Two-Dimensional Liquid Chromatography Tandem-Mass Spectrometry Untangles the Deep Metabolome of Marine Dissolved Organic Matter

Stilianos Papadopoulos Lambidis[1], Tilman Schramm[1], Karoline Steuer-Lodd[1], Shane Farrell[1,2], Paolo Stincone[1], Robin Schmid[3], Irina Koester[4], Ralph Torres[4], Lihini Aluwihare[4], Carsten Simon[5], Daniel Petras[1,*]

1 University of Tuebingen, Interfaculty Institute of Microbiology and Medicine, Tuebingen, Germany
2 Bigelow Laboratory for Ocean Sciences, East Boothbay, USA
3 Institute of Organic Chemistry and Biochemistry of the CAS, Prague, Czech Republic
4 University of California San Diego, Scripps Institution of Oceanography, La Jolla, USA
5 Helmholtz-Centre for Environmental Research, Leipzig, Germany
*Corresponding Author, daniel.petras@uni-tuebingen.de

Dissolved organic matter (DOM) is one of the most complex chemical mixtures and plays a central role in biogeochemical cycles across our ecosphere. Despite its importance, DOM remains poorly understood at the molecular level. Over the last decades, significant efforts have been made to decipher the chemical composition of DOM by high-resolution mass spectrometry (HRMS) and liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS). Yet, the complexity and high degree of non-resolved isomers still hamper the full structural analysis of DOM. To overcome this challenge, we adapted a two-dimensional (2D) LC approach consisting of two reversed-phase dimensions with orthogonal pH, followed by MS/MS data acquisition and molecular networking. The 2D chromatography approach mitigates the complexity of DOM, enhancing both the quality of MS/MS spectra and spectral annotation rates. Applying our approach to analyze coastal surface DOM from Southern California (USA), we annotated in total more than 600 structures via MS/MS spectrum matching, which was up to 90% more than in iterative 1D LC-MS/MS analysis with the same total run time. Our data provide an unprecedented view into the molecular composition of coastal DOM, highlighting the potential of 2D LC-MS/MS approaches to decipher ultra-complex mixtures.

Key Words: Dissolved Organic Matter, 2D Chromatography, 2D-LC-MS/MS, Tandem Mass Spectrometry, Environmental Metabolomics, Molecular Networking

INTRODUCTION

Dissolved organic matter (DOM) is one of the most complex natural chemical mixtures. Despite its ubiquity and major role in biogeochemical cycling, DOM remains poorly understood on the

structural level. Especially in marine systems, DOM is of fundamental importance, representing the largest and most actively cycling pool of reduced carbon[1–3]. For example, DOM affects nutrient retention[4,5], trace metal complexation[6], long-term carbon sequestration[3], and aerosol formation[7–9]. Missing molecular information, however, hampers our understanding of ecosystem functioning and potential feedback loops under climate change scenarios[1,3,10].

Many studies investigate the molecular composition of DOM by direct injection (DI) ultrahigh resolution mass spectrometry (HR-MS) via Fourier Transform Ion Cyclotron Resonance (FT-ICR) or Orbitrap mass spectrometers. While these techniques allow unprecedented resolution of isobaric molecular detail in DOM, complementary coupled techniques such as tandem mass spectrometry (MS/MS)[11–13], liquid chromatography-mass spectrometry (LC-MS)[14–17], and ion mobility mass spectrometry (IMS)[18,19] all suggest a high degree of non-resolved isomeric complexity. Although prior attempts to resolve this complexity via LC-MS have shown promising results, they typically resolve the higher abundant components of DOM and thus are more suitable for freshly produced metabolites and xenobiotics and less for the diverse refractory components of DOM that are present at lower concentrations. Besides inherent problems of subsampling due to a limited number of possible MS/MS scans per LC-MS/MS run, a central bottleneck is the annotation of the resulting spectra[20]. Here, the use of molecular networking (MN) offers a promising avenue to resolve this complexity[21]. MN allows for the internal comparison of MS/MS spectra based on their spectral similarity as well as the matching of spectra against existing datasets and growing community-based reference libraries[22–24]. This way, large LC-MS-MS/MS datasets can be processed, and similar molecular features can be grouped to propagate chemical class level annotations[20,21,25,26]. By making use of the concept of annotation propagation, comprehensively annotated reference samples could lay the foundation for a detailed exploration of DOM data sets, as reference MS/MS spectra for most compounds present in DOM are still limited[25,27].

DOM analyzed by LC-MS/MS setups suffers from insufficient chromatographic resolution, leading to so-called DOM “mountains”, “humps”, or “unresolved complex mixture” of mass features[15,28,29]. Furthermore, the structural level annotation of DOM often remains less than 10% of detected features[14,20,30]. This reflects not only the insufficient coverage of MS/MS duty cycles and chemodiversity in public spectral libraries, but also the insufficient separation of molecular features, leading to the acquisition of low-quality “chimeric” tandem mass spectra[13,14,31]. To improve the chromatographic resolution of DOM, multidimensional separations via gas chromatography[32,33] and liquid chromatography[29,34] have been attempted, but did not yet lead to high-throughput routines.

Among the available 2D-setups, separation by two reversed-phase LC steps at varying pH have shown to be most effective due to the high peak capacity, orthogonality, and high compatibility of mobile phases with evaporation and electrospray ionization[35–37]. Such 2D setups have been widely used in the proteomics field to improve the identification depth of peptides[38]. For example, Wang and coworkers report a 2.2 and 2.7-fold increase in protein and proteoform identifications, respectively, compared to a traditional 1D LC-MS approach[36]. Similarly, McIlvin & Saito reported an online 2D LC-MS setup that led to a 1.5 - 2-fold higher protein identification rate in marine meta-proteome samples, arguing that the diverse assemblages of organisms and ecological interactions led to a wide array of low-abundance protein features[39]. Combined high pH/ low pH separations for the MS/MS analysis of marine DOM represents a promising application in

environmental metabolomics, as this field faces similar problems as in (meta)proteomics samples, such as high sample complexity and large numbers of rare features. The orthogonality of pH over the two chromatographic dimensions might be especially suitable for DOM, as many compounds contain carboxylic acids or amines, which would drastically change their protonation and thereby their retention to the stationary phase with the different pH.

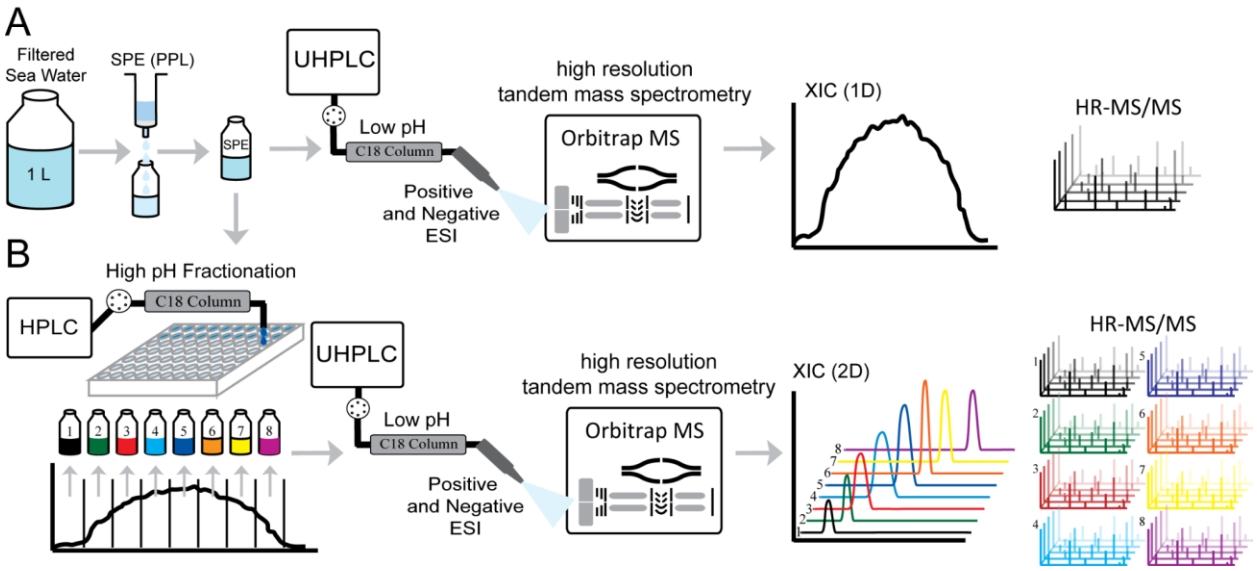

Figure 1: Schematic workflow of 1D and 2D-LC-MS/MS analysis. (A) 1D LC-MS/MS setup. Filtered seawater is extracted using a PPL-cartridge and separated in a UHPLC with a C18 column and low pH mobile phase, before positive and negative electrospray ionization (ESI) and subsequent high-resolution tandem mass spectrometry. Exemplary extracted ion chromatogram (XIC) indicating insufficient separation of isobaric compounds ("DOM-Mountain"). (B) 2D-LC-MS/MS setup. After solid-phase-extraction, samples were separated using an HPLC with high pH mobile phase and a C18 column (1D) and fractionated into multiple fractions. Fractions were further separated in a UHPLC with a low pH mobile phase and a C18 column before subsequent positive and negative ESI into a high-resolution tandem mass spectrometer. Exemplary XICs showing separation of isobaric compounds.

