Utilization of Synthetic Oxygen-Functionalized, Carboxylate-Rich Alicyclic Molecule Analogues in Liquid Chromatography and Tandem Mass Spectrometric Analysis of Dissolved Organic Matter

Jeffrey Hawkes[1], Agnes Flygare[1], Lindon W. K. Moodie[2], Alexander J. Craig[1,2*]

[1]Department of Chemistry BMC, Uppsala University, Uppsala 752 37, Sweden

[1]Department of Medicinal Chemistry, Uppsala University, Uppsala 752 37, Sweden

*Corresponding author: alexander.craig@uu.se

Keywords

Dissolved organic matter, mass spectrometry, liquid chromatography, tandem mass spectrometry, chemical synthesis

Synopsis

Here, we have prepared a series of carboxylate rich alicyclic molecules with a variety of oxygen-based functionalities, to compare their analytical data against that of dissolved organic matter (DOM) standards. Liquid chromatography and tandem mass spectrometry data highlights how retention time and fragmentation data can be used to understand how differently eluting DOM formulas behave.

Abstract

Dissolved organic matter (DOM) is one of the most complex chemical mixtures known, with its chemical composition long puzzling biogeochemists. Identifying the chemical structures within DOM is essential for unraveling its origins and environmental fate. However, DOM's complexity has impeded structural elucidation, with accurate functional group compositions from recalcitrant DOM poorly represented in the synthetic and isolative literature. Consequently, hypothesized DOM compounds are derived from models that inadequately represent true structures. To address this, carboxylic acid-only CRAM analogues were previously synthesized but failed to replicate the extensive fragmentation observed in marine DOM during tandem mass spectrometry (MS2).

Here, we prepared CRAM analogues with numerous oxygen-functionalities to enable more diverse fragmentation pathways. Liquid chromatography studies showed functional group composition better predicted LC polarity than O/C ratio, and that alcohols represented early-eluting DOM profiles, while ethers ketones and lactones better represented central isomers. MS2 studies revealed α-hydroxy ketones and 1,2-diols led to the greatest backbone fragmentation, but remained less extensive than that of DOM. Ether and ester functionalities were labile even at low fragmentation energy, suggesting that such groups are likely contributors to core marine DOM carbon backbones, and contribute to the extensive fragmentation observed in all MS2 experiments.

Introduction

Dissolved organic matter (DOM) is one of the largest reservoirs of organic carbon on the planet, and plays a vital role in nutrient transport between aquatic microorganisms. Composed predominantly of small organic molecules derived from the decomposition of life, DOM represents one of the most chemically complex mixtures known. This molecular diversity has long impeded efforts to accurately characterize the chemical structures it contains, limiting our understanding of its origins and environmental role.[1] Particularly enigmatic is the predominantly marine recalcitrant DOM (RDOM), which persists for millennia and exhibits remarkable physical, chemical, and biological stability.[2] In contrast, the labile DOM (LDOM) fraction is rapidly turned over within hours to days, and primarily consists of transient biological metabolites.[3] While metabolites have been directly identified within

LDOM,[4–8] the functional group composition of RDOM is unlike compounds known from biosynthetic pathways, presenting a significant challenge when it comes to unravelling its structural composition, and ultimately its source and fate.

Efforts to constrain the chemical classes of RDOM have led to descriptions such as carboxylate-rich alicyclic molecules (CRAM),[9] material derived from linear terpenoids (MDLT),[10] acetyl- or hetero-polysaccharides,[9,11,12] and poly-carboxylic acid polyaromatic molecules.[13] However, these classifications are primarily based on the nuclear magnetic resonance (NMR) and high-resolution mass spectrometry (HRMS) data of bulk DOM. While NMR and HRMS (including tandem MS (MS2)) are the gold standards for small-molecule structural analysis, their application to the complex mixtures found in DOM presents significant challenges. NMR spectroscopy, even with advanced 2D techniques, provides broad and poorly differentiated regions for DOM.[14,15] Similarly, HRMS and MS2 techniques struggle to differentiate isomers with identical molecular formulas, even when paired with liquid chromatography (LC) or ion mobility.[16–18] Consequently, the proposed structural classes reflect averaged representations of dominant functional groups and carbon types, rather than precise depictions of functional group arrangements or backbone connectivity. To accurately determine RDOM fluxes and constrain its chemical origins – whether lignin,[19,20] terpenoid,[9,10,21] or polysaccharide-derived[11,12] – greater insight into these specific structural features is critical.

Previously, we disclosed compounds (Figure 1a, **1**-**4**) that mimicked the key features of hypothesized CRAM, including fused alicyclic rings, multiple carboxylic acids, and a predominantly reduced carbon backbone.[22] Subsequent higher-energy collisional dissociation (HCD) MS2 studies revealed distinct fragmentation patterns, with low-energy fragmentation leading predominantly to losses of H_2O and CO_2, consistent with the carboxylic acid functionalities observed in both isolated compounds and DOM. However, high-energy HCD fragmentation caused extensive breakdown of DOM backbone structures, while the carbon scaffold of our decalin backbone CRAMs remained largely intact. Notably, these low energy HCD results were similar to several prior collision induced dissociation (CID) MS2 studies in the field.[17,23–25] In both cases, lower fragmentation energies primarily fragment the most labile DOM functional groups, ultimately limiting in-depth structural information, and suggesting that higher energy beam-type fragmentation methods allow for more extensive investigation of the carbon scaffolds within DOM.

Given the extensive high-energy HCD breakdown in DOM, we speculated that varied oxygen functionalities might enable additional points of charge stabilization, which could facilitate further carbon-carbon bond fragmentations.[26–28] Additionally, for almost any given molecular formula in DOM, peaks exist across a broad range of LC retention times, but little defines what constitutes the functional group features of an early or late eluting isomer. To test these factors, we prepared twelve synthetic compounds based on our initial CRAM decalin scaffold which also contain additional oxygen functionalities, including alcohols, 1,2 diols, ketones, α-hydroxy ketones, ethers, lactones, and esters. This allows for direct comparison between these two sets of molecules, but also provides relative comparison between materials where the primary difference is the inclusion of different functional groups. As such, the information gained from LC and MS2 studies is generally transferable to other carbon scaffolds, leading to insights into the relative differences in functional group type and abundance in isomers of natural DOM.

Methods

Extensive details are available for all experimental methods in the Supporting Information (SI).

Synthetic Information

Synthetic methods are described in the SI, along with spectral data for final CRAM analogues (^{1}H, ^{13}C, COSY, HSQC, and HMBC NMR, and LCMS, LCMS2, and charged aerosol detection (CAD) spectra). Where noted, compounds were prepared as diastereomeric mixtures, hindering full NMR assignment. Fortunately, the nature of this work generating extensive MS2 profiles of these compounds helps to reinforce structural assignments, and CAD data helps in defining their minimum purities.

Reference Materials

The reference standards SRFA and TRM-0522 was used as received and dissolved in 5% acetonitrile in water to a concentration of 10 mg/ml for injection by LCMS.

