereas "labile" precursors fragmented to relative completeness showed a wide range of matches (**Figure S-9**).

The combination of these observations leads us to conclude that the Δm matching approach presented herein does not yield random matches, although the pairwise Δm calculation (precursor ion *m/z* - product ion *m/z*) would theoretically allow for such an artifact. Instead, it reveals molecular detail of a biogeochemical signal. A random matching result to Δm's of seemingly wrong precursor compositions (e.g., loss of S from a sulfur-free precursor; four CO_2 losses from a precursor with only seven oxygen atoms) would be expected if the calculated Δm values were either derived from noise or reactions in the collision cell, and not from an inherently structured biogeochemical signal from precursors that fragment individually. This is a notable finding as it suggests that it will be possible to deconvolute chimeric mass spectra from IPIMs in the future. These findings suggest that the "match assignment" of higher-order indicative Δm's may reveal differences in DOM molecular composition not visible from MS^1 inspection alone.[24,75]

Lastly, the positive evaluation also shows that mass difference matching is not only a valuable approach to recalibrate FTMS datasets of complex organic mixtures[76] but can also serve to check formula annotation routines. Most precursors in our study were successfully annotated with a molecular formula containing the major elements C, H, N, O and S, and this was substantiated by matching to respective Δm's of correct mass and elemental composition. However, a minor number of unannotated formulas did indicate the presence of P and Cl (but not F,

Br or I, which were also part of the Δm library list), which may be taken as a sign that these atoms should be included for better coverage of elemental composition (i.e., prioritization) in our specific sample context (**Figure S-10**).[3]

DOM ecosystem imprints revealed. Δm matching revealed unexpected higher-order mass differences present in DOM (Δm features used, see **Table S-16**). Among the most prominent indicative features was the methyl radical loss[22,33,46] which matched to oxygen-poor DOM precursors (n = 19, average O/C = 0.3, **Figure 2d**). The distribution of $CH_3^{\bullet}$-yielding precursors was paralleled by CH_2 and CO losses, i.e., implying similarities between precursors expelling these Δm's (**Figure 2b, c**), e.g., condensed structures with aliphatic, lactone, or quinone moieties.[69] The methyl radical loss is an expected diagnostic Δm of methoxylated aromatic rings such as present in lignin (**Note S-4**), but was also matched to highly condensed DOM precursors not viewed classically as "lignin-like" (**Figure 2d**, red square).[19,22,33,46] Methyl radical loss has also been noted from methyl- or methoxy-substituted aromatic structures in positive ESI conditions.[45] The presence of methoxy functionalities in soil DOM likely reflects the high transformation potential of non-soluble organic materials by the decomposer community.[46] Ester-linked carboxylated phenols (**#7**, **#11**) and O-glycosides (**#9**, **#12**, **#13**, **#14**) all cleaved central O-linkages at low energy, leading to the loss of hydrogen-rich substructures[31,36,48] and were matched to DOM precursors (**Figure 2e**). These high-mass Δm's (e.g., $C_8H_6O_3$ or $C_6H_{10}O_5$; **Figure S-3e, f**) are likely no combinations of the more dominant oxygen-rich neutral losses (CO, H_2O, or CO_2) due to their low O/C and O/H ratios, but this must be further tested, e.g., with model mixtures. Aliphatic side chains, for example, prevail as O-poor substituents of cyclic core structures in DOM and could also contribute.[77,78] Unexpectedly, gallate Δm's did not match with precursors in the anticipated "tannic" structural domain but with those in the center of the Van Krevelen plot (**Figure 2f**, red square). The "tannic" identity of previously Van Krevelen-classified forest ecosystem markers[57] in this particular DOM sample could not be confirmed by our tandem FTMS spectra. Matching frequencies to Δm equivalents of indicative ring cleavage series of flavon-3-ols (i.e., flavonoid aglycones[47,48], 28 matches in DOM), flavan-3-ols (i.e., catechin[50–52], 50 matches), and benzoic-acid-related Δm's followed the same trend as gallate Δm's (**Figure 2f, Figure 3**). Likewise, Δm equivalents of highly indicative polyol losses[47,48] matched to 25 DOM precursors (**Figure 2e**) in the central Van Krevelen plot despite the absence of "carbohydrate-like" precursors (**Figure 2e**, red square).