Making use of this concept, we adapted a 2D separation with a first (1D) offline high pH fractionation followed by a second (2D) low pH separation in both negative and positive ESI ionization, followed by data-dependent acquisition (DDA) of MS/MS spectra. Applying and comparing these workflows to our standard 1D LC-MS/MS approaches of a coastal DOM sample, we show that the 2D approach effectively reduces sample complexity while enhancing tandem MS data quality and coverage, leading to a higher coverage and annotation rate of DOM.

MATERIALS AND METHODS

Sample collection and processing. As a DOM reference sample, we collected 200 L of surface seawater in a 20-liter bucket from the Ellen Browning Scripps Memorial Pier (32°52'001.500N 117°15'026.900W) in La Jolla, Southern California, USA, on the 26th of February 2021 between 11:00 and 19:00 PDT. Seawater was filtered through 0.7 µm GF/F filters. 25 L aliquots of the

sample were acidified with 26 mL of hydrochloric acid (37% p.a., trace metal grade, J.T.Baker, pH 2). 25 L of Seawater were then extracted via PPL cartridges with a bed mass of 5 g (Bond Elut, Agilent) according to previous studies[40,41]. Before use, the PPL cartridges were activated with 30 mL of MeOH and then rinsed with 30 mL H_2O (pH 2, LCMS grade) and 30 mL MeOH (LCMS grade) followed by an equilibration with 30 mL H_2O (pH 2, LCMS grade). For sample loading, acidified seawater was pulled through 8 SPE PPL cartridges in parallel. We used a vacuum SPE station (Agilent 20 port SPE station) to maintain a flow rate of approximately 20 mL/min/cartridge for a total loading time of 20 h. A process blank was collected using 4 L of acidified H_2O (pH2, LCMS grade) onto a 5 g cartridge using the same SPE protocol as above.
After sample loading, the cartridges were desalted with 60 mL H_20 (pH 2, LCMS grade) and dried under N_2 gas. After drying, the cartridges were eluted with 20 mL MeOH per cartridge resulting in a total of 200 mL eluted sample to which an internal standard mix was added that contained 5 µg each of domoic acid, kanic acid, isoxaben, irgarol, imazapyr, heroin, methamphetamine and cocaine. The pooled sample was then aliquoted to 100 individual 2mL HPLC vials. All vials were dried down in a vacuum centrifuge overnight at room temperature, resulting in 1.8 mg of total organic matter per vial.

Prefractionation by high pH reversed-phase liquid chromatography. Separation of samples in the 1st dimension was performed using a 1260 Infinity II Semi-preparative HPLC (Agilent, Santa Clara, USA) system with mobile phase A: H_2O + 0.05% NH_4OH and mobile phase B: ACN + 0.05% NH_4OH, with the mobile phases having a pH around 10. For the chromatographic separation, a reversed-phase C18 porous core column (Kinetex EVO C18, 150 x 4.6 mm, 2,6 µm particle size and 100 A pore size, Phenomenex, Torrance, USA) was used. The flow rate was 1 ml/min, and the elution was performed with a linear gradient, from 0-8 min from 95-50 % A, then from 8-10 min 50-1 % A, holding 1 % A for 3 min to washout the column followed by a 3 min re-equilibration phase at 95% A. A total of 250 µL was injected in 10 iterative HPLC runs (each 25 µl) and fractionated using the integrated fraction collector. Fractions were collected in a time-dependent manner into a 96-well deep well plate (Brand, Wertheim, Germany), changing the fraction after every minute, yielding 10 fractions. For an easier workflow, we also collected the first two fractions of the washout, 11 and 12, so that the same fractions are in the same column of the plate. Corresponding fractions between runs were pooled into vials before evaporating the solvent in a speedvac (Concentrator plus, Eppendorf, Hamburg, Germany). The samples were redissolved in 100 µl of 80:20 MeOH:H_2O.

Low pH reversed-phase liquid chromatography tandem mass spectrometry. DOM samples and the fractions (2nd dimension) were separated, using the same LC gradient as in the prefractionation, but different mobile phases, with mobile phase A: H_2O + 0.1% formic acid and mobile phase B: ACN + 0.1% formic acid (pH ~ 2) were used. The eluting molecules were detected by the coupled Q Exactive HF Orbitrap mass spectrometer. DOM constituents were ionized in positive or negative electrospray ionization (ESI) mode. ESI settings were as described previously[14,17]: In short, gas flows (in AU) were set to 52 AU (sheath gas), 14 AU (auxiliary gas), and 0 AU (sweep gas); spray voltage was set to 3.5 kV, inlet capillary was heated to 320 °C; S-Lens level was at 50%. MS1 spectra were acquired with a scan range of 150-1,500 *m/z* at a nominal resolution of 140,000 at *m/z* 200 and automatic gain control (AGC) set to one million

charges. Maximum injection time was set to 100 ms. MS/MS spectra were acquired at a lower AGC target of 300,000 charges with a minimum 10% C-trap filling. Precursors were selected within 1 *m/z* window width. Collision-induced dissociation (HCD) was conducted stepwise at 25, 35 and 45% and at charge state (z) of 1. MS/MS acquisition was triggered at peak apex within 2 – 15 seconds after the first precursor detection. Precursors were then dynamically excluded for 5 seconds, as well as precursors with unassigned charge states or isotope peaks. Retention time drifts were controlled with a quality control mix of six standards evenly distributed in the chromatogram (Sulfamethazine, Sulfamethizole, Sulfachloropyridazine, Sulfadimethoxine, Amitriptyline, and Coumarin-314) at the beginning and the end of the sequence.

LC-MS/MS feature detection and molecular networking. MS1 features were extracted with MZmine 3[42] at a signal threshold of 30,000 and 0.6 seconds minimum peak width. MS1 features were then transformed to extracted ion chromatograms (XICs) at a minimum signal threshold of 90,000 and 5 ppm mass tolerance. Likewise, thresholds for deconvolution were set at 30,000 (baseline level) and 90,000 (minimum peak height). Maximum peak length was 2 min. Isotope peaks were grouped at tolerances of 5 ppm for mass and 0.1 min for retention time. Alignment of XICs between samples was conducted using similar tolerance limits, and only XICs grouped with > 1 isotope peaks were considered for further analysis. The aligned XIC list was checked for duplicates at the above mass (5 ppm) and retention time (0.1 min) tolerances. From MZmine 3 the precursor purity information was exported and used to evaluate the presence of chimeric spectra. Classical Molecular Networking (MN) was used to evaluate the spectra quality and annotate by matching MS/MS spectra with the GNPS libraries[22]. The MZmine3 output quant table and .mgf files were used to create Feature-Based Molecular Networking (FBMN) in GNPS[23]. Detailed settings and urls of processed jobs are provided in Table S1-S4.

In-silico spectrum annotation. Sirius (5.6.3)[43] was used for in-silico annotation of tandem mass spectrometry data. Using the Sirius module, molecular formulas were computed by matching experimental with predicted isotopic patterns from the fragmentation trees analysis of MS/MS. Parameters for SIRIUS are as follows: Instrument: Orbitrap, MS/MS ppm: 5, Isotope scorer: ignore, Candidates stored: 10, Min candidates per Ion: 1, Databases used: no selections, Possible Ionizations: pos/neg, Tree timeout: 0, Compound timeout: 0, Use heuristics above m/z: 300, Use heuristics only above m/z: 650, Only molecular formulas present in the Bio Databases were considered. In-silico structure annotations was performed with CSI:FingerID[44] and structures from the Bio Database.

Visualization of molecular networks. The visualization and analysis of the networks was carried out using Cytoscape (Version 3.9.1)[45]. For that, the obtained table from the feature-based network analysis was thinned out, eliminating redundant SMILES-codes and structural formulas. Nodes represent detected compounds and their drawn chemical structure. Chemical similarity was calculated using Tanimoto similarity (cut-off 0.8), creating a connection between the nodes using the ChemViz add-on.

RESULTS AND DISCUSSION

Improved chromatographic resolution extends the depth of annotated molecules in marine DOM. To improve chromatographic resolution and analytical depth for the molecular characterization of DOM, we developed an offline 2D UHPLC-MS/MS method. Improved chromatographic resolution should lead to cleaner and more unique MS2 spectra, especially in the data-dependent acquisition (DDA) mode used here. In DDA, the compounds are selected for isolation and fragmentation by their intensity information from the previous MS1 scan, so with higher chromatographic resolution, different compounds are selected for isolation and fragmentation. The spectra should be cleaner, as fewer isomers and compounds with similar masses should be present in the quadrupole isolation window (here, 1 *m/z*).