Liquid-Chromatography Mass-Spectrometry

Liquid chromatography was conducted with a Thermo Vanquish UPLC at a flow rate of 0.5 ml/min from 5% to 100% acetonitrile in water with 0.1% formic acid. Mass spectrometry was performed using an Orbitrap Q Exactive (Thermo Fisher), with MS2 experiments being conducted at normalized collision energies of 35 and 75 by higher energy collision dissociation (HCD) experiments using the PRM method.

Results and discussion

Synthetic Design

Molecules were designed based on the parameters Hertkorn *et al* described for hypothesized CRAM,[9] which are fused alicyclic compounds with reduced carbon backbones functionalized with several carboxylic acids, which we also used to design our previously disclosed CRAMs that were functionalized only with carboxylic acids (hereafter referred to as COOH-CRAM). Similarity of any new compounds to the first set was highly desirable, as while this specific decalin type structure could be poorly represented in real DOM samples, *relative* comparisons between different functional group composition and orientation are likely to be transferrable between carbon scaffolds. In addition to this, ease of synthetic access was at a high priority, as the required hydrolysis to CRAM products from ester equivalents was expected to be more problematic with the inclusion of other potentially sensitive functionalities. As such, targets were based predominantly on the modification of the ester equivalent of alkene CRAM **1** (Figure 1a) where possible, minimizing any potential differences in relative acid regio- and stereochemistry, and allowing for bulk modification of a readily accessible starting material.

Choice of oxygen group incorporation centered on functionalities that have been described within RDOM chemical classes in the literature, as well as those consistent with the 1D and 2D NMR spectral data of DOM. Alcohols were ideal candidates, described as part of CRAM,[20] MDLT,[10] and HPS,[11,12] and also frequently identified as contributors from NMR experiments. Similarly, ketones have been suggested as minor contributors to CRAM, and while they typically display little or no contribution to the ^{13}C NMR spectra of marine DOM, they are present in riverine samples.[9,15,20,29,30] Ethers and esters seemed straightforward inclusions,[31,32] with their presence in multiple biopolymers known to contribute to DOM (i.e. tannins and lignin), identification of methyl ethers and longer alkyl-ethers in marine samples,[15] and the virtual indistinguishability of alcohols from ethers and carboxylic acids from carboxylic esters in ^{1}H and ^{13}C spectra as complex as those of DOM. While common in terrestrial and riverine DOM, phenol and other aromatic functionalities were excluded, as the requirement for several atoms per functionality and low NMR contribution of aromatics in marine DOM limit the number of aromatic molecules than can contribute to most molecular formulas within samples.

Thus, compounds **5**-**16** were synthesized (SI pages 9-27) in addition to the previously obtained COOH-CRAMs **1**-**4**. Di-acids **5** and **6** were included to allow for comparisons between O/C ratios and functional group compositions, as well as comparisons between di, tri, and tetra carboxylic acid compounds.

Lactone **7**, isolated as a by-product from the SPE extraction of a large-scale preparation of **1**, provides a cyclic ester with an identical molecular formula to **1**. It should be noted that this compound was isolated using preparative high-performance liquid chromatography (preparative HPLC), and exists as two closely eluting diastereomers. Mono-alcohols **8** and **9** are regioisomers of one molecular formula ($C_{13}H_{18}O_7$), providing data for how small positional changes can affect LC and MS2 experimental outcomes. Of note, the tertiary alcohol **9** was prepared directly from **7** and as such is a much simpler diastereomeric mixture than secondary alcohol **8**.

Ether **10**, ketone **11**, and α-hydroxy ketone **12** are the sole regioisomers of their respective class, with the positional change in oxygen functionality of ether **10** relative to the other compounds coming as a result of synthetic challenges. It should be noted that α-hydroxy ketone **12** proved relatively unstable in both water and methanol over the course of 24-72 hours, and should be considered more useful as the combination of an alcohol and ketone in LC and MS2 experiments, than suggestion that specific α-hydroxy ketones are suitable long-term stable DOM candidates. The final set of compounds are diols **13**-**16**. Diol **13** is the only O_8 diol, and was isolated through preparative HPLC due to challenges with extraction into organic solvents during reaction work-up. Diols **14** are regioisomeric compounds that were inseparable in prior synthetic steps. Diols **15** and **16** exist as part of one mixture, where some portion of the esters represented in **16** were unhydrolyzable before increasingly harsh conditions led to decomposition of the reaction mixture. However, extracted ion chromatograms and LC and MS2 studies allow for simple differentiation in the context of this work, and the incorporation of a linear ester functionality (as opposed to the cyclic lactone **7**) was seen as a useful comparison.

Retention time investigation

Liquid chromatography (LC) was performed using a 5-100% ACN gradient in H_2O with 0.1% formic acid over 10 minutes, allowing assessment of retention time as a measure of polarity. It is important to note that the different molecular formulas corresponding to the newly synthesized compounds have expected differences in retention profiles within DOM, and importantly, the extracted ion chromatogram for a given formula can vary drastically from the total ion chromatogram. For this reason, we present results in the context of the distribution of retention times for the same molecular formula in TRM (Figure 1b, SI page 28). We use a novel metric, percent cumulative intensity (%CI, Figure 1c), which is determined as the percentage of summed intensity that has eluted for the equivalent molecular formula in TRM at the retention time of the compound (Figure 1d, e). %CI was also calculated for SRFA for comparison (SI page 28).