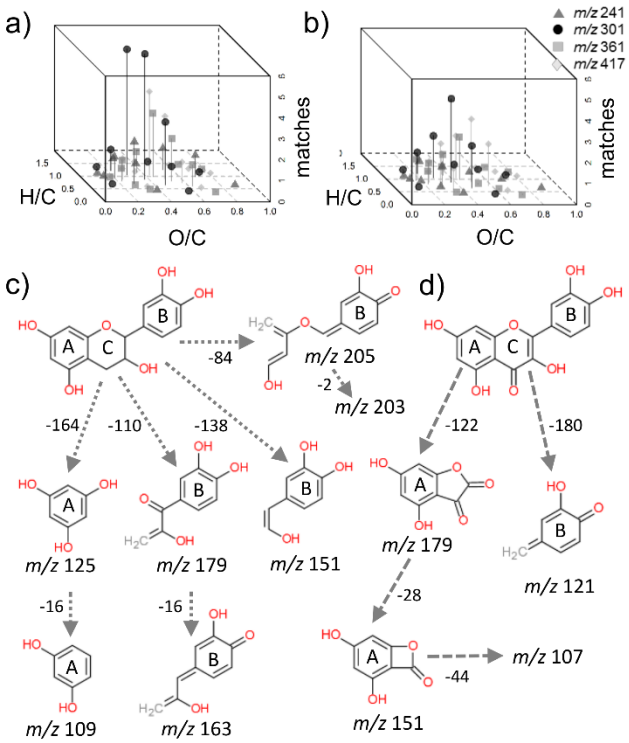


Figure 3. 3D-Van Krevelen plots showing matches against a) six Δm's indicative of flavan-3-ol scaffolds and b) four Δm's indicative of flavon-3-ol scaffolds (**Table S-16**). c) Scheme showing the major neutral precursors and products of the suggested fragmentation pathway of a flavan-3-ol (shown is catechin, #10). Related Δm's are given as nominal *m/z* (**Table S-7**).[50–52] d) Similar scheme of the suggested fragmentation pathway for a flavon-3-ol (shown is quercetin, core structure in #13).[47]

We interpret the successful matching of indicative Δm's in forest topsoil DOM as a remaining source imprint of primary or recycled organic remains from plants, soil animals and microorganisms.[6,9,10,12,18] We acknowledge that their low abundance (**Figure S-3**) agrees with rapid vanishing of biochemical imprints during initial decay.[23,31,32,36] This view has, however, emerged from common and ubiquitous DOM signals (precursor ions, product ions, and Δm's), which represent only 60 – 70% of the information.[22,23] Our observation of high numbers in rare but indicative matched Δm features in DOM shows the importance of the missing information for models of DOM chemodiversity.

All in all, our findings indicate large deviations between Δm matching patterns and expected structural domains in Van Krevelen space, and thus question our recent understanding and means of interpretation of DOM chemistry (**Figure 2**; **Figure S-4**), which will be further discussed in the following two sections. Chimeric tandem mass spectra pose significant challenges for structure annotation in DOM[30], but can be used to mine Δm patterns of knowns and those matching with unknown precursors in complex DOM samples by simple deconvolution. Further

tests with model mixtures are needed to reveal the rules of simultaneous precursor fragmentation experiments and to improve identification from mixtures, also by applying complementary techniques.[25,26]

Structural differences among DOM precursors. We used two-way clustering to compare precursor Δm matching profiles (**Figure S-12**). Six clusters of precursors (A – F) and seven Δm clusters were differentiated (**Figure S-13a-e**, **Table S-17**, **Note S-5**). All in all, precursors were clearly clustered according to number of matches and thus initial ion abundance, but the type of Δm matches differed strongly as well, especially between the first four precursor clusters (A – D; **Figure S-12**). Clustering of Δm features reflected the major differences between non-indicative and indicative features, and the matching of O-rich (and O-poor) precursors and Δm features (such as polyol equivalents or $CH_3^{\bullet}$) related to differences in fragmentation sensitivity as described above. Precursor pairs linked through a formal CH_4 vs. O exchange[40,41] often showed high similarities (e.g., $C_{16}H_{14}O_6$ and $C_{17}H_{18}O_5$) that contrasted with other members of the same series (e.g., $C_{15}H_{10}O_7$), but ion abundance was not the primary driver of this effect. Consequently, each IPIM covered 2-3 CH_4 vs. O series that were spread across 1-4 precursor clusters, reflecting the large differences in matching.