Our workflow relies on the fractionation of the DOM sample with a C18 column with high pH mobile phase with 0.05% NH_4OH as a modifier for the first dimension. Technical duplicates of corresponding fractions were then injected into a second dimension with a C18 column and low pH mobile phase with 0.1% formic acid, before MS/MS data acquisition (20 injections per ESI+ and ESI- mode). To compare the 2D results to our classic 1D LC-MS/MS workflow, we injected the same sample in 20 consecutive LC-MS/MS runs (ESI+ and ESI- mode) yielding the overall same MS/MS analysis time and theoretical maximum number of spectra.

To assess possible improvements in the chromatographic resolution, we first compared the total ion chromatograms (TICs) of a classical 1D LC-MS/MS run and our 2D approach and subsequently compared extracted ion chromatograms (XICs) of representative high abundant masses (470.1653; 500.1753; 579.2068 m/z in pos mode and 381.1195; 395.1351; 411.1301 m/z in neg mode) as shown in Figure 2.

The TIC obtained from the 1D runs (Figure 2 A and B, light gray) shows the aforementioned “DOM-Mountain”, showing poor resolution of the peaks, whereas the TICs obtained from the technical replicates from the 2D runs show more defined peaks that also resemble the fraction order from the first dimension. Importantly, the retention time of the main peaks in all fractions is clearly distinct, indicating a constant and reproducible interaction of the DOM component with the stationary phase and likely less colloidal interactions between DOM itself[29,46–48].

The effect of this sample decomplexation gets more pronounced when looking at specific XICs. In Figure 2, C-E the XICs of three highly abundant masses in positive mode, m/z 470.1653 (+/- 3 ppm (Figure 2 C)), m/z 500.1753 (+/- 3 ppm (Figure 2 D)), and m/z 579.2068 (+/- 3 ppm (Figure 2 E)) are shown. The XIC obtained from the 1D run is shown in gray in the background and the XICs obtained from the 2D run are overlaid. While the 1D XICs show poor chromatographic resolution and a “Mountain-like” shape again, the XICs obtained from the 2D run show at least partially resolved peaks, with detection of molecules with the same m/z in multiple, different fractions. While this effect is more pronounced in the positive mode, clearer peaks and a better resolution are also visible in the negative mode. This suggests improved resolution of isomeric and isobaric compounds by the orthogonal 2D LC, which is crucial to increase MS/MS purity and, ultimately, MS/MS based identification.

Evaluation of Improved Molecular Networking and Spectral Quality. We employed two forms of molecular networking, classical molecular networking (MN) and feature-based molecular networking (FBMN). Briefly, classical molecular networking is a widely used bioinformatics tool

for the visualization and annotation of non-targeted MS data. In addition to spectral matching against reference spectra, it aligns all experimental spectra of a data set against one another and connects related molecules by their spectral similarity. This is achieved by clustering MS2 spectra and using the obtained consensus MS2 spectra for MN[49]. We employed classical molecular networking using the GNPS infrastructure and created classical molecular networks either from the 20 2D runs or from the 20 1D runs (Figure 3).

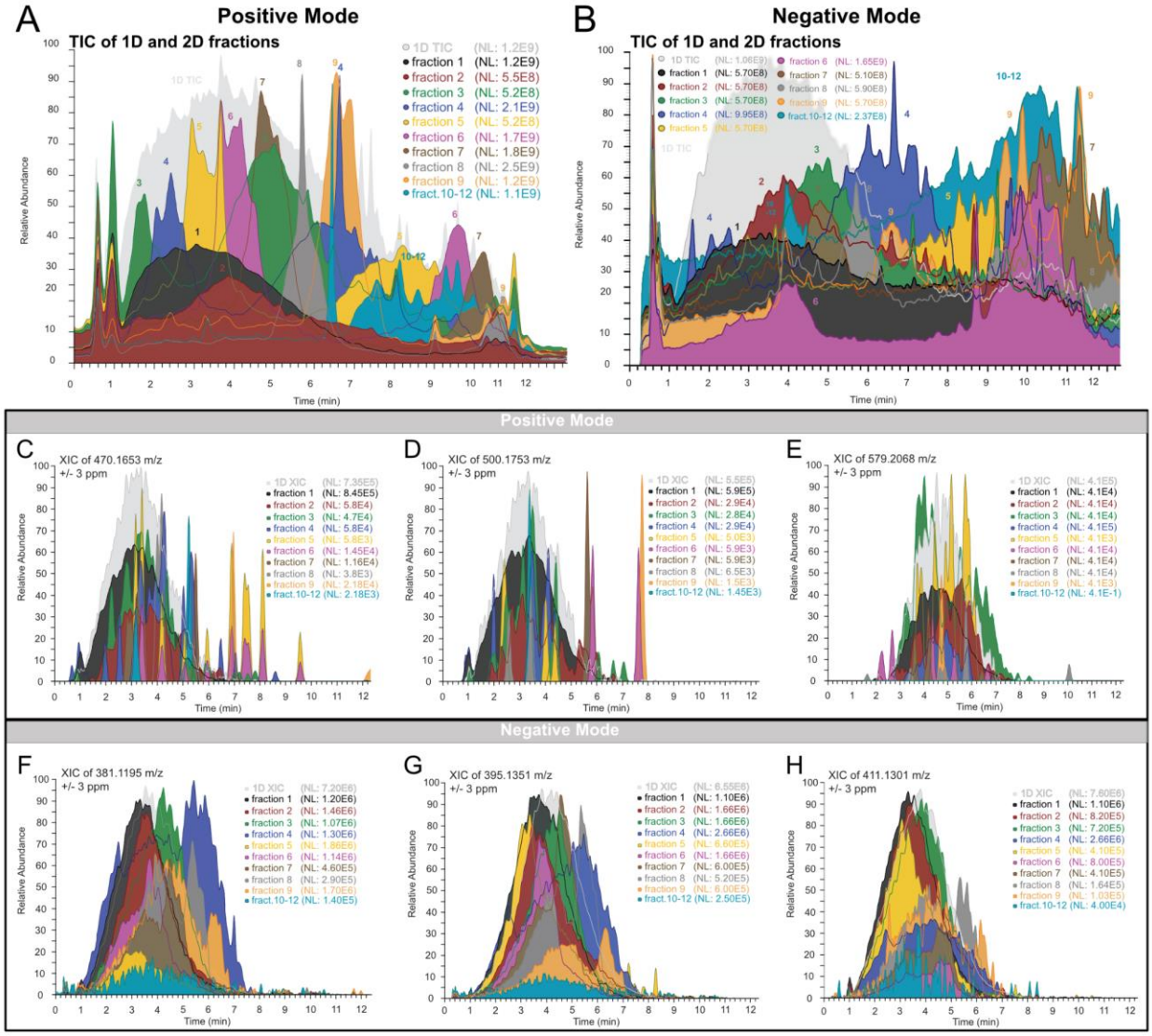

Figure 2: Comparison of 1D and 2D Total and Extracted Ion Chromatograms showing improved separation of DOM. (A) Overlay of the Total Ion Chromatograms (TICs) of the single fractions of the 2D analysis in positive mode (light gray, in background). (B) Overlay of the TICs of the single fractions of the 2D analysis in negative mode (light gray, in background). (C-E) Overlay of the Extracted Ion Chromatograms (XIC) for (C) m/z 470.1653, (D) m/z 500.1753, and (E) m/z 579.2068 of the single fractions of the 2D analysis over the XIC of the 1D analysis in positive mode (gray, in background). (F-H) Overlay of the XIC for (F) m/z 381.1195 (G) m/z 395.1351 (H) m/z 411.1301 of the single fractions of the 2D analysis over the XIC of the 1D analysis in negative mode (light gray, in background)

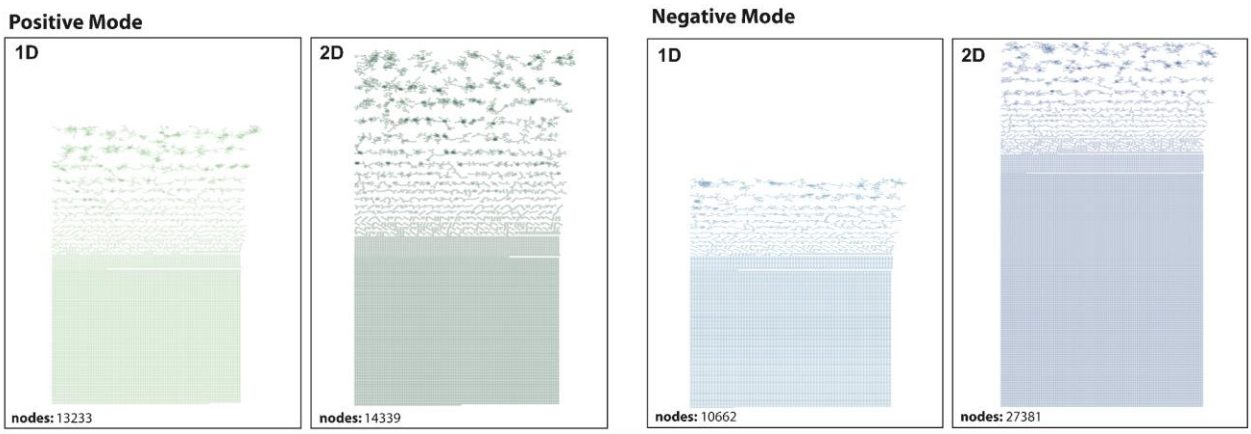

Figure 3: 1D and 2D LC-MS/MS based molecular networks showing larger numbers of annotated features and nodes for the 2D setup. Classic molecular networks of both positive mode (green, left side) and the negative mode (blue, right side) with the 1D and 2D method (lighter and darker colors, respectively). Data contained 20 replicate injections for 1D and duplicated injections for the 10 fractions for 2D (20 total injections).