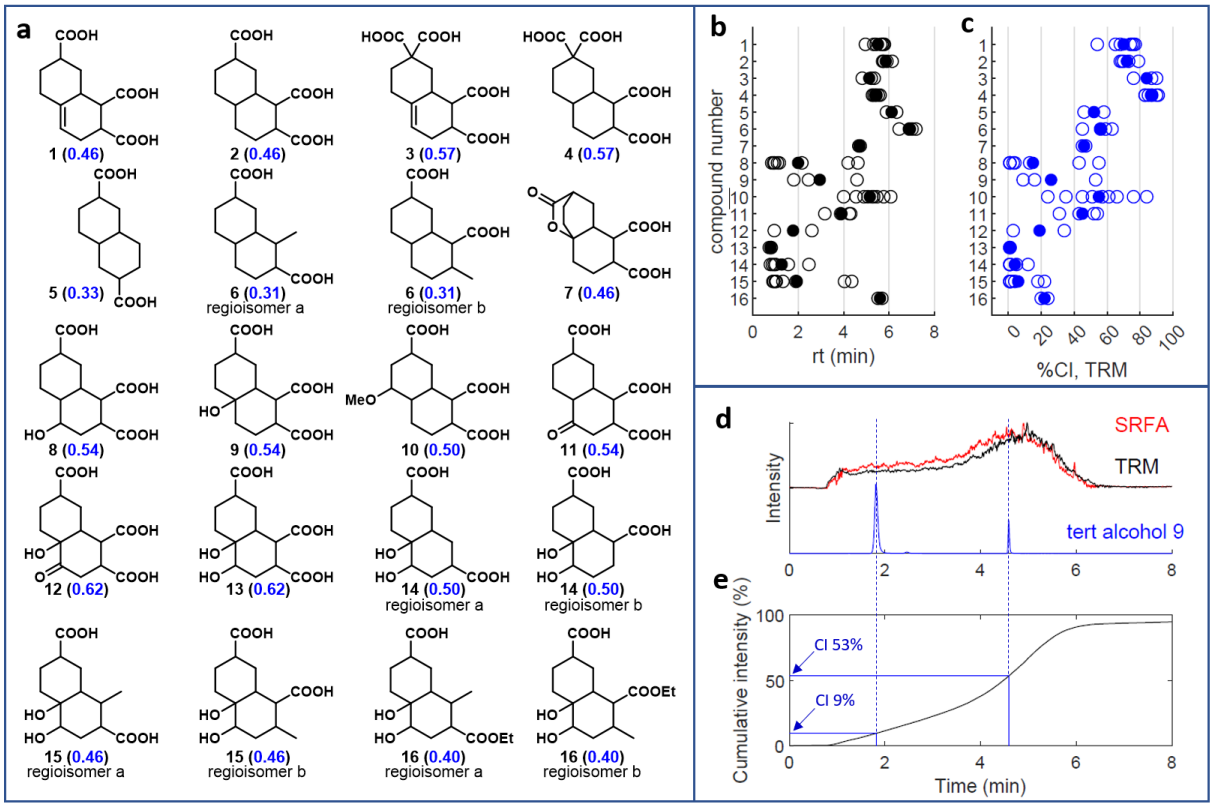

Figure 1: a) *Structures of CRAM analogues with O/C ratios listed in blue **1**-**16**. **b)** Compound retention times. **c)** Compound %CI plotted by compound number. All detected isomers are shown as unfilled circles, mean of detected isomers are shown as filled circles. **d)** Depiction of the %CI metric, XIC of m/z 285.098 in TRM (black), SRFA (red) and the tertiary alcohol **9**. **e)** Trace of cumulative intensity for TRM as a black line, which eventually reaches 100%, and shows the equivalent points at which the two most intense isomers of **9** elute, so that the %CI value can be read from the y axis.*

To first provide a benchmark for relative elution profiles, we examined the retention times of our COOH-CRAMs **1**-**6**. Changing the number of acids on this scaffold made little difference in absolute retention time, with all isomers of di-acid **5**, methyl-diacid **6**, triacid alkene **1**, triacid alkane **2**, tetraacid alkene **3**, and tetraacid alkane **4** eluting between 4.82 and 6.14 minutes (Figure 1b) (i.e. over less than a 15% change in ACN concentration). These tri- and tetra-acids elute late in their corresponding %CI, with compounds **1** and **2** eluting between 63% and 87% CI, and compounds **3** and **4** eluting between 76% and 91%. Conversely, diacids **5** and **6** eluted between 45% and 63% CI, indicating that for lower O/C ratios COOH-CRAM exhibit profiles more like the center of their corresponding main %CI peak. While O/C ratios have previously been used to indicate or characterize relative hydrophobicity and hydrophilicity in DOM LC studies,[33–35] this result indicates that in isolation they are a poor indicator of retention time in reverse-phase LC (RPLC). Where a more oxygen rich formula in DOM (i.e. SI Figures S5 and S13, pages 31 and 33) on average elutes earlier than a less oxygen rich formula (i.e. SI Figure S6 and Figure 15, pages 32 and 34), there is still significant overlap between these elution profiles. The relatively small change in retention of COOH-CRAMs **1**-**6**, but relatively large difference in O/C ratio (0.31-0.57), shows that functionality needs to be considered prominently when considering diverse retention times. We believe that the use of standards such as those employed here offer a future avenue to determine which functional group compositions elute early or late within LC experiments, for example in conjunction with metrics such as %CI.

To explore this idea further, we examined the alcohol, ketone, and diol analogues of CRAM tri-acid **2**. The inclusion of an alcohol in the structures of **8** and **9** significantly decreased retention time relative

to the parent tri-acid **2**, with the earliest isomers eluting very early at 1% CI for secondary alcohol **8**, the latest eluting by 55% for tertiary alcohol **9**, the average at 15% for **8** and 26% for **9**. Similarly, the inclusion of a ketone in compound **11** significantly decreased retention times (first isomer = 31% CI, last = 54%, average = 45%). Finally, the incorporation of a 1,2 diol in compound **13** led to extremely LC polar compounds, with all isomers eluting in the first 1% of CI. These **13** isomers provide a close O/C comparison to tetraacids **3** and **4**, with all three compounds containing eight oxygen atoms, and diols and acids bearing thirteen and fourteen carbon atoms respectively (O/C ratios; **13** = 0.62, **3**/**4** = 0.57).

To reinforce that functional group composition was a stronger indicator of retention that O/C ratio, two additional diols **14** and **15** were prepared. These compounds contained six oxygen atoms and either twelve or thirteen carbon atoms respectively, allowing for more direct comparison to triacids **1** and **2** ($C_{13}O_6H_{16}$ and $C_{13}O_6H_{18}$ respectively). The isomers of **14** and **15** also had very low %CI values, with 11 out of 14 isomers eluting within 5% CI, and the latest eluting at 12% for **14**, and 22% for **15**. The lower O/C ratios of these compounds compared to diol **13** (**14** = 0.5, **15** = 0.46, **13** = 0.62), but only slightly increased average CIs (**14** = 4%, **15** = 6%, **13** = 1%), reinforce that functional group composition is the strongest indicator of retention time in LC. Finally, α-hydroxy ketone **12** was prepared to examine a molecule containing both a ketone and an alcohol. This compound again eluted early relative to its tetra-acid counterparts (isomer 1 = 3% CI, isomer 2 = 34%, average = 19%).

Ultimately, these alcohol and ketone bearing compounds exemplify the relatively strong effect of functional group composition vs molecular formula composition and O/C ratio. However, even a single alcohol provides significant LC polarity increases, indicating that switching a carboxylic acid functionality for two alcohols is not a functional means to explain the majority of observed isomers in mixtures like TRM. While ketones appear initially as strong candidates for mid-eluting isomers, their ^{13}C contributions to terrestrial DOM are low (as low as 1 in 38 C atoms),[20] and in marine DOM are virtually non-existent (as low as 1 in 110 C atoms)[9,15] and, in our opinion, are often integrated from noise in marine DOM. As such, they can explain some portion of observed isomers, but additional functionality must be required to describe the retention profile of any specific molecular formula within DOM. With a lack of aromatic functionality in marine DOM (six C atoms are required, between 1 in 20 and 1 in 25 C atoms are aromatic across several marine depths, as such at most 1 in 20 molecules contain an aromatic functionality), ethers and esters remained as the last prominent oxygen functionalities prevalent in all types of DOM.