High congruence of fragmentation patterns among sets of DOM precursors has been interpreted as a sign of similarly substituted but slightly differing core structures.[33,41] The wide differences in matching reported herein, however, show that this model may fall short in describing the full complexity of DOM fragmentation, especially looking at rare but highly informative structural signatures.

The strong differences between precursors were also apparent for members of the same "structural domain" that have been postulated based on Van Krevelen plots. Seventeen precursors that matched to at least one of the 42 indicative Δm's of the reference compound set (**Table S-7**) were classified as "lignin-like" formulas according to Minor et al. (2014)[56] and were grouped into four different precursor clusters (A – D; **Figure S-12**, **Table S-18**) that differed widely in matching (especially in O-rich flavonoid-, gallate and polyol equivalents, Δm cluster 4; small O-poor aliphatic equivalents, Δm cluster 5; and phenylpropanoid equivalents, Δm cluster 6). Six S-containing and ten N-containing precursors were also classified as lignin-type formulas (**Table S-19**) despite the absence of N and S in lignin-like structures, which reiterates the need to use these classifications with caution.[54,75,79] Although these precursors did not match to any of the indicative reference compound Δm's, they matched with many of the S- and N-containing GNPS-derived Δm features (spanning 35 – 118 S-containing and 57 – 247 N-containing Δm's) that

represented on average 90 ± 4% of all matches per precursor (**Table S-19**). This not only indicates high specificity and robustness of matching, but also revealed large differences among these precursors in terms of number and type of potential structural links. This shows the high potential of Δm matching to reveal hidden structural detail of heteroatom-containing formulas in DOM.

Negative-mode ESI CHNO precursor ions generally show few neutral N losses in aquatic DOM and thus have been interpreted as alicyclic or aromatic heterocyclic N such as in imide, pyridinic or pyrrolic moieties that are substituted with carboxyl and hydroxyl groups.[69,73] In line with these earlier reports, we found no evidence of nitrate esters (HNO_3 loss, Δm = 62.9956) in soil DOM. However, a majority of N-containing precursors (here, all within ranges $C_{10\text{-}23}H_{6\text{-}26}N_2O_{1\text{-}11}$, n=27) showed a link to N_2 (Δm = 28.0061 Da, 92%), N_2O (44.0011 Da, 92%), and CH_4N_2 (44.0374 Da, 77%), and multiple other N losses. Such a diversity of potential N losses contradicts with previous reports, but many N compounds yield fragments in negative ion ESI-MS.[80] Loss of N_2 could indicate direct cleavage under negative ESI conditions, possibly from azo/diazo-functionalities. Lemr et al. (2000) have shown that cleavage of azo/ diazo-N in metal azo-complexes was possible directly (MS^2) or indirectly ($MS^{>2}$) as N_2 or in other reduced forms (e.g., CH_3N, $C_3H_3N_2$, or CHN).[81]

S-containing precursors (here, all within ranges $C_{9\text{-}24}H_{6\text{-}34}O_{2\text{-}12}S_1$, n=23) matched with Δm's indicative of sulfonic acids: SO_2 (Δm = 63.9619, 4% of all S precursors), SO_3 (79.95681, 60%) and H_2SO_3 (81.97246, 35%). Against previous reports, however, we also found potential direct losses of S (31.97207, 65%) which can originate from reduced sulfur functionalities, such as thiophenes, thioethers, sulfoxides and thioesters.[74] Other reduced S Δm's were also commonly matched, including CS (43.97207, 78%) and CH_2OS (61.98263, 74%; possibly as a combination $CO+H_2S$), which have been observed in positive ionization mode via atmospheric pressure photoionization (APPI) in aromatic reference compounds.[82] This may indicate a more diverse set of S-containing molecules in soil as compared to the deep ocean, where oxidized species seem to dominate.[74] Matched Δm's containing S and > 3 C atoms always contained oxygen atoms as well, which indicates that extensive S-containing aliphatic chains were likely no common structural unit in our DOM sample; alternatively, they may have been missed due to low ionization or because they resisted fragmentation.[82] Further tests with N- and S-containing reference compounds and DOM samples are necessary to reveal the diversity and identity of dissolved organic nitrogen and sulfur molecules in soil in detail.