As a first result, the classic molecular networks obtained from the 2D analysis show more nodes overall, but also more complex networks, having more connected nodes than the 1D networks, as shown in Figure 3. The difference is far more pronounced in the negative mode, where an additional strong increase in self-looped nodes (nodes not connected to other nodes) is visible. This result could be indicative of the improvement of carboxylic acid-rich refractory components that are better ionized and typically more dominant in negative ESI mode. Larger and more complex molecular networks indicate a better overall spectral quality of MS2 spectra, as chimeric spectra would be more random and thus less likely to be similar to other spectra (e.g., single nodes).

To assess the spectral quality, we used the precursor purity checker module available in MZmine 3, which provides the precursor purity of each MS/MS feature, giving a score between 0 and 1, with 1 being a “clean” spectrum, which means that no other precursor was present within the isolation window in MS1 survey scan closest to the MS/MS scan. The spectral purity scores for all spectra between the 1D and 2D experiments are visualized in Figure 4. The median spectral purity in the positive mode increased over two-fold, from 0.25 for the 1D analysis to 0.72 for the 2D analysis, whereas the improvement in the negative mode was not as pronounced, increasing only from 0.53 to 0.81. Overall, in the distribution, a clear shift is visible in both modes: In the 1D analysis, the distribution shows a high abundance of spectra having low purity, which shifts towards higher purity in the 2D analysis. This shows a clear improvement of the spectral purity by the improved chromatographic resolution obtained through the 2D LC approach. Interestingly, we observed a general higher spectral purity of MS/MS in negative mode, which are mainly attributed to a set of features that had high precursor purity (~ 1.0) in both the 1D and 2 D analysis. Yet, the general trend of improved precursor purity in the 2D run also holds true for the remaining features (< 0.99) in negative mode.

To investigate if the improved precursor purity relates to improved chromatographic separation and peak shape, we plotted the precursor purity against the full width at half maximum (FWHM) of the chromatographic peak which showed a higher density of cleaner spectra (precursor purity

> 0.8) in the 2D runs (Figure 4). Interestingly, in both the 1D and 2D runs a general relation between high purity spectra and sharper peaks (FWHM <0.1 min) could be observed. Yet, we noted a shift in this pattern in the 2D runs, where an increasing amount of high purity MS/MS spectra come from chromatographic features with wider peaks (FWHM > 0.1 min).

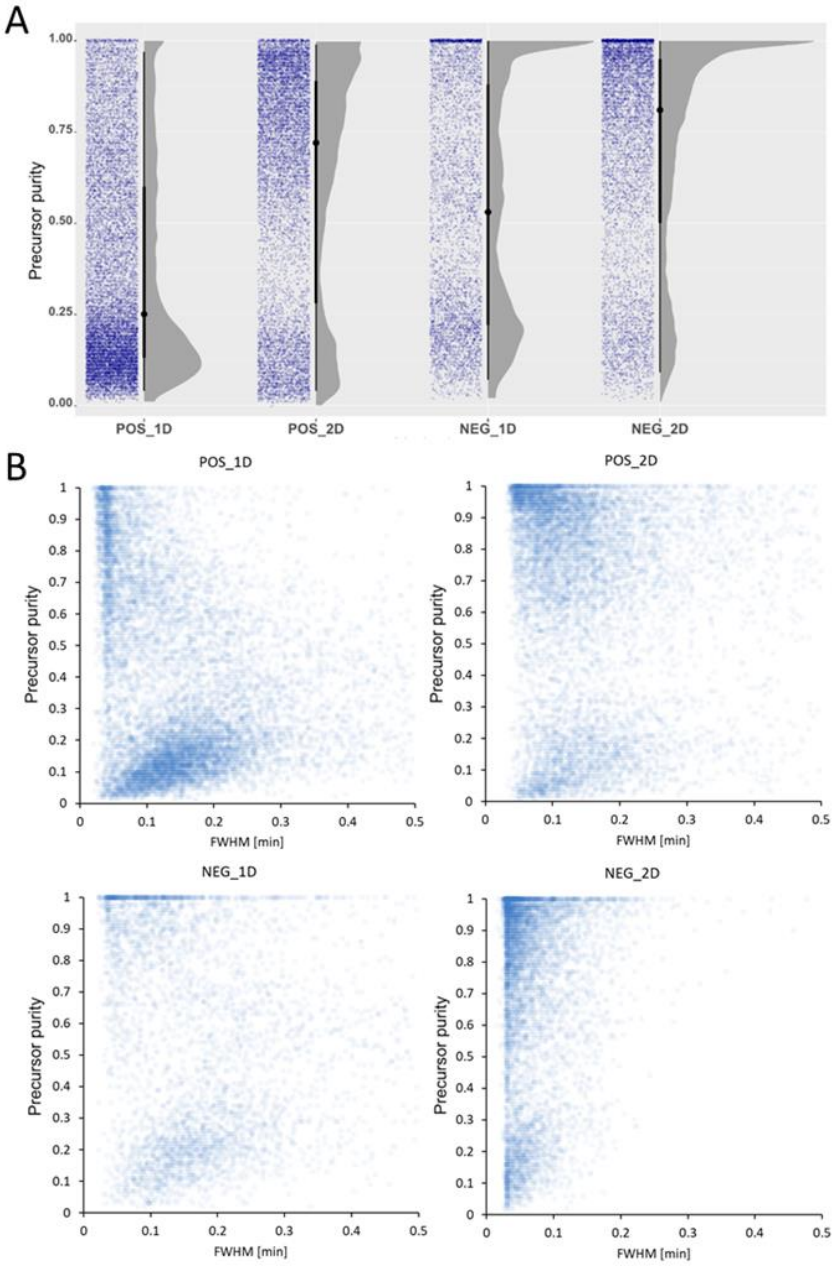

Figure 4: Distribution of MS/MS precursor purity. MS/MS precursor purity was evaluated using the precursor purity checker-module in MZmine 3.5.0, which ranks the purity between 1 and 0, with 1 representing a "clean" precursor selection for MS/MS. (A) shows raincloud plots of precursor purity. The single spectra are depicted as dots, the distribution is depicted as a half-violin plot, the black dot indicates the median, the thick black line the interquartile range (IQR) and the thin black line the 95% confidence interval. (B) Scatter plots show the relation between precursor purity and the full width at half maximum (FWHM) of the chromatographic peaks.

To further assess differences in spectral quality at the global data set level, we analyzed parameters concerning identification and networking efficiency (Figure 5 and S1). To also visualize the effect of the iterative 1D analysis, we analyzed the 20 1D runs in an increasing order (1, 2, 3, 4 runs etc. and the 2 D analysis via classical molecular networking (Figure 5 A+B). As expected, in the line plot we see that with an increasing amount of 1D runs, an increase in the number of total Library IDs from 48 to 291 in positive mode and from 5 to 71 in negative mode, can be observed. Strikingly, for the 2D approach (a total of 20 runs in the 2nd dimension with the same total run time as 20 1D runs), a total of 501 Library IDs were obtained in positive mode and 186 in negative mode, increasing the Library IDs by 72% or even 162%, respectively. Taken together, the overall increase in Library IDs is 90% (from 362 to 687) in the combined positive and negative mode data. The annotation rate of all features in negative mode is comparable between the 20 1D runs and the 2D approach, whereas the annotation rate in positive mode is at 2.2% with 20 1D runs and increases to 3.5% for the 2D approach.