Thus, ether **10** and lactone **7** were the next compounds investigated. Ether **10** showed a broad diversity in polarity between isomers, with CI's between 24% and 84%, with an average of 55%. This was particularly notable to us, as it allows for a range of retentions lower than that of the COOH-CRAMs **1**-**6**, while not displaying the same propensity for the drastic retention time reduction seen in alcohols **8**-**9** and **12**-**16**. Furthermore, regions that can indicate ether functionalities are highly prevalent in the NMR spectra of terrestrial DOM and RDOM, with ^{1}H NMR chemical shifts between 2.9-4.1 ppm, and ^{13}C NMR chemical shifts between 47-90 ppm containing predominantly alcohols, ethers, and esters, while methyl ethers have been shown to be prevalent in some marine DOM samples.[15,36] Conversely, less variety in RT was observed for lactone **7** in comparison to ether **10**. However, only two isomers of **7** were produced (vs. nine for **10**), which limits broad conclusions.

The comparatively reduced retention time of **7** was somewhat surprising, as both triacid alkene **1** and lactone **7** share the same molecular formula ($C_{13}O_6H_{16}$), but all isomers of **7** are more polar than all isomers of **1**. Our initial expectation was that the additional acid of alkene **1** would provide a greater capacity for H-bonding, and thus decreased retention time. However, it appears that at least for the decalin scaffold examined in this paper, that the incorporation of a cyclic ester can lead to decreased retention time relative to its corresponding acid equivalent. This may be due to a conformational

difference, as the extra cyclisation in the lactone moves the hydrophobic regions toward the center of the molecule, and positions the oxygen atoms more externally. One additional compound bearing an ester did exist within our dataset – however, it was obtained as an impurity during the hydrolysis to form diol **15**. Ester **16**, as such, contains an acid, two alcohols, and an ethyl ester, with isomers eluting at 5.52 and 5.69 minutes. It is notable that the absolute retention time of this compound is significantly higher than the other diols examined, indicating that carboxylic acid-alcohol intramolecular interactions may play a key role in their very low LC retention times. However, they still follow the same relative CI trend of all diols, with isomers of **16** eluting at 20% and 24% compared to the equivalent CI of TRM.

MS2 Investigation

Fragmentation studies were performed using higher energy collisional dissociation (HCD) at 35 V (low energy) and 75 V (high energy). Additional fragmentation metrics are provided (SI, pages 29-30) to reinforce the qualitative trends discussed here through qualitative observation of various diagnostic fragment peaks. For COOH-CRAMs **1-4**, we previously showed that low energy fragmentation led predominantly to neutral losses of H_2O, CO_2, and minor losses of CO ('functional group' fragments), as well as trace 'fingerprint' fragments from decomposition of the carbon backbone. High energy fragmentation provided high intensities of functional group fragments, as well as slightly increased fingerprint fragmentation for alkanes **2** and **4**, and moderate fingerprint fragmentation of alkenes **1** and **3**. Added to this initial set of acid-only compounds, diacids **5** and **6** show comparatively decreased fragment intensity at both low and high energy, with parent ions being the dominant peak at both low and high energy (Figure 2a; shown for **5**). This is in stark contrast with our previously disclosed CRAM-analogues, none of which retained their parent ions under high energy fragmentation, and likely suggests that intramolecular proximity of functionalities to one another (Figure 2b) promotes the sequential loss of H_2O and CO_2 typically associated with the fragmentation of DOM.[28,37,38]

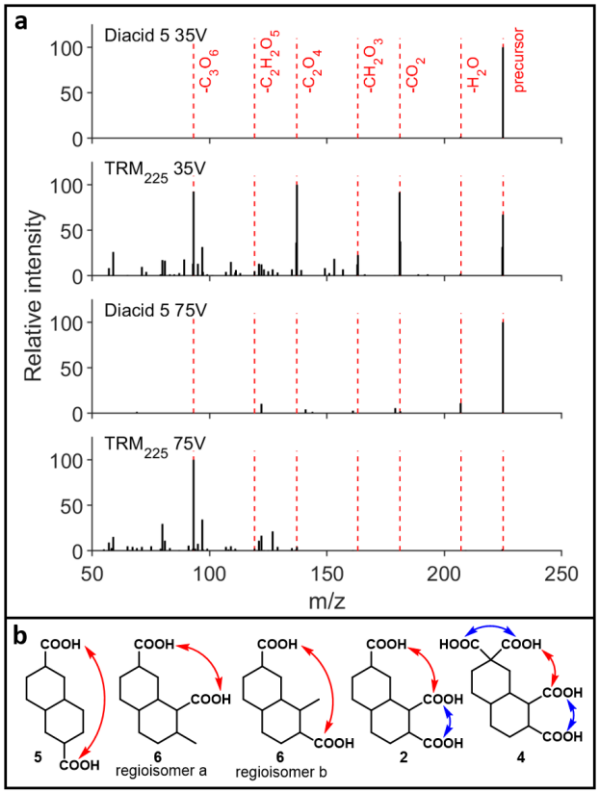

Figure 2: a) *HCD fragmentation data of **5** at a) 35 and b) 75 V, vs same mass in TRM (35 (c), 75 V (d).* ***b)*** *Relative proximity of acids **5** vs **6** vs **2** vs **4** (red = far, blue = near).*

Comparison of acids **5** and **6** with TRM highlights similar trends to the previously examined acids **1**-**4**.[22] In DOM, an increased number of neutral H_2O and CO_2 losses are observed during the low energy fragmentation of the same masses, alongside moderate fingerprint fragmentation (Figure 2a). When fragmented at high energy, DOM shows very low or no intensity of functional group fragments, and instead undergoes extensive fragmentation to fingerprint ions. As we previously highlighted,[22] it is important to note that utilizing an orbitrap for these types of fragmentations leads to the isolation of ions across a single unit mass, and it is possible that the diversity of fragments seen here arises from the fact that several different parent isomers are being fragmented. Nevertheless, the stability of the parent ion and functional group fragments for diacids **5** and **6** reinforces that only carboxylic acid functionalized compounds do not fragment the same way as DOM.

For triacid alcohols **8** and **9**, low energy fragmentation showed the sequential neutral losses of two CO_2 and one H_2O (Figure 3). Similarly, at low energy, diacid diols **14** and **15** showed sequential losses of up to two CO_2 and up to two H_2O (SI pages 50-51), while triacid diols **13** showed losses of up to three CO_2 and up to two H_2O (Figure 4). Pleasingly, fragmentation at high energy for alcohol **8** and diols **13**-**15** led to the observation of extensive fingerprint fragmentation (**8** and **9** in Figure 3, **13** in Figure 4; **14**, **15** in SI pages 50-51). However, it is important to note that for all of these compounds, several functional group fragments remained, with prominent ions in every high energy fragmentation corresponding to the decalin backbone after all but one oxygen atom has fragmented (ketene anion **i/ii** in Figure 3b; oxy-anion **iii/iv** in Figure 4b). For tertiary alcohol **9**, fragmentation was less extensive (Figure 3a). While this could be due to the relative instability of a formal or partial negative charge at a tertiary center relative to a secondary center, investigation of the well-resolved individual isomers of alcohol **8** showed significantly fewer fragments compared to the total HCD spectrum. As such, the relative simplicity of the high energy fragmentation of tertiary alcohol **9** is most likely due to its diastereomeric 'simplicity' in comparison to secondary alcohol **8**.