Ion abundance is linked to Δm matching frequency and structural diversity. In contrast to the expectation that indicative Δm's would reflect structural domains, they showed most frequent matching in the central part of the Van Krevelen plot (**Figure 2**, **Figure 3**). This structural domain has been assigned to ubiquitous and abundant lignin and carboxyl rich aromatic molecule (CRAM)-related precursor structures (**Figure S-4a**)[57,58,78] and parallels with a maximum in potential underlying chemodiversity.[83] We thus used our data to test whether the number and type of matched Δm features are suitable variables to reveal such proposed chemodiversity patterns in DOM by combining it with structural suggestions.

We found significant positive correlations between the numbers of structural suggestions and Δm matches per precursor or precursor cluster (**Figure 4)**, and this was also true for specific Δm features and the related scaffold types (**Figure S-13f-k, Table S-20**). We acknowledge that natural product databases are far from being complete. We took this effect into account by 1) extending our Δm database from 14 reference compounds to 11280 tandem mass spectra available through GNPS, and 2) mining for structure suggestions in several databases including in-silico structures predicted from enzymatic reactions (see method section). These extensions increased the strength of the correlation (**Figure 4**). Matching frequency and initial ion abundance were thus strongly related to the number, type, and diversity of structure suggestions. Precursors with low mass defects showed exceptionally few structural hits, likely indicating bias in natural product databases (**Figure S-14**).[19]

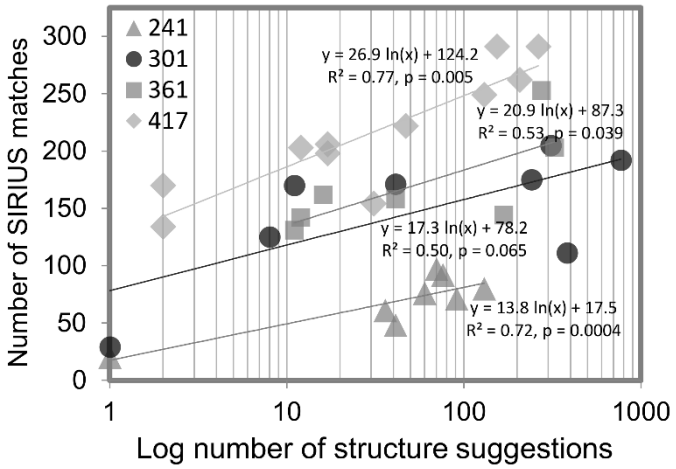


Figure 4. The relation between the number of structure suggestions and Δm matches with an extended set of tandem MS fingerprints of reference compounds in GNPS (11280 spectra in negative ESI mode, including 35722 unique Δm features). Matching frequency and number of hits were both positively correlated to precursor ion abundance (**Figure S-13**). We included

only features detected at least three times across all spectra for matching (n=9981 Δm's). Note the logarithmic scale on the x-axis.

At this point, it is not clear whether an increased matching frequency is due to a better S/N of a DOM precursor and its product ions, or if it indicates high chemodiversity. The number of matched indicative Δm's assessed in this study may be interpreted as a first, very rough measure to account for underlying molecular complexity due to the low number of precursors tested. The general agreement between types of Δm's and structure suggestions however suggests that Δm matching may reflect structural properties even in complex mixtures of precursor ion species such as in DOM (**Figure 4**, **Figure S-13f-k**). In support of that, our observations agree well with theoretical considerations on the probability distribution of structural diversity in two-dimensional Van Krevelen space.[24,30,53,83] Along with recent progress on the aspect of ionization effects in complex mixtures[84], our results encourage further studies on the Δm matching behavior of synthetic mixtures of knowns, variation among unknowns across DOM chemotypes, and the improved bioinformatic exploitation of chimeric (LC-) $FTMS^n$ data of complex organic mixtures.[85–87]

CONCLUSION

We here present a novel "Δm matching" approach to improve the analysis and interpretation of chimeric tandem mass spectra from ultra-complex mixtures ubiquitously found in nature, at the example of DOM. DOM is the most mobile and elusive form of carbon in soils and mediates many of the fundamental processes that maintain functional soils. Our approach allows to exploit a large source of hitherto untapped molecular and structural information that, if routinely assessed, will enable new insights in these fundamental processes. Our results suggest that ultrahigh-resolution tandem mass spectra (MS^2) from DOM precursor mixtures, commonly described as "chimeric" MS^2 data, can be deconvoluted to yield individual precursor fingerprints of potential structural composition. We report hundreds of Δm features for the first time in soil DOM, allowing a glimpse into the complex chemistry of dissolved organic molecules in soil, including organic sulfur and nitrogen compounds, and identify elements that may need to be included into molecular formula annotation routines (phosphorus, chlorine). Number and types of Δm matches vary largely among precursors in one sample, thereby suggesting even stronger differences between samples or treatments that are to be studied in future. Most importantly, our data provide timely experimental proof that the Van Krevelen plot – the most widely used approach to interpret DOM molecular composition data – needs to be