Interestingly, the 2D runs did not yield a higher number of total spectra, and for positive mode, the overall spectra decreased by 37.4% (from 124010 to 77712) and by 4.8% (from 97236 to 92591) in negative mode. Conversely, the total number of nodes (clustered spectra) increased slightly from 13233 to 14339 in positive mode and significantly from 10662 to 27381 in negative mode. We assume the larger number of total spectra in the 1D runs could be explained with the general higher signal throughout the LC gradient, so the number of MS/MS spectra acquired in every DDA duty cycle will be maxed out. For the 2D runs on the other hand, as the sample is decomplexed, in some areas of the LC gradient, the topN (5 iterative MS2 events after MS1 survey scan) of the DDA duty cycle will not be maxed out and thus less total MS/MS spectra will be acquired. While a decrease in the number of total spectra could indicate less total coverage, we argue that it could also indicate an improvement regarding MS/MS sub-sampling and less redundant spectra. Investigating the average amount of clustered spectra per network Node (i.e. redundant MS/MS spectra) we see a clear decrease from an average of 9.4 spectra per node in the 1D run vs. an average of 5.4 spectra per node in the 2D runs in positive mode as well as in negative mode with 9.1 vs. 3.4 spectra per node.

Assessing the number of networks that include at least one identified compound showed that, although the consecutive runs led to an increase (from 57 to 105 in positive and from 13 to 23 in negative mode), the 2D analysis was able to identify at least one compound in 128 networks for positive mode, and 42 for negative mode respectively (Figure S3). This shows a broader distribution of identified compounds throughout the molecular networks, which, as connected nodes share high structural similarity and have a known mass difference, facilitates structure elucidation and compound identification of unknowns.

Comparing the unique library IDs (based on unique library entries with smiles codes) between the 1D and 2D LC-MS/MS based molecular networks (Figure S2), an average of ~ 16 % where only annotated in the 1D analysis, whereas ~ 31 % were identified in both analyses and ~ 52% only in the 2D analysis.

To further assess the chemical space of the compounds detected, we used FBMN and SIRIUS. Briefly, FBMN differs from MN in that it uses processed spectral information and therefore includes MS1 information like retention time or ion mobility, thereby facilitating the distinction between isomers giving the same MS2 spectra, and spectral annotation in general, as well as incorporating relative quantitative information, which enables robust downstream metabolomics statistical

analysis. While more nodes were obtained with FBMN using the 1D analysis in positive mode (20220 vs 12758 in 2D), more library IDs could be annotated with the 2D analysis (1003 vs 1176 in 2D). In the negative mode, the 2D analysis led to more nodes and more annotations (12419 vs 15545 nodes in 2D and 117 vs 251 annotations). This can also be seen in the annotation rate obtained from the FBMN, almost doubling between the 1D and the corresponding 2D analysis (4.96% to 9.22% in positive mode and 0.94% to 1.61% in negative mode (Table S5)).

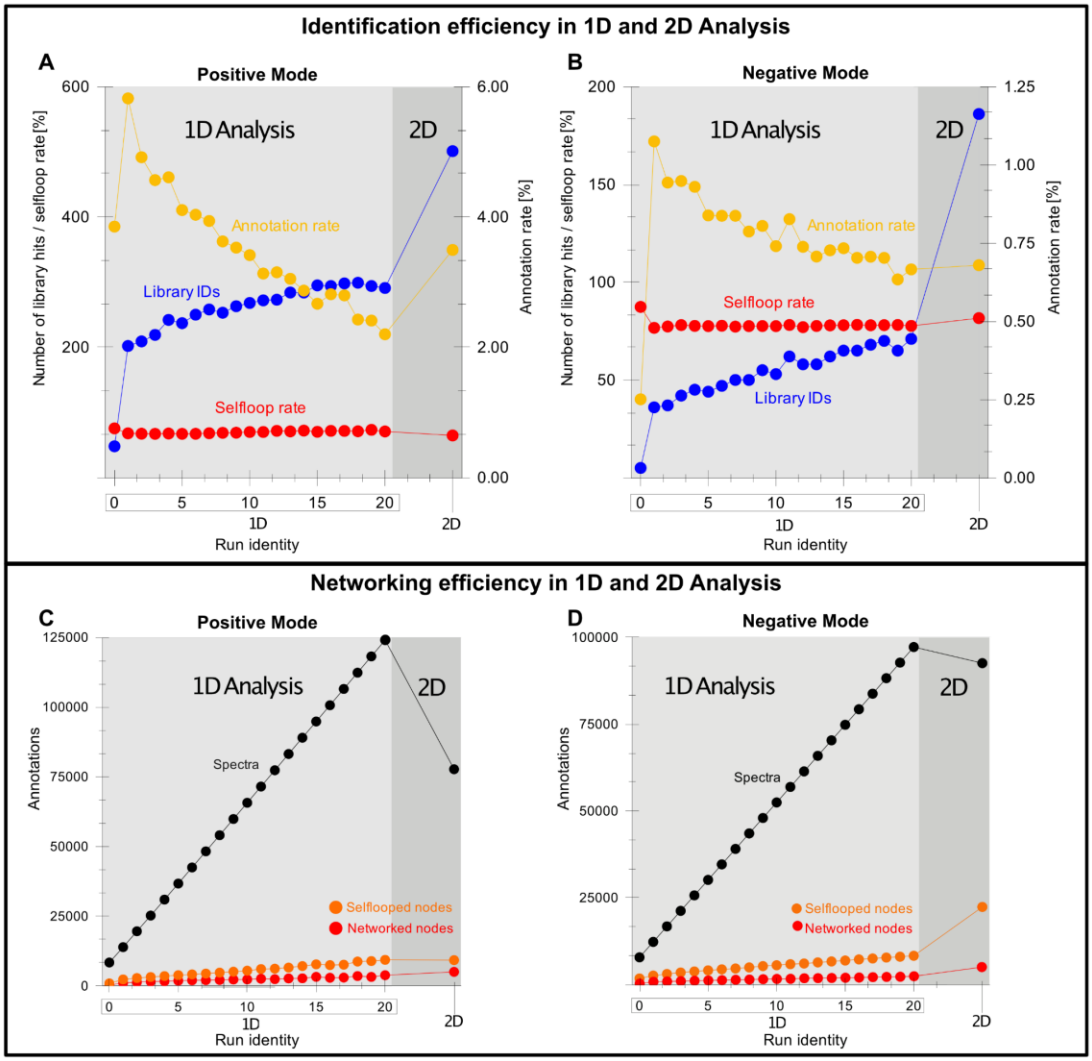

Figure 5: Increase in Identification and Networking efficiency between 1D and 2D analysis. (A,B) Comparison of annotation efficiency in 1D and 2D analysis in positive mode (A) and negative mode (B). (C,D) Comparison of Networking efficiency in 1D and 2D analysis in positive mode (C) and negative mode (D).

We exported the output of the FBMN to SIRIUS, an in-silico annotation tool, predicting the molecular formula by calculating possible fragmentation trees for MS2 spectra, and used the obtained molecular formulas. We obtained 17869 formulas for the 1D analysis, 11543 for the 2D analysis, and in negative mode 10527 formulas were predicted for the 1D analysis and 12055 for 2D. In negative mode, more formulas were predicted for the 2D analysis. Looking at H:C and O:C

ratios of molecular formulas in van Krevelen plot (Figure S3), the negative mode data shows a broader distribution of formulas compared to the 1D analysis, whereas the van Krevelen plots for the positive mode look overall similar in their distribution. Investigating the score distribution of the SIRIUS result, we observed no significant improvement of SIRIUS scores nor the CSI:FingerID confidence scores between 1d and 2D runs (Figure S4), indicating that MS/MS spectral purity might have less of an effect for molecular formula calculation with SIRIUS, which also relies on MS1 mass accuracy and isotope patterns.

Untangling the Deep Metabolome of Marine Dissolved Organic Matter. To obtain a status quo of annotated chemical space of coastal DOM, we combined the structural level 2 identification (spectral matches to MS/MS libraries[50]) obtained from the FBMN using 2D analysis in positive and negative mode and removed redundant annotations. After the removal of redundant nodes, often originating from library entries having isomeric or canonical smiles codes, 393 and 99 uniquely annotated compounds in positive mode and negative mode were level 2 IDs, respectively.

An overview of all level 2 annotated compounds in the 2D analyses is provided in supplemental tables S5-S8. Expanding the library search to putative analogs (spectra with high similarity but different precursor masses), we obtained 1746 analog hits (level 3 ID, meaning spectral similarity to known compounds of a chemical class[50]) for the compounds detected in positive mode, and 222 in negative mode. In order to display the annotated chemical space, we generated a structure-based network using the Tanimoto similarity metric, shown in Figure 6 (a high-resolution version of the network is provided in the SI). The network yielded some larger (> 50 nodes) and multiple smaller (< 20 Nodes) sub-networks, as well as individual unconnected nodes (152). As expected, the nodes in these clusters share a high structural similarity and/or are analog matches with similar frequently occurring delta mass patterns.

The largest cluster contained 557 compounds, mainly highly oxidized molecules such as di- and tricarboxylic acids and fatty acids. The second largest cluster, containing 351 nodes, also harbors highly oxidized compounds, having 326 compounds that contain carboxylic acids as well as aromatic and alicyclic rings, which are structural properties representatives for carboxyl-rich alicyclic molecules (CRAM), a common class of refractory DOM components (Figure 6D).