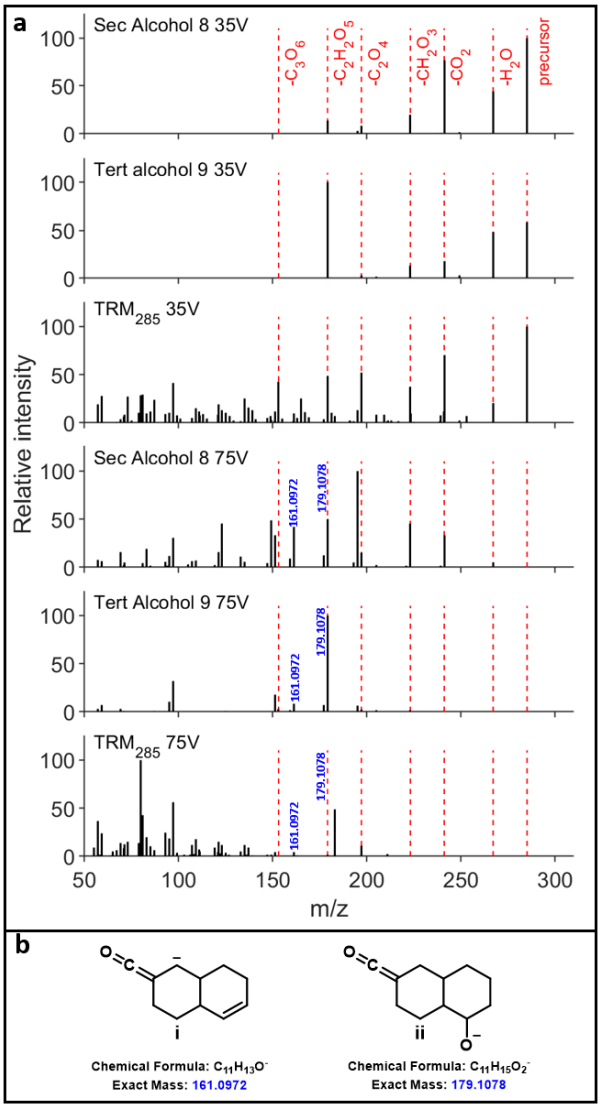

Figure 3: a) *top: 35V HCD fragmentations of sec alcohol* ***8****, tert alcohol* ***9****, TRM285, bottom: 75V HCD fragmentations of sec alcohol* ***8****, tert alcohol* ***9****, TRM285.* ***b)*** *idealized functional group fragments (****i/ii****) for alcohols* ***8*** *and* ***9****.*

Turning to comparison of the alcohols and diols to DOM at *m/z* 285, similar patterns emerge as for the COOH-CRAMs. At low energy, DOM exhibits the same functional group fragments, while undergoing significantly greater fingerprint fragmentation across all masses corresponding to alcohols **8**-**9** and diols **13**-**15** (SI, pages 42-43, 48-51). This is true even for triacid diol **13**, which alongside ketone **11,** α-hydroxy ketone **12** and diol **16** underwent the most extensive fragmentation of all our synthetic compounds (SI, page 29). At higher energy, functional group fragments are again almost entirely lost for all alcohol and diol masses in TRM (*m/z* 285 and 301). This suggests that while the incorporation of these additional oxygen functionalities may promote the increased fragmentation seen in natural samples, it does not enable the breakdown extent seen in DOM, highlighted by the presence of decalin-type fragments of 161.0972 and 179.1078 for **8** and **9** that are essentially absent in DOM (Figure 3a). As the LC data suggested that alcohols are likely only present as relatively early eluting isomers of any specific formula (vide supra), it was also examined whether fragmentation data corresponding to later LC retention led to greater conservation of higher mass fragments. However, investigation of three separate areas within DOM for the mass corresponding to diol **13** showed essentially no variation in average fragment mass, suggesting that other structural features must contribute to this extensive DOM fragmentation (SI page 55).

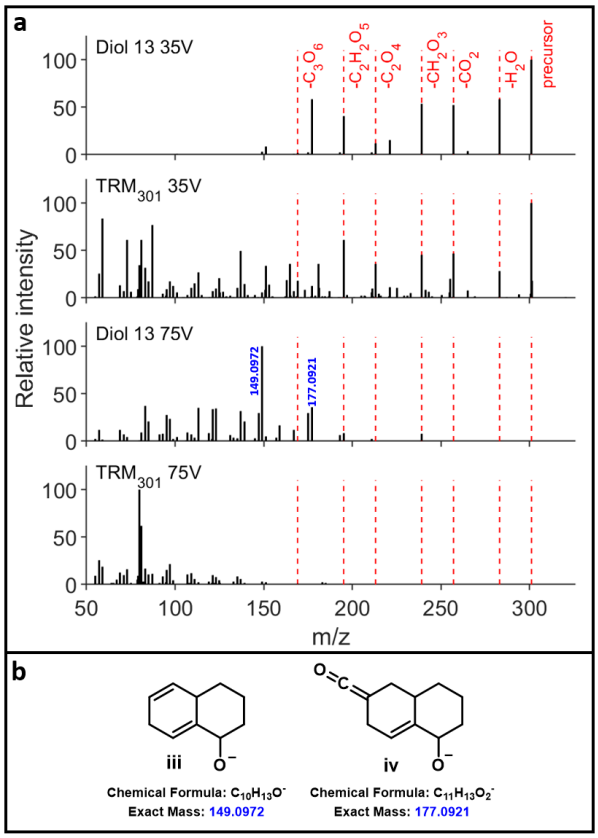

Figure 4) a) *HCD fragmentation of diol **13** and TRM301.* ***b)*** *idealized functional group fragments (**iii/iv**) for diol **13**.*

For lactone **7**, ketone **11**, and α-hydroxy ketone **12**, the same trends as for the other isolated compounds are seen in low and high energy HCD spectra. Low energy spectra show functional group fragments that correspond predominantly to water, CO_2, and CO losses, which are also seen in the low energy fragmentation spectra of TRM (alongside mild fingerprint fragmentation). At high energy, smaller functional group fragments are observed for **7**, **11**, and **12**, while DOM undergoes extensive breakdown to fingerprint fragments. It should be noted that ketone **11** emerged as one of the best matches to DOM based on fragmentation metrics (SI page 29). It demonstrated one of the most comparable extents of fragmentation to TRM, and exhibited the highest number of fragmentation peaks observed in the same experiment. This adds to the relative consistency of the retention profile of **9** to DOM (vide supra). However, it should be stressed that the relatively strong retention of these functional group fragments of **11**), and relative lack of them in the same nominal mass in TRM (SI page 45-46, limits how much decalin ketones such as **11** are likely to contribute to marine mixtures.

For ester acid diol **16**, the majority of functional group fragments seen in both low and high energy HCD fragments are absent within DOM (35V HCD shown in Figure 5a). However, it should be noted that early within the fragmentation spectra of **16**, a neutral loss of C_2H_4 (ethene) corresponding to fragmentation of the ester is observed immediately from the parent ion during low energy HCD fragmentation, and from the first H_2O loss fragment (299.1508 to 271.1189, 281.1379 to 253.1081 Figure 5a, idealized fragmentation in Figure 5b). Subsequent fragments are more typical, corresponding to losses of CO_2, H_2O, and CO. Perhaps more enticingly, ether **10** shows neutral losses of CH_3OH (255.1239 to 223.0978 in 35V HCD shown in Figure 5a, idealized fragmentation in Figure 5b) during low energy HCD fragmentation, observable in TRM, with subsequent losses to 205.0872 ($-H_2O$) and 161.0974 ($-H_2O$, CO_2) and reliant on the initial loss of CO_2 and CH_3OH.