used with extreme caution for structural interpretation. Although the presence of indicative Δm features indicate that precursors may be linked to certain structural features, their gradual and monotonous matching patterns (i.e., trends in CO_2, CH_2 and CO losses, and highest matching going along with structural diversity and ion abundance) strongly suggest a dominant randomization during decomposition in soil. It thus seems warranted to assume that soil DOM chemistry diverges largely from what is covered in natural product databases based on plant and microbial samples. The emergence of the extraordinary chemodiversity of DOM or related complex mixtures requires novel forms of experimentation that include top-down approaches by studying DOM transformation in the lab and field, but also bottom-up experiments that mimic complex mixtures based on known compounds or well-studied systems such as plant or microbial extracts. The Δm matching approach presented herein opens exciting avenues for hypothesis testing on DOM transformation in soils, for example regarding the impact of enzymatic treatments, microbial decomposition, or nutrient recycling. Together with the constantly growing MS databases such as Mass Bank or GNPS and comprehensive chromatographic and ion mobility decomplexation methods, MS^2 Δm matching will provide fundamental insights for the deconvolution of chimeric spectra and ultimately the hidden molecular diversity of dissolved organic matter in soils and beyond.

ASSOCIATED CONTENT

Data and Software Code Accessibility

All MS/MS data can be found on the Mass spectrometry Interactive Virtual Environment (MassIVE) under the following links: ftp://massive.ucsd.edu/MSV000087117/ (DOM data) and ftp://massive.ucsd.edu/MSV000087133/ (reference compound data). Other data associated to this manuscript is available online free of charge from the PANGAEA Data Publisher under the following link: xxx.

Supporting Information

The supporting information contains twenty tables and fourteen figures, five additional notes (incl. supplementary experimental section), and sixty-four references.

Table S-1: List of reported or proposed DOM Δm features from MS1 studies. Table S-2: Information on reference compounds and solutions used in this study. Table S-3: Instrument settings for fragmentation experiments. Table S-4: Recalibration peaks used for reference compound measurements. Table S-5: Precursor and major product ions of the 14 reference compounds. Table S-6: Results of reference compound data analysis with SIRIUS and CSI:FingerID. Table S-7: List of all 50+5 Δm features extracted from the reference compound dataset. Table S-8: Properties of non-fragmented isolated precursor ion mixtures

(IPIMs). 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: List of 234 GNPS Δm features that reoccurred in CHO formulas. Table S-14: List of 45 GNPS Δm features that reoccurred CHNO formulas. Table S-15: List of 25 GNPS Δm features that reoccurred CHOS formulas. Table S-16: Lists of Δm values used for analysing matching patterns in Van Krevelen space. Table S-17: Properties and Δm matching behavior of precursor clusters. Table S-18: Lignin-like precursor formulae, their molecular properties and clustering. Table S-19: Lignin-like precursor formulae containing N or S. Table S-20: Overview of suggested structures and their major scaffold categories. Figure S-1: Overview of reference compounds used in the study. Figure S-2: Error assessment of reference compound Δm's. Figure S-3: Orbitrap tandem MS of DOM (exemplary spectra). Figure S-4: Distribution of known structures in chemical space (C, H, O, and m/z). Figure S-5: Comparison of IPIMs Δm matches to indicative and non-indicative Δm's. Figure S-6: Demonstrating the presence of a fragmentation sensitivity continuum in DOM. Figure S-7: The number of Δm matches vs. precursor ion abundance. Figure S-8: Numbers of Δm matches of DOM precursors (IPIMs 241, 361, and 417). Figure S-9: The number of Δm matches vs. precursor fragmentation sensitivity. Figure S-10: Matching against the full list of GNPS Δm's (evaluation of approach). Figure S-11: Changes in Δm matching frequency upon widening of tolerance window. Figure S-12: Δm matching profiles and their similarities (two-way cluster analysis). Figure S-13: Links between precursor clusters, matching efficiency and structure suggestions. Figure S-14: Effect of mass defect on the number of structure suggestions. Note S-1: Supplementary experimental details. Note S-2: Detailed description of pure substance fragmentation behavior. Note S-3: Behavior of non-responsive DOM precursor ions. Note S-4: Potential esterification of DOM by methanol during SPE and storage. Note S-5: Matching behavior of precursor clusters.