Overall, the annotated compounds include 918 (37,3%) nitrogen containing compounds. Hundreds of metabolites were detected within those classes, including disaccharides, fatty acids, peptides, and bile acids. In addition to the spiked standards, we could annotate a multitude of lipids and amino acids as well as terpenes such as loliolide, and the signaling compound dihydrojasmonic acid. In addition to the vast amount of natural products, dozens of xenobiotics including herbicides such as simetryn, and 2-hydroxyatrazine, the hydrolysis product of atrazine, drugs such as amitriptyline, drug metabolites such as benzoylecgonine and hydroxybupropion, as well as the industrially used triallyl cyanurate were annotated as level 2 IDs (mirror matches of the here mentioned compounds are provided in Figure S5 in the SI or can be inspected for all annotated compounds through the GNPS links provided in Table S2).

Notably, most of the mentioned xenobiotics were only found in the 2D analysis and would have remained unannotated using our conventional 1D LC-MS/MS approach. Together, the compounds annotated in this study (as level 2 IDs) as well as large amounts of derivatives (level

3 IDs) provide an unprecedented view into the chemical composition and vast diversity and human influences on the coastal DOM pool.

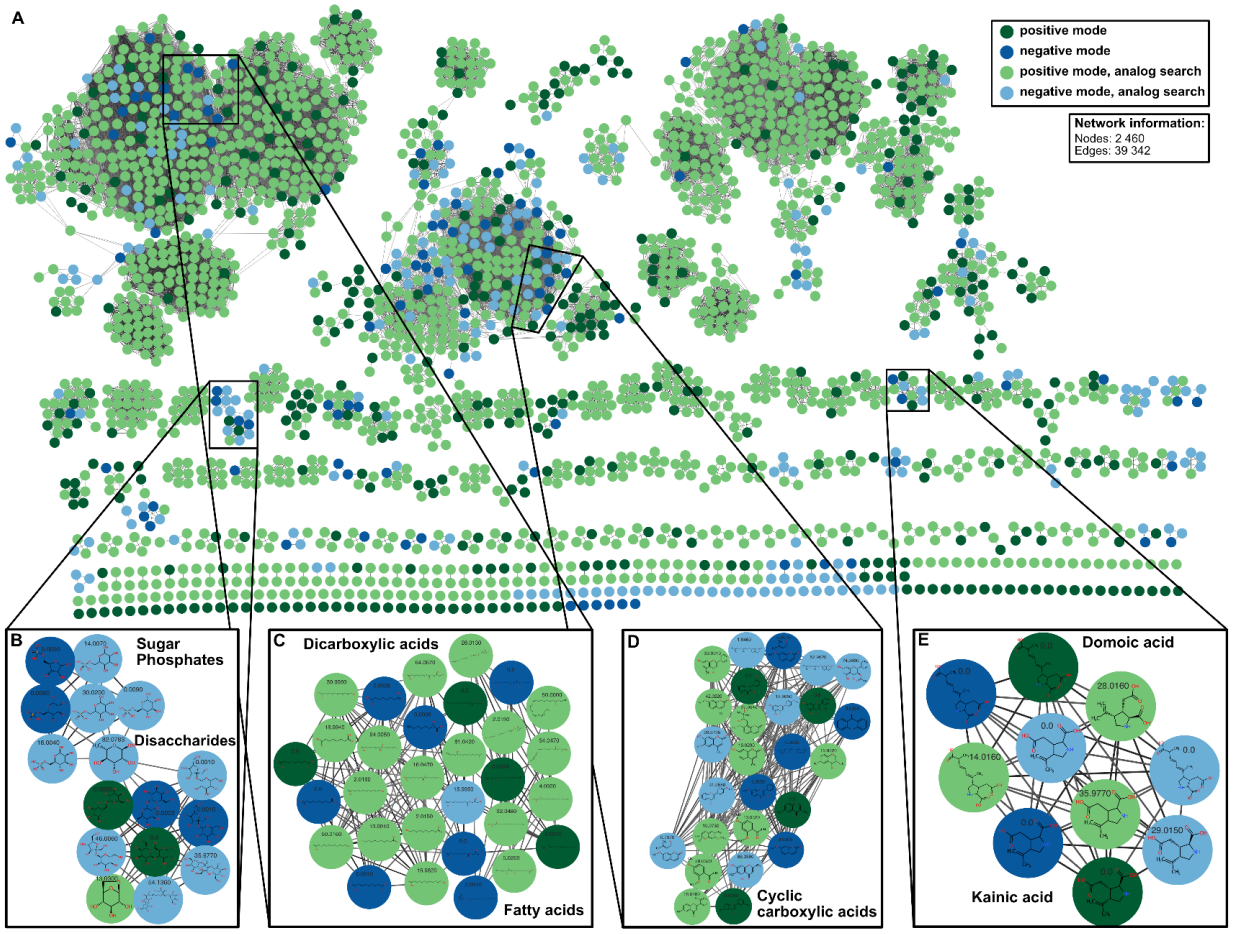

Figure 6: Structure-similarity network of annotated features (level 2 and level 3 IDs) from coastal DOM. A high resolution image of the network, including chemical structures, is available in the SI. (A) The figure shows the chemical similarity network between the annotated features using the Tanimoto similarity metric. Each node represents an annotated compound. The colors of the nodes indicate the mode of analysis (green, positive mode; blue, negative mode). After removal of redundant nodes, the network was expanded including the analog search provided by GNPS advanced library search, which is indicated by the respective lighter nodes. The connections between the nodes were calculated based on the chemical similarity between the compound's structural fingerprints using a threshold of 0.8. The network contains 393/99 unique library matches and 1746/222 analog library matches in the positive/negative ion mode. (B-E) Zoomed subnetworks, for C and D only a subset of the subnetwork is shown. (B) Zoom on a subnetwork containing sugar phosphates and disaccharides. (C) Zoom on a subset of nodes containing dicarboxylic acids and fatty acids. (D) Zoom on a subset of nodes containing carboxylic acids and aromatic and alicyclic rings. (E) Zoom on a subnetwork containing domoic and kainic acid derivatives.

Practical Considerations to Implement 2D-LC-MS/MS in Routine Analysis Workflows. A key advantage of the suggested protocol is its ease of integration with high-throughput 1D setups to both conserve fast data acquisition and improved annotation rates[27], and does not require changes in terms of hardware and gradient time to improve peak capacities[51]. This outweighs the

comparatively small additional effort of an added fractionation step of representative or pooled sample(s). Through the different pH in the mobile phases, different functional groups are influenced differently and change their interaction with the stationary phase. This leads to an orthogonal separation, increasing the chromatographic resolution, especially for closely related compounds. While offline 2D-LC-MS/MS approaches are inherently more time and work intensive compared to a routine LC-MS/MS approach, they clearly increase the chemical coverage and analytical depth. A possible strategy to implement 2D-LC-MS/MS analysis into large scale DOM studies, at a feasible total runtime, could be a hybrid 1D-2D approach. In such a workflow, the 2D analysis would be performed on representative or pooled sample sets to obtain high quality MS/MS spectra and increased depth, and the quantitative comparison of these features between all samples would be performed on MS1 level of the 1D runs. An important point here is that the chromatographic condition of the 1D run and the 2nd dimension of the 2D runs must be identical to align features by retention time and MS1.

CONCLUSION

To elucidate molecular signatures that drive global natural processes, such as global carbon cycling, including human impacts, it is essential to be able to detect and identify also very low abundant molecules to capture a more complete picture. With both, iterative injections via DDA-based LC-MS/MS as well as 2D-LC-MS/MS, we could increase the spectral coverage and annotation of marine DOM, through which we obtained unprecedented structural insights into the chemical space of an exemplary coastal water sample. We demonstrate that 2D-LC-MS/MS analysis, using reversed phase C18 UHPLC and orthogonal mobile phase modifier and pH, effectively improves MS/MS spectral purity and spectrum library matching results from DOM in comparison to standard non-targeted LC-MS/MS. Specifically, the obtained Library IDs increased by 90% and the median MS/MS spectral purity increased by 188% and 53% in positive and negative mode respectively. The improved data quality and depth of the 2D analysis can be clearly seen by the higher MS/MS purity, increased number of library annotations as well as larger network size in comparison to the 1D analysis.

In our case study with a coastal DOM sample, we could increase the obtained library IDs, as well as the networked nodes in positive and negative mode, thereby improving the molecular insights into DOM.

There is little doubt that with the constant increase of public DOM LC-MS/MS datasets, reference samples and the inclusion of new *in silico* annotation tools, molecular insight into DOM will further improve. We envision that the here presented 2D LC-MS/MS method will be broadly applicable in the field and will contribute to a deeper annotation of DOM in the future. Together with FAIR sharing of raw and processed data, we hope that these analytical advances will contribute to a better understanding of biogeochemical dynamics and human influences on DOM in the ocean as well as other ecosystems.