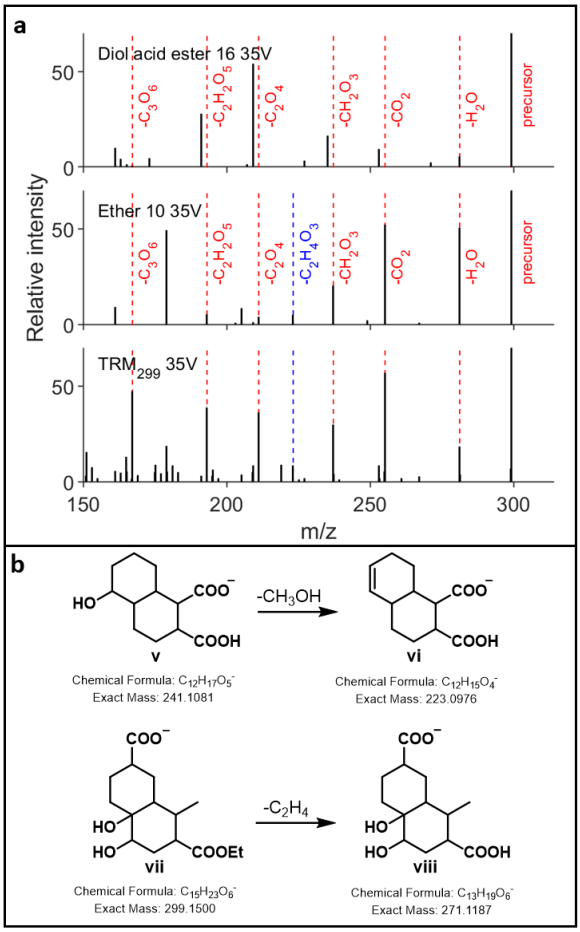

Figure 5: a) *35V HCD spectra of diol acid ester (top) ether (mid) correpsonding DOM mass 299 (bottom), blue dashed line indicates CO_2 and CH_3OH loss.* ***b)*** *Idealized CH_3OH losses of ether* ***10*** *(a to b), ethene losses of ester* ***16*** *(c to d).*

The presence of these methyl ethers in TRM is not unexpected, with this functionality existing as a clear derivative of lignin and tannin type structures, and their presence being documented at several ocean depths from both 1D and 2D NMR techniques.[15] However, we believe that the neutral loss of CH_3OH for ether **10**, as well as the neutral loss of ethene for ester **16** is an important indicator as to why the fragmentation spectra of our synthetic molecules and DOM vary so markedly, especially at higher energies. The extensive fingerprint fragmentation of DOM in HCD experiments suggests that molecules within DOM must have backbones that break into significantly smaller fragments than the fused alicyclic structures presented here. Here, the incorporation of esters and ethers into core carbon backbones would remedy this lack of fingerprint fragmentation for fused alicyclic structures. Their presence could enable ring-opening type fragmentations (Figure 6a), or the possibility for small fragments linked through ethers or esters to cleave from larger structures (Figure 6b), allowing for the generation of the diverse and plentiful fingerprint fragments observed in natural DOM samples. To exemplify this, masses corresponding to small $C_7H_xO_y$ fragments (ix, x, xi, Figure 6c, ketenes are $-H_2O$ loss products of carboxylic acids) that were seen in all HCD75 spectra gathered in TRM regardless of parent ion mass (exemplified in Figure 6d, visible in SI pages 39-52 for all masses) have been imagined as idealized breakdown products of these pathways.

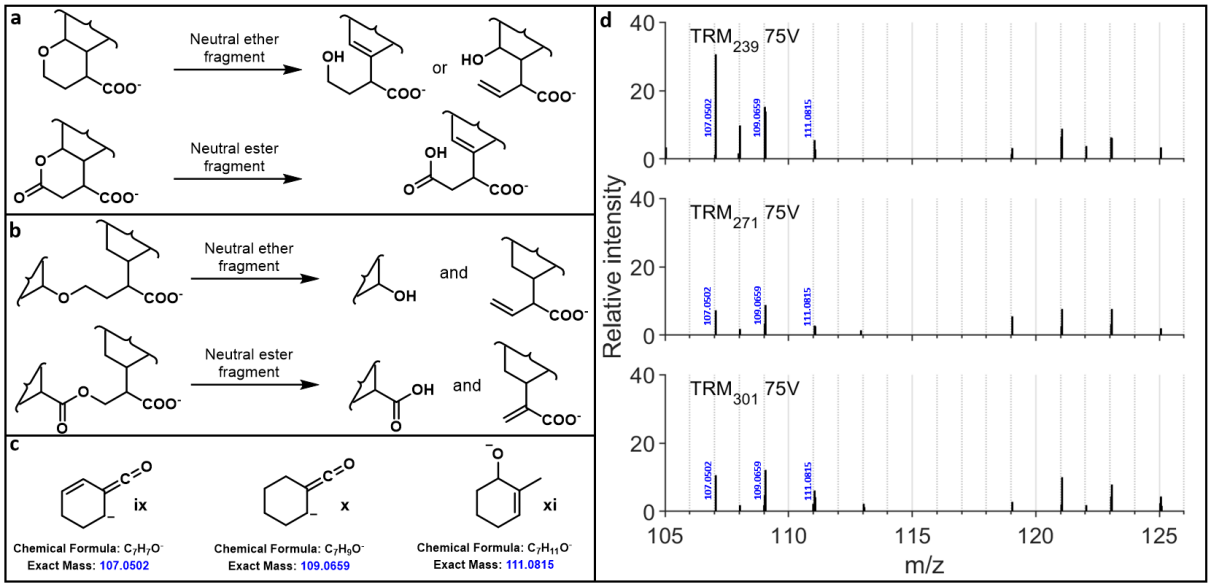

Figure 6: a) *hypothetical DOM cyclic ester and ether fragmentations;* ***b)*** *hypothetical DOM linear ether and ester fragmentations;* ***c)*** *idealized DOM fragments from masses observed in all TRM HCD fragmentation spectra gathered in this work;* ***d)*** *zoomed HCD75 fragmentation spectra showing masses 107.0502, 109.0659, and 111.0815, repeating mass set from 119 to 125 m/z also shown.*

Environmental Implications and Conclusions

Twelve CRAM analogues with varying oxygen functional groups and O/C ratios were prepared and analyzed by LCMS and MS2, for comparison with representative DOM samples. LC analysis found that adding alcohol groups significantly increased the polarity of CRAM-like compounds. As a result, in natural samples, alcohol-containing compounds are likely to appear as early-eluting isomers for a given molecular formula. This effect was especially strong for 1,2-diol compounds, showing that replacing an acid group with two alcohol groups, and adding a C-C double bond is unlikely to explain the full diversity of molecular isomers found in DOM for a given formula. Instead, functional groups like ketones, ethers, and esters provided better matches across the full range of relative retention times observed in natural DOM mixtures. The relative lack of ketone functionalities in marine DOM NMR spectra suggests that ethers and esters are the best functional groups to account for retention time diversity of CRAM-like molecular formulas in RDOM. Additionally, we have developed a new metric, % cumulative intensity, that is useful for the comparison of retention times of isolated compounds to DOM, and believe should be useful for documenting and comparing retention times been individual DOM molecular formulas and between natural samples.