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

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

Author Contributions

The manuscript was drafted by CS and 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.

Synopsis

We present a novel approach to reveal previously disregarded but important structural information in a highly complex mixture of soluble organics extracted from soil.

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 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).

REFERENCES

(1) 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.

(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) 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.

(4) 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.

(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) 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**, No. February, 1–14. https://doi.org/10.1111/1365-2745.13556.

(7) Delgado-Baquerizo, M.; Maestre, F. T.; Reich, P. B.; Jeffries, T. C.; Gaitan, J. J.; Encinar, D.; Berdugo, M.; Campbell, C. D.; Singh, B. K. Microbial Diversity Drives Multifunctionality in Terrestrial Ecosystems. *Nat. Commun.* **2016**, *7*,

10541. https://doi.org/10.1038/ncomms10541.

(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) 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.

(10) Kästner, M.; Miltner, A. SOM and Microbes - What Is Left from Microbial Life. In *The future of soil carbon: Its conservation and formation*; Garcia, C., Nannipieri, P., Hernandez, T., Eds.; Academic Press: Waltham, 2018; pp 125–163. https://doi.org/10.1016/B978-0-12-811687-6.00005-5.

(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.; Pulleman, M.; Sukkel, W.; van Groenigen, J. W.; Brussaard, L. Soil Quality – A Critical Review. *Soil Biol. Biochem.* **2018**, *120*, 105–125. https://doi.org/10.1016/j.soilbio.2018.01.030.

(12) 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. https://doi.org/10.1038/s41561-020-0612-3.

(13) Hawkes, J. A.; Kew, W. High-Resolution Mass Spectrometry Strategies for the Investigation of Dissolved Organic Matter. In *Multidimensional Analytical Techniques in Environmental Research*; Duarte, R., Duarte, A., Eds.; Elsevier Inc., 2020; pp 71–104. https://doi.org/10.1016/b978-0-12-818896-5.00004-1.

(14) Cooper, W. T.; Chanton, J. C.; D'Andrilli, J.; Hodgkins, S. B.; Podgorski, D. C.; Stenson, A. C.; Tfaily, M. M.; Wilson, R. M. A History of Molecular Level Analysis of Natural Organic Matter by FTICR Mass Spectrometry and The Paradigm Shift in Organic Geochemistry. *Mass Spectrom. Rev.* **2021**, 1–25. https://doi.org/10.1002/mas.21663.

(15) 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.

(16) Hertkorn, N.; Harir, M.; Cawley, K. M.; Schmitt-Kopplin, P.; Jaffé, R. Molecular Characterization of Dissolved Organic Matter from Subtropical Wetlands: A Comparative Study through the Analysis of Optical Properties, NMR and FTICR/MS. *Biogeosciences* **2016**, *13*, 2257–2277. https://doi.org/10.5194/bgd-12-13711-2015.

(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) 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. https://doi.org/10.1038/s41561-019-0417-4.

(19) 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.

(20) Malik, A. A.; Roth, V.-N.; Hébert, M.; Tremblay, L.; Dittmar, T.; Gleixner, G. Linking Molecular Size, Composition and Carbon Turnover of Extractable Soil Microbial Compounds. *Soil Biol. Biochem.* **2016**, *100*, 66–73. https://doi.org/10.1016/j.soilbio.2016.05.019.

(21) Hawkes, J.; Sjöberg, P. J. R.; Bergquist, J.; Tranvik, L. Complexity of Dissolved Organic Matter in the Molecular Size Dimension: Insights from Coupled Size Exclusion Chromatography Electrospray Ionisation Mass Spectrometry. *Faraday Discuss.* **2019**, *218*, 52–71. https://doi.org/10.1039/c8fd00222c.

(22) 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.

(23) 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.

(24) 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.

(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) 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.

(27) Hartmann, A. C.; Petras, D.; Quinn, R. A.; Protsyuk, I.; Archer, F. I.; Ransome, E.; Williams, G. J.; Bailey, B. A.; Vermeij, M. J. A.; Alexandrov, T.; Dorrestein, P. C.; Rohwer, F. L. Meta-Mass Shift Chemical Profiling of Metabolomes from Coral Reefs. *Proc. Natl. Acad. Sci.* **2017**, *114*, 11685–11690. https://doi.org/10.1073/pnas.1710248114.