DATA AVAILABILITY

Raw data was converted to .mzML format with MSConvert using centroid mode. All raw and converted data can be found on the Mass spectrometry Interactive Virtual Environment (http://massive.ucsd.edu/) with the accession number MSV000092520. Classic molecular Network analysis was performed through the GNPS data analysis environment (gnps.ucsd.edu) with the parameters found in the SI (Table S1). Urls to the jobs submitted to GNPS with the obtained results can be found in the SI (Table S2).

AUTHOR CONTRIBUTIONS

DP conceptualized and designed the study. CS, IK and DP, performed initial proof of concept experiments. RRT, LIA and DP provided materials, collected the sample, and performed the extraction. SPL, PS and DP performed the MS data acquisition. SPL, TS, KSL, SF and DP performed data analysis and generated figures. RS contributed software. SPL, CS and DP wrote the Manuscript. All authors read, edited and approved the final manuscript.

NOTES

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

DP was supported by the Deutsche Forschungsgemeinschaft through the CMFI Cluster of Excellence (EXC 2124) and the Collaborative Research Center CellMap (TRR 261). PS was supported by the European Union Horizon 2020 Program through a Marie Skłodowska-Curie fellowship (101108450-MeStaLeM).

REFERENCES

(1) Moran, M. A.; Kujawinski, E. B.; Stubbins, A.; Fatland, R.; Aluwihare, L. I.; Buchan, A.; Crump, B. C.; Dorrestein, P. C.; Dyhrman, S. T.; Hess, N. J.; Howe, B.; Longnecker, K.; Medeiros, P. M.; Niggemann, J.; Obernosterer, I.; Repeta, D. J.; Waldbauer, J. R. Deciphering Ocean Carbon in a Changing World. *Proc Natl Acad Sci U S A* **2016**, *113* (12), 3143–3151. https://doi.org/10.1073/pnas.1514645113.

(2) Vorobev, A.; Sharma, S.; Yu, M.; Lee, J.; Washington, B. J.; Whitman, W. B.; Ballantyne, F.; Medeiros, P. M.; Moran, M. A. Identifying Labile DOM Components in a Coastal Ocean through Depleted Bacterial Transcripts and Chemical Signals. *Environ Microbiol* **2018**, *20* (8), 3012–3030. https://doi.org/10.1111/1462-2920.14344.

(3) Zhang, C.; Dang, H.; Azam, F.; Benner, R.; Legendre, L.; Passow, U.; Polimene, L.; Robinson, C.; Suttle, C. A.; Jiao, N. Evolving Paradigms in Biological Carbon Cycling in the Ocean. *National Science Review* **2018**, *5* (4), 481–499. https://doi.org/10.1093/nsr/nwy074.
(4) Kujawinski, E. B.; Longnecker, K.; Barott, K. L.; Weber, R. J. M.; Kido Soule, M. C. Microbial Community Structure Affects Marine Dissolved Organic Matter Composition. *Frontiers in Marine Science* **2016**, *3*.
(5) Wang, W.-L.; Moore, J. K.; Martiny, A. C.; Primeau, F. W. Convergent Estimates of Marine Nitrogen Fixation. *Nature* **2019**, *566* (7743), 205–211. https://doi.org/10.1038/s41586-019-0911-2.
(6) Decho, A. W.; Gutierrez, T. Microbial Extracellular Polymeric Substances (EPSs) in Ocean Systems. *Front Microbiol* **2017**, *8*, 922. https://doi.org/10.3389/fmicb.2017.00922.
(7) Beaupré, S. R.; Kieber, D. J.; Keene, W. C.; Long, M. S.; Maben, J. R.; Lu, X.; Zhu, Y.; Frossard, A. A.; Kinsey, J. D.; Duplessis, P. Oceanic Efflux of Ancient Marine Dissolved Organic Carbon in Primary Marine Aerosol. *Science Advances* **2019**, *5* (10), eaax6535.
(8) Malfatti, F.; Lee, C.; Tinta, T.; Pendergraft, M. A.; Celussi, M.; Zhou, Y.; Sultana, C. M.; Rotter, A.; Axson, J. L.; Collins, D. B.; Santander, M. V.; Anides Morales, A. L.; Aluwihare, L. I.; Riemer, N.; Grassian, V. H.; Azam, F.; Prather, K. A. Detection of Active Microbial Enzymes in Nascent Sea Spray Aerosol: Implications for Atmospheric Chemistry and Climate. *Environ. Sci. Technol. Lett.* **2019**, *6* (3), 171–177. https://doi.org/10.1021/acs.estlett.8b00699.
(9) Ossola, R.; Tolu, J.; Clerc, B.; Erickson, P. R.; Winkel, L. H. E.; McNeill, K. Photochemical Production of Sulfate and Methanesulfonic Acid from Dissolved Organic Sulfur. *Environ Sci Technol* **2019**, *53* (22), 13191–13200. https://doi.org/10.1021/acs.est.9b04721.
(10) Kujawinski, E. B. The Impact of Microbial Metabolism on Marine Dissolved Organic Matter. *Ann Rev Mar Sci* **2011**, *3*, 567–599. https://doi.org/10.1146/annurev-marine-120308-081003.
(11) 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.
(12) Leyva, D.; Tariq, M. U.; Jaffé, R.; Saeed, F.; Lima, F. F. Unsupervised Structural Classification of Dissolved Organic Matter Based on Fragmentation Pathways. *Environ Sci Technol* **2022**, *56* (2), 1458–1468. https://doi.org/10.1021/acs.est.1c04726.
(13) Simon, C.; Dührkop, K.; Petras, D.; Roth, V.-N.; Böcker, S.; Dorrestein, P. C.; Gleixner, G. Mass Difference Matching Unfolds Hidden Molecular Structures of Dissolved Organic Matter. *Environ Sci Technol* **2022**, *56* (15), 11027–11040. https://doi.org/10.1021/acs.est.2c01332.
(14) Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. High-Resolution Liquid Chromatography Tandem Mass Spectrometry Enables Large Scale Molecular Characterization of Dissolved Organic Matter. *Frontiers in Marine Science* **2017**, *4*.
(15) Patriarca, C.; Bergquist, J.; Sjöberg, P. J. R.; Tranvik, L.; Hawkes, J. A. Online HPLC-ESI-HRMS Method for the Analysis and Comparison of Different Dissolved Organic Matter Samples. *Environ Sci Technol* **2018**, *52* (4), 2091–2099. https://doi.org/10.1021/acs.est.7b04508.
(16) Han, L.; Kaesler, J.; Peng, C.; Reemtsma, T.; Lechtenfeld, O. J. Online Counter Gradient LC-FT-ICR-MS Enables Detection of Highly Polar Natural Organic Matter Fractions. *Anal Chem* **2021**, *93* (3), 1740–1748. https://doi.org/10.1021/acs.analchem.0c04426.
(17) Petras, D.; Minich, J. J.; Cancelada, L. B.; Torres, R. R.; 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.
(18) 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.
(19) Leyva, D.; Tose, L. V.; Porter, J.; Wolff, J.; Jaffé, R.; Fernandez-Lima, F. Understanding the Structural Complexity of Dissolved Organic Matter: Isomeric Diversity. *Faraday Discuss* **2019**, *218* (0), 431–440. https://doi.org/10.1039/c8fd00221e.
(20) Koester, I.; Quinlan, Z. A.; Nothias, L.-F.; White, M. E.; Rabines, A.; Petras, D.; Brunson, J. K.; Dührkop, K.; Ludwig, M.; Böcker, S.; Azam, F.; Allen, A. E.; Dorrestein, P. C.; Aluwihare, L. I. Illuminating the Dark Metabolome of Pseudo-Nitzschia–Microbiome Associations. *Environmental Microbiology* **2022**, *24* (11), 5408–5424. https://doi.org/10.1111/1462-2920.16242.
(21) Petras, D.; Caraballo-Rodríguez, A. M.; Jarmusch, A. K.; Molina-Santiago, C.; Gauglitz, J. M.; Gentry, E. C.; Belda-Ferre, P.; Romero, D.; Tsunoda, S. M.; Dorrestein, P. C.; Wang, M. Chemical Proportionality within Molecular Networks. *Anal. Chem.* **2021**, *93* (38), 12833–12839. https://doi.org/10.1021/acs.analchem.1c01520.
(22) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.
(23) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin H, C.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Feature-Based Molecular

Networking in the GNPS Analysis Environment. *Nat Methods* **2020**, *17* (9), 905–908. https://doi.org/10.1038/s41592-020-0933-6.
(24) Wang, M.; Jarmusch, A. K.; Vargas, F.; Aksenov, A. A.; Gauglitz, J. M.; Weldon, K.; Petras, D.; da Silva, R.; Quinn, R.; Melnik, A. V.; van der Hooft, J. J. J.; Caraballo-Rodríguez, A. M.; Nothias, L. F.; Aceves, C. M.; Panitchpakdi, M.; Brown, E.; Di Ottavio, F.; Sikora, N.; Elijah, E. O.; Labarta-Bajo, L.; Gentry, E. C.; Shalapour, S.; Kyle, K. E.; Puckett, S. P.; Watrous, J. D.; Carpenter, C. S.; Bouslimani, A.; Ernst, M.; Swafford, A. D.; Zúñiga, E. I.; Balunas, M. J.; Klassen, J. L.; Loomba, R.; Knight, R.; Bandeira, N.; Dorrestein, P. C. Mass Spectrometry Searches Using MASST. *Nat Biotechnol* **2020**, *38* (1), 23–26. https://doi.org/10.1038/s41587-019-0375-9.
(25) Quinlan, Z. A.; Koester, I.; Aron, A. T.; Petras, D.; Aluwihare, L. I.; Dorrestein, P. C.; Nelson, C. E.; Wegley Kelly, L. ConCISE: Consensus Annotation Propagation of Ion Features in Untargeted Tandem Mass Spectrometry Combining Molecular Networking and In Silico Metabolite Structure Prediction. *Metabolites* **2022**, *12* (12), 1275. https://doi.org/10.3390/metabo12121275.
(26) Stephens, B.; Stincone, P.; Petras, D.; English, C.; Opalk, K.; Giovannoni, S.; Carlson, C. Oxidation State of Bioavailable Dissolved Organic Matter Influences Bacterioplankton Respiration and Growth Efficiency. **2023**.
(27) da Silva, R. R.; Wang, M.; Nothias, L.-F.; van der Hooft, J. J. J.; Caraballo-Rodríguez, A. M.; Fox, E.; Balunas, M. J.; Klassen, J. L.; Lopes, N. P.; Dorrestein, P. C. Propagating Annotations of Molecular Networks Using in Silico Fragmentation. *PLoS Comput Biol* **2018**, *14* (4), e1006089. https://doi.org/10.1371/journal.pcbi.1006089.
(28) Sandron, S.; Rojas, A.; Wilson, R.; Davies, N. W.; Haddad, P. R.; Shellie, R. A.; Nesterenko, P. N.; Kelleher, B. P.; Paull, B. Chromatographic Methods for the Isolation, Separation and Characterisation of Dissolved Organic Matter. *Environ Sci Process Impacts* **2015**, *17* (9), 1531–1567. https://doi.org/10.1039/c5em00223k.
(29) 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.
(30) Zhang, F.; Harir, M.; Moritz, F.; Zhang, J.; Witting, M.; Wu, Y.; Schmitt-Kopplin, P.; Fekete, A.; Gaspar, A.; Hertkorn, N. Molecular and Structural Characterization of Dissolved Organic Matter during and Post Cyanobacterial Bloom in Taihu by Combination of NMR Spectroscopy and FTICR Mass Spectrometry. *Water Res* **2014**, *57*, 280–294. https://doi.org/10.1016/j.watres.2014.02.051.
(31) Stincone, P.; Shah, A. K. P.; Schmid, R.; Graves, L.; Lambidis, S. P.; Torres, R.; Xia, S.-N.; Minda, V.; Aron, A.; Wang, M.; Hughes, C. C.; Petras, D. Evaluation of Data Dependent MS/MS Acquisition Parameters for Non-Targeted Metabolomics and Molecular Networking of Environmental Samples - Focus on the Q Exactive Platform. ChemRxiv March 20, 2023. https://doi.org/10.26434/chemrxiv-2023-l8n67.
(32) 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.
(33) Spranger, T.; Pinxteren, D. van; Reemtsma, T.; Lechtenfeld, O. J.; Herrmann, H. 2D Liquid Chromatographic Fractionation with Ultra-High Resolution MS Analysis Resolves a Vast Molecular Diversity of Tropospheric Particle Organics. *Environ Sci Technol* **2019**, *53* (19), 11353–11363. https://doi.org/10.1021/acs.est.9b03839.
(34) Woods, G. C.; Simpson, M. J.; Simpson, A. J. Oxidized Sterols as a Significant Component of Dissolved Organic Matter: Evidence from 2D HPLC in Combination with 2D and 3D NMR Spectroscopy. *Water research* **2012**, *46* (10), 3398–3408.

(35) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Two-Dimensional Separation of Peptides Using RP-RP-HPLC System with Different PH in First and Second Separation Dimensions. *J Sep Sci* **2005**, *28* (14), 1694–1703. https://doi.org/10.1002/jssc.200500116.
(36) Wang, Z.; Ma, H.; Smith, K.; Wu, S. Two-Dimensional Separation Using High-PH and Low-PH Reversed Phase Liquid Chromatography for Top-down Proteomics. *Int J Mass Spectrom* **2018**, *427*, 43–51. https://doi.org/10.1016/j.ijms.2017.09.001.
(37) Baghdady, Y. Z.; Schug, K. A. Online Comprehensive High PH Reversed Phase × Low PH Reversed Phase Approach for Two-Dimensional Separations of Intact Proteins in Top-Down Proteomics. *Anal Chem* **2019**, *91* (17), 11085–11091. https://doi.org/10.1021/acs.analchem.9b01665.
(38) Vinh, J. Proteomics and Proteoforms: Bottom-up or Top-down, How to Use High-Resolution Mass Spectrometry to Reach the Grail; 2019; pp 529–567. https://doi.org/10.1016/B978-0-12-814013-0.00017-X.
(39) McIlvin, M. R.; Saito, M. A. Online Nanoflow Two-Dimension Comprehensive Active Modulation Reversed Phase-Reversed Phase Liquid Chromatography High-Resolution Mass Spectrometry for Metaproteomics of Environmental and Microbiome Samples. *J Proteome Res* **2021**, *20* (9), 4589–4597. https://doi.org/10.1021/acs.jproteome.1c00588.
(40) Cancelada, L.; Torres, R.; Garrafa-Luna, J.; Dorrestein, P.; Aluwihare, L.; Prather, K. A.; Petras, D. Assessment of PPL Solid-Phase Extraction and Non-Targeted Tandem Mass Spectrometry for the Analysis of Xenobiotics in Seawater. **2022**, *2022*, OS12B-0733.
(41) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. A Simple and Efficient Method for the Solid-phase Extraction of Dissolved Organic Matter (SPE-DOM) from Seawater. *Limnology and Oceanography: Methods* **2008**, *6* (6), 230–235.
(42) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.; Murray, K. J.; Hoffmann, N.; Lu, M. Integrative Analysis of Multimodal Mass Spectrometry Data in MZmine 3. *Nature biotechnology* **2023**, *41* (4), 447–449.
(43) 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. *Nature methods* **2019**, *16* (4), 299–302.
(44) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI: FingerID. *Proceedings of the National Academy of Sciences* **2015**, *112* (41), 12580–12585.
(45) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.
(46) Woods, G. C.; Simpson, M. J.; Koerner, P. J.; Napoli, A.; Simpson, A. J. HILIC-NMR: Toward the Identification of Individual Molecular Components in Dissolved Organic Matter. *Environ Sci Technol* **2011**, *45* (9), 3880–3886. https://doi.org/10.1021/es103425s.
(47) Baalousha, M.; Motelica-Heino, M.; Coustumer, P. L. Conformation and Size of Humic Substances: Effects of Major Cation Concentration and Type, PH, Salinity, and Residence Time. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2006**, *272* (1), 48–55. https://doi.org/10.1016/j.colsurfa.2005.07.010.
(48) A. Hawkes, J.; R. Sjöberg, P. J.; Bergquist, J.; J. Tranvik, L. Complexity of Dissolved Organic Matter in the Molecular Size Dimension: Insights from Coupled Size Exclusion Chromatography Electrospray Ionisation Mass Spectrometry. *Faraday Discussions* **2019**, *218* (0), 52–71. https://doi.org/10.1039/C8FD00222C.
(49) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; van der Hooft, J. J. J.; Ernst, M.; Kang, K. B.; Aceves, C. M.; Caraballo-Rodríguez, A. M.; Koester, I.; Weldon, K. C.; Bertrand, S.; Roullier, C.; Sun, K.; Tehan, R. M.; Boya P, C. A.; Christian, M. H.; Gutiérrez,

M.; Ulloa, A. M.; Tejeda Mora, J. A.; Mojica-Flores, R.; Lakey-Beitia, J.; Vásquez-Chaves, V.; Zhang, Y.; Calderón, A. I.; Tayler, N.; Keyzers, R. A.; Tugizimana, F.; Ndlovu, N.; Aksenov, A. A.; Jarmusch, A. K.; Schmid, R.; Truman, A. W.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Reproducible Molecular Networking of Untargeted Mass Spectrometry Data Using GNPS. *Nat Protoc* **2020**, *15* (6), 1954–1991. https://doi.org/10.1038/s41596-020-0317-5.

(50) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed Minimum Reporting Standards for Chemical Analysis. *Metabolomics* **2007**, *3* (3), 211–221. https://doi.org/10.1007/s11306-007-0082-2.

(51) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Analytical chemistry* **2005**, *77* (19), 6426–6434.