MS2 data revealed similar functional group fragmentation patterns between the newly synthesized compounds and representative DOM molecular formulas during low-energy HCD experiments. Data from comparison between di-, tri-, and tetra- carboxylic acids suggests that for reliable functional group fragmentation, proximity between groups is essential. Notably, at low energies, CH_3OH and C_2H_4 losses were observed from an ether and an ester functionalized CRAM, respectively. Similar CH_3OH fragment losses from the ether were also observed in fragmentation the same nominal mass in TRM-0522. High-energy HCD experiments caused extensive fragmentation of DOM into low *m/z* masses, indicative of carbon backbone degradation. In contrast, the synthetic compounds retained fragments of some functional groups even at higher energies, and the extent of further backbone breakdown varied greatly among the synthesized compounds. Among the synthesized compounds, diols and an α-hydroxy ketone broke down into the smallest average fragment masses, most closely resembling DOM

fragmentation. However, these compounds still did not fragment to the same extent as DOM under either low- or high-energy HCD conditions.

While these diols and α-hydroxy ketones were the best fragmentation matches to DOM, they suffer drawbacks when considering retention profiles (diols), or natural ^{13}C NMR data (α-hydroxy ketones). Instead, we propose ethers and esters are excellent candidates to explain the difference in fragmentation data between the synthesized compounds and DOM. They align well with the center of the main reverse-phase DOM LCMS peak, and underwent additional functional group fragmentations at low and high HCD energies that might explain extensive DOM fingerprint fragmentation. A single one of either of these functionalities incorporated into the carbon backbones of CRAM-like molecules would enable more extensive fragmentation than for all-carbon alicyclic systems, at relatively low total abundance within any given DOM sample. The ether and ester functionalities fit well into established NMR data, with specific signals masked by the dominant alcohol and carboxylic acid functionalities of bulk DOM. Future investigations will focus on the preparation of compounds suitable to test whether both fused cyclic ethers and esters, as well as ether or ester linked carbocyclic systems more accurately align with the HCD fragmentation data of DOM.

Acknowledgements

The work was funded by FORMAS (grant number 2021-00543). LM acknowledges the Uppsala antibiotic centre.

Supporting Information: Supporting Information file contains a description of the synthetic methods, expanded general methods, synthetic procedures, tabulated metrics for fragmentation and cumulative intensity data, extracted ion chromatograms for TRM-0522 and SRFA, tandem mass spectrometry data for final compounds, and NMR and LC-MS-CAD data for all synthetic intermediates and final compounds.

Credit system for author contributions

Term	Definition	Initials
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims	AC, JH
Methodology	Development or design of methodology; creation of models	AC, JH
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components	JH
Validation	Verification, whether as a part of the activity or separate, of the overall replication/ reproducibility of results/experiments and other research outputs	AC, JH
Formal analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data	AC, JH

Term	Definition	Initials
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection	AC, AF
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools	JH, LM
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse	AC, JH
Writing - Original Draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)	AC
Writing - Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre-or postpublication stages	AC, AF, JH, LM
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/ data presentation	AC, JH
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team	AC, JH
Project administration	Management and coordination responsibility for the research activity planning and execution	AC
Funding acquisition	Acquisition of the financial support for the project leading to this publication	JH, LM

References

(1) Nebbioso, A.; Piccolo, A. Molecular Characterization of Dissolved Organic Matter (DOM): A Critical Review. *Anal. Bioanal. Chem.* **2013**, *405* (1), 109–124.

(2) Hansell, D. A. Recalcitrant Dissolved Organic Carbon Fractions. *Annu. Rev. Mar. Sci.* **2013**, *5* (1), 421–445.

(3) Moran, M. A.; Ferrer González, F.; Fu, H.; Nowinski, B.; Olofsson, M.; Powers, M.; Schreier, J.; Schroer, W.; Smith, C.; Uchimiya, M. The Ocean's Labile DOC Supply Chain. *Limnol. Oceanogr.* **2022**, *67*.

(4) Yamashita, Y.; Tanoue, E. Chemical Characteristics of Amino Acid-Containing Dissolved Organic Matter in Seawater. *Org. Geochem.* **2004**, *35* (6), 679–692.