(28) 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*, 129450. https://doi.org/10.1016/j.chemosphere.2020.129450.

(29) 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.

(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) 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.

(32) 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.

(33) 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.

(34) 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.

(35) 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.

(36) 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.

(37) 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.

(38) 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.

(39) 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.

(40) 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.

(41) 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.

(42) 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.

(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) 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.

(45) Smyth, W. F.; Ramachandran, V. N.; O'Kane, E.; Coulter, D. Characterisation of Selected Drugs with Nitrogen-Containing Saturated Ring Structures by Use of Electrospray Ionisation with Ion-Trap Mass Spectrometry. *Anal. Bioanal. Chem.* **2004**, *378*, 1305–1312. https://doi.org/10.1007/s00216-003-2414-z.

(46) 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.

(47) 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.

(48) 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.

(49) Gross, G. G. Biosynthesis of Ellagitannins: Old Ideas and New Solutions. In *Chemistry and Biology of Ellagitannins*; Quideau, S., Ed.; World Scientific: London, 2009; pp 94–118. https://doi.org/10.1142/9789812797414_0003.

(50) Yuzuak, S.; Ballington, J.; Xie, D.-Y. HPLC-QTOF-MS/MS-Based Profiling of Flavan-3-Ols and Dimeric Proanthocyanidins in Berries of Two Muscadine Grape Hybrids FLH 13-11 and FLH 17-66. *Metabolites* **2018**, *8* (4), 57. https://doi.org/10.3390/metabo8040057.

(51) 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.

(52) Galaverna, R. S.; Sampaio, P. T. B.; Barata, L. E. S.; Eberlin, M. N.; Fidelis, C. H. V. Differentiation of Two

Morphologically Similar Amazonian Aniba Species by Mass Spectrometry Leaf Fingerprinting. *Anal. Methods* **2015**, *7* (5), 1984–1990. https://doi.org/10.1039/c4ay02598a.

(53) 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.

(54) 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.

(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. https://doi.org/10.1021/acs.analchem.8b00529.

(56) 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.

(57) 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.

(58) 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* (September), 1–16. https://doi.org/10.3389/feart.2018.00138.

(59) 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. https://doi.org/10.1021/acs.analchem.9b05659.

(60) 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.

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

(62) 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.

(63) 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.

(64) 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.

(65) 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.

(66) 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.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35*, 521–526. https://doi.org/10.1093/nar/gkl923.

(67) 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.

(68) 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*, 61. https://doi.org/10.1186/s13321-016-0174-y.

(69) 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.

(70) Poon, G. K. Analysis of Catechins in Tea Extracts by Liquid Chromatography-Electrospray Ionization Mass Spectrometry. *J. Chromatogr. A* **1998**, *794* (1–2), 63–74. https://doi.org/10.1016/S0021-9673(97)01050-9.

(71) Rockenbach, I. I.; Jungfer, E.; Ritter, C.; Santiago-Schübel, B.; Thiele, B.; Fett, R.; Galensa, R. Characterization of Flavan-3-Ols in Seeds of Grape Pomace by CE, HPLC-DAD-MS n and LC-ESI-FTICR-MS. *Food Res. Int.* **2012**, *48* (2), 848–855. https://doi.org/10.1016/j.foodres.2012.07.001.

(72) 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.

(73) 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.

(74) 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.

(75) 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.; 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.; Uhrín, 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. https://doi.org/10.1039/c9fd90045d.

(76) 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.

(77) 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.

(78) 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.

(79) Minor, E. C.; Li, H.; Oyler, A. R.; Swenson, M. M.; Mattson, B. M. Reply to the 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), 497–498. https://doi.org/10.1039/C4EM00676C.

(80) 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.

(81) 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.

(82) 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.

(83) 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.

(84) Patriarca, C.; Balderrama, A.; Može, M.; Sjöberg, P. J. R.; Bergquist, J.; Tranvik, L. J.; Hawkes, J. A. Investigating the Ionization of Dissolved Organic Matter by Electrospray. *Anal. Chem.* **2020**, *92*, 14210–14218. https://doi.org/10.1021/acs.analchem.0c03438.

(85) 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.

(86) 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.

(87) 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.* **2020**. https://doi.org/10.1038/s41587-020-0740-8.