(5) Amon, R. M.; Benner, R. Combined Neutral Sugars as Indicators of the Diagenetic State of Dissolved Organic Matter in the Arctic Ocean. *Deep-Sea Res. I: Oceanogr. Res. Pap.* **2003**, *50* (1), 151–169.
(6) Aristilde, L.; Guzman, J. F.; Klein, A. R.; Balkind, R. J. Compound-Specific Short-Chain Carboxylic Acids Identified in a Peat Dissolved Organic Matter Using High-Resolution Liquid Chromatography–Mass Spectrometry. *Org. Geochem.* **2017**, *111*, 9–12.
(7) Xiao, M.; Wu, F.; Liao, H.; Li, W.; Lee, X.; Huang, R. Characteristics and Distribution of Low Molecular Weight Organic Acids in the Sediment Porewaters in Bosten Lake, China. *J. Environ. Sci.* **2010**, *22* (3), 328–337.
(8) Papadopoulos Lambidis, S.; Schramm, T.; Steuer-Lodd, K.; Farrell, S.; Stincone, P.; Schmid, R.; Koester, I.; Torres, R.; Dittmar, T.; Aluwihare, L. Two-Dimensional Liquid Chromatography Tandem Mass Spectrometry Untangles the Deep Metabolome of Marine Dissolved Organic Matter. *Environ. Sci. Technol.* **2024**, *58* (43), 19289–19304.
(9) Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a Major Refractory Component of Marine Dissolved Organic Matter. *Geochim. Cosmochim. Acta.* **2006**, *70* (12), 2990–3010.
(10) Arakawa, N.; Aluwihare, L.; Simpson, A.; Soong, R.; Stephens, B.; Lane-Coplen, D. Carotenoids Are the Likely Precursor of a Significant Fraction of Marine Dissolved Organic Matter. *Sci. Adv.* **2017**, *3*, e1602976.
(11) Aluwihare, L.; Repeta, D. A Comparison of the Chemical Characteristics of Oceanic DOM and Extracellular DOM Produced by Marine Algae. *Mar. Ecol. Prog.* **1999**, *186*, 105–117.
(12) Aluwihare, L. I.; Repeta, D. J.; Pantoja, S.; Johnson, C. G. Two Chemically Distinct Pools of Organic Nitrogen Accumulate in the Ocean. *Science* **2005**, *308* (5724), 1007–1010.
(13) Dittmar, T.; Koch, B. P. Thermogenic Organic Matter Dissolved in the Abyssal Ocean. *Mar. Chem.* **2006**, *102* (3–4), 208–217.
(14) Mitschke, N.; Vemulapalli, S.; Dittmar, T. NMR Spectroscopy of Dissolved Organic Matter: A Review. *Environ. Chem. Lett.* **2023**, *21* (2), 689–723.
(15) Hertkorn, N.; Harir, M.; Koch, B. P.; Michalke, B.; Schmitt-Kopplin, P. High-Field NMR Spectroscopy and FTICR Mass Spectrometry: Powerful Discovery Tools for the Molecular Level Characterization of Marine Dissolved Organic Matter. *Biogeosciences* **2013**, *10* (3), 1583–1624.
(16) Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the Structural Complexity of Dissolved Organic Matter: Isomeric Diversity. *Faraday Discuss.* **2019**, *218*, 431–440.
(17) Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J.; Tranvik, L. J.; Bergquist, J. Extreme Isomeric Complexity of Dissolved Organic Matter Found across Aquatic Environments. *Limnol. Oceanogr. Lett.* **2018**, *3* (2), 21–30.
(18) 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.
(19) DiDonato, N.; Hatcher, P. G. Alicyclic Carboxylic Acids in Soil Humic Acid as Detected with Ultrahigh Resolution Mass Spectrometry and Multi-Dimensional NMR. *Org. Geochem.* **2017**, *112*, 33–46.
(20) Li, S.; Harir, M.; Bastviken, D.; Schmitt-Kopplin, P.; Gonsior, M.; Enrich-Prast, A.; Valle, J.; Hertkorn, N. Dearomatization Drives Complexity Generation in Freshwater Organic Matter. *Nature* **2024**, *628* (8009), 776–781.
(21) Lam, B.; Baer, A.; Alaee, M.; Lefebvre, B.; Moser, A.; Williams, A.; Simpson, A. Major Structural Components in Freshwater Dissolved Organic Matter. *Environ. Sci. Technol.* **2008**, *41*, 8240–8247.
(22) Craig, A. J.; Moodie, L. W.; Hawkes, J. A. Preparation of Simple Bicyclic Carboxylate-Rich Alicyclic Molecules for the Investigation of Dissolved Organic Matter. *Environ. Sci. Technol.* **2024**, *58* (16), 7078–7086.
(23) Zark, M.; Dittmar, T. Universal Molecular Structures in Natural Dissolved Organic Matter. *Nat. Commun.* **2018**, *9* (1), 3178.
(24) 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.

(25) Osterholz, H.; Niggemann, J.; Giebel, H.-A.; Simon, M.; Dittmar, T. Inefficient Microbial Production of Refractory Dissolved Organic Matter in the Ocean. *Nat. Commun.* **2015**, *6* (1), 7422.
(26) Cheng, C.; Yang, M.; Wu, Z.; Wang, Y.; Zeng, F.; Wu, W.; Guan, S.; Guo, D. Fragmentation Pathways of Oxygenated Tetracyclic Triterpenoids and Their Application in the Qualitative Analysis of Ganoderma Lucidum by Multistage Tandem Mass Spectrometry. *Rapid. Commun. Mass. Spectrom.* **2011**, *25* (9), 1323–1335.
(27) Zhang, J.; Feng, E.; Li, W.; Sheng, H.; Milton, J. R.; Easterling, L. F.; Nash, J. J.; Kenttämaa, H. I. Studies of the Fragmentation Mechanisms of Deprotonated Lignin Model Compounds in Tandem Mass Spectrometry. *Anal. Chem.* **2020**, *92* (17), 11895–11903.
(28) Demarque, D. P.; Crotti, A. E.; Vessecchi, R.; Lopes, J. L.; Lopes, N. P. Fragmentation Reactions Using Electrospray Ionization Mass Spectrometry: An Important Tool for the Structural Elucidation and Characterization of Synthetic and Natural Products. *Nat. Prod. Rep.* **2016**, *33* (3), 432–455.
(29) Abdulla, H. A.; Minor, E. C.; Dias, R. F.; Hatcher, P. G. Changes in the Compound Classes of Dissolved Organic Matter along an Estuarine Transect: A Study Using FTIR and 13C NMR. *Geochim. Cosmochim. Acta.* **2010**, *74* (13), 3815–3838.
(30) Mitschke, N.; Vemulapalli, S. P. B.; Dittmar, T. Dissolved Organic Matter Contains Ketones Across a Wide Range of Molecular Formulas. *Environ. Sci. Technol.* **2024**, *58* (35), 15587–15597.
(31) Leenheer, J. A.; Wershaw, R. L.; Reddy, M. M. Strong-Acid, Carboxyl-Group Structures in Fulvic Acid from the Suwannee River, Georgia. 2. Major Structures. *Environ. Sci. Technol.* **1995**, *29* (2), 399–405.
(32) Leenheer, J. A.; Wershaw, R. L.; Reddy, M. M. Strong-Acid, Carboxyl-Group Structures in Fulvic Acid from the Suwannee River, Georgia. 1. Minor Structures. *Environ. Sci. Technol.* **1995**, *29* (2), 393–398.
(33) Stenson, A. C. Reversed-Phase Chromatography Fractionation Tailored to Mass Spectral Characterization of Humic Substances. *Environ. Sci. Technol.* **2008**, *42* (6), 2060–2065.
(34) Sleighter, R. L.; Caricasole, P.; Richards, K. M.; Hanson, T.; Hatcher, P. G. Characterization of Terrestrial Dissolved Organic Matter Fractionated by pH and Polarity and Their Biological Effects on Plant Growth. *Chem. Biol. Techn. Agric.* **2015**, *2*, 1–19.
(35) Li, Y.; Harir, M.; Uhl, J.; Kanawati, B.; Lucio, M.; Smirnov, K. S.; Koch, B. P.; Schmitt-Kopplin, P.; Hertkorn, N. How Representative Are Dissolved Organic Matter (DOM) Extracts? A Comprehensive Study of Sorbent Selectivity for DOM Isolation. *Wat. Res.* **2017**, *116*, 316–323.
(36) Felgate, S. L.; Craig, A. J.; Moodie, L. W.; Hawkes, J. Characterization of a Newly Available Coastal Marine Dissolved Organic Matter Reference Material (TRM-0522). *Anal. Chem.* **2023**, *95* (16), 6559–6567.
(37) Kanawati, B.; Schmitt-Kopplin, P. Exploring Rearrangements along the Fragmentation of Glutaric Acid Negative Ion: A Combined Experimental and Theoretical Study. *Rapid Commun. Mass. Spectrom.* **2010**, *24* (8), 1198–1206.
(38) Grossert, J. S.; Fancy, P. D.; White, R. L. Fragmentation Pathways of Negative Ions Produced by Electrospray Ionization of Acyclic Dicarboxylic Acids and Derivatives. *Can. J. Chem.* **2005**, *83* (11), 1878–1890.

TOC Graphic: