Supporting Information

Marine dissolved organic matter shares thousands of molecular formulae yet differs structurally across major water masses

Michael Seidel[1*], Sahithya Phani Babu Vemulapalli[1], Daniel Mathieu[2], Thorsten Dittmar[1,3]

[1] Research Group for Marine Geochemistry, Institute for Chemistry and Biology of the Marine Environment (ICBM), University of Oldenburg, 26129 Oldenburg, Germany
[2] Magnetic Resonance Spectroscopy, NMR Applications, Bruker BioSpin GmbH, 76287 Rheinstetten, Germany
[3] Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), 26129 Oldenburg, Germany

[*]Correspondence:
Michael Seidel
m.seidel@uni-oldenburg.de

Contents: 3 Tables, 6 Figures, References
This document contains supporting information for the manuscript titled above and contains 13 pages with additional method descriptions, 6 Figures, 3 Tables, and references. Figures include 1) superimposed ^{1}H NMR spectra, 2) a ^{1}H-^{1}H COSY spectrum with key structural feature groups, 3) ^{1}H-^{1}H COSY spectra of NEqPIW SPE-DOM recorded on 500 and 800 MHz NMR spectrometers with different cryo probes, 4) downfield ^{1}H-^{1}H COSY spectra of the five marine SPE-DOM samples, 5) Bray-Curtis dissimilarity matrixes of NMR and FT-ICR-MS structural and molecular formulae level data, and 6) superimposed simulated ^{1}H-^{1}H COSY spectra of four selected DOM compounds. Tables include number of common spectral features and Bray-Curtis dissimilarity values of ^{1}H-^{1}H COSY spectra and from FT-ICR-MS mass spectra.

Ultrahigh-resolution mass spectrometry

Molecular formulae were determined using a solariX XR FT-ICR-MS with a 15 Tesla magnet (Bruker Daltonics, USA) with an electrospray ionization source (ESI, Bruker Apollo II). SPE-DOM extracts were mixed with ultrapure water to yield a 1:1 water-methanol ratio (*v/v*) and a DOC concentration of 5 mg C L^{-1}. Samples were directly infused into the ESI unit with a flow rate of 2 µL min^{-1} with the capillary voltage set to 4.5 kV in negative mode. Ions were accumulated for 0.2 s in the quadrupole unit prior to transfer into the ICR cell, recording 1000 scans in broadband mode using 8 megaword data sets and a scan-range of 94 to 2000 Da (one hour measurement time per sample). Mass spectra were internally calibrated with a reference mass list covering the full detected mass range (mass accuracy <0.1 ppm). Molecular formulae above the method detection limit (MDL) [1] were assigned using ICBM-OCEAN [2]. MDL was 2.5, sample junction was in fast join mode (0.5 ppm sample tolerance) and a recalibration tolerance of 0.5 ppm. Minimum signal to MDL ratio as backbone for recalibration was 5 using mean recalibration mode. Molecular formulae were assigned with $^{12}C_{1-50}{}^{1}H_{1-200}{}^{16}O_{1-50}{}^{14}N_{0-4}{}^{32}S_{0-2}P_{0-1}$ with a tolerance of 0.5 ppm in the mass range from 100 to 1000 Da while excluding isotope ratio mismatches above signal to MDL ratios of isotope formulae >6. Isotope tolerance was set to 1000 ‰, the homologous series network approach was applied (CH_2, CO_2, H_2, H_2O, and O), molecular formulae containing nitrogen (N), sulfur (S), and phosphorus (P) simultaneously were excluded as well as all singlet molecular formulae. Molecular formulae in the samples were normalized to the sum of total FT-ICR-MS signal intensities and the modified aromaticity index (AI_{mod}) and double bond equivalents (DBE) [3, 4] as well as intensity-weighted molar contents of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), sulfur (S) and phosphorus (P) as well as intensity-weighted molar ratios of oxygen-to-carbon (O/C), hydrogen-to-carbon (H/C), AI_{mod} and DBE were calculated for each

sample as in Seidel, et al. [5]. Instrument variability of the FT-ICR-MS analyses was accounted for by repeated measurements of an in-house standard of North Pacific (NEqPIW) DOM.

Table S1: Number of common spectral features (gray cells) and Bray-Curtis dissimilarity (blue cells) of SPE-DOM samples derived from the two-dimensional symmetrized ^{1}H-^{1}H COSY NMR spectra by comparing the data from peak matching with a precision of 0.1 ppm in both the frequency coordinates (F1, F2).

	EMW (209)	**NADW (177)**	**AABW (164)**	**NEqPIW (221)**	**STSW (169)**
EMW (209) [1]		130	132	120	74
NADW (177)	0.36 [2]		122	109	70
AABW (164)	0.34	0.32		106	61
NEqPIW (221)	0.51	0.51	0.52		108
STSW (169)	0.67	0.68	0.68	0.50	

[1] Number of total spectral features in the data set.
[2] Value of 0: the two data sets share all features; value of 1: data sets do not share any features.

Table S2: Number of common spectral features (gray cells) and Bray-Curtis dissimilarity (blue cells) of SPE-DOM samples derived from the two-dimensional symmetrized ^{1}H-^{1}H COSY NMR spectra by comparing the data from peak matching with a precision of 0.01 ppm in both the frequency coordinates (F1, F2).

	EMW (990)	**NADW (670)**	**AABW (1102)**	**NEqPIW (994)**	**STSW (845)**
EMW (990) [1]		145	230	112	55
NADW (670)	0.81 [2]		162	70	48
AABW (1102)	0.79	0.82		125	80
NEqPIW (994)	0.90	0.93	0.90		154
STSW (845)	0.95	0.95	0.93	0.85	

[1] Number of total spectral features in the data set.
[2] Value of 0: the two data sets share all features; value of 1: data sets do not share any features.

Table S3: Number of common molecular formulae (gray cells) and Bray-Curtis dissimilarity values (blue cells) in SPE-DOM samples derived from the FT-ICR-MS mass spectra.

	EMW (4652)	**NADW (4265)**	**AABW (3345)**	**NEqPIW (4471)**	**STSW (4120)**
EMW (4652) [1]		3836	3177	3823	3564
NADW (4265)	0.09 [2]		2976	3556	3299
AABW (3345)	0.11	0.13		2845	2770
NEqPIW (4471)	0.20	0.20	0.25		3732
STSW (4120)	0.17	0.20	0.22	0.10	

[1] Number of total molecular formulae in the data set.
[2] Value of 0: the two data sets share all features; value of 1: data sets do not share any features.

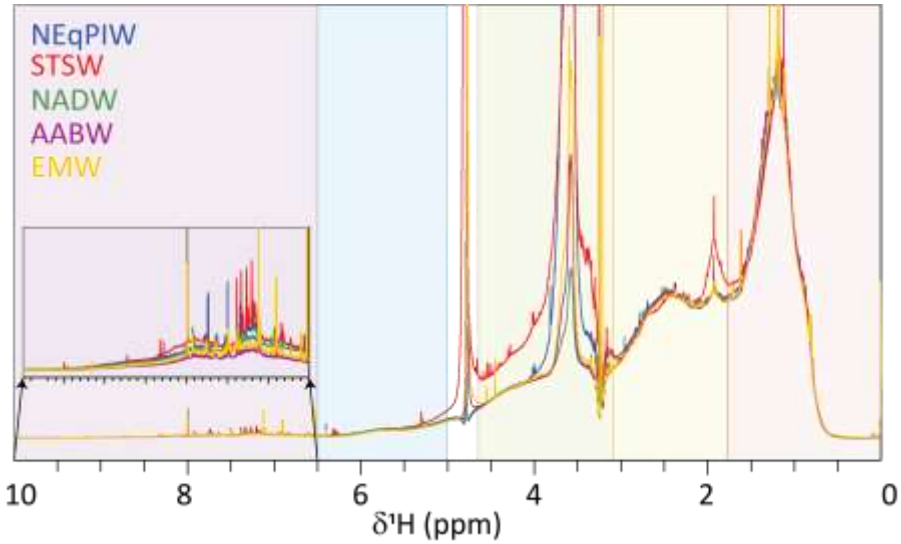

Figure S1: Superposition of ^{1}H NMR spectra of marine SPE-DOM (1 mg of each in 50 µL CD_3OD) samples recorded on Bruker Avance Neo 800 MHz spectrometer equipped with a 1.7 mm MicroCryoProbe. Broad humps with only a few resolved resonances are characteristic features of 1D ^{1}H NMR spectra of complex DOM mixtures. Yet, in the section of unsaturated protons which represent olefinic and aromatic structures, several signals can be differentiated. Measurement time was ca. two hours for each sample. Colored shapes depict main structural features: 0.0 - 1.9 ppm aliphatic; 1.9 - 3.1 ppm acetate and CRAM; 3.1 - 4.9 ppm carbohydrate and methoxy; 5.3 - 6.5 ppm olefinic; 6.5 - 10.0 ppm aromatic. Colors of the ^{1}H NMR spectra are in accordance with the colors of the sample locations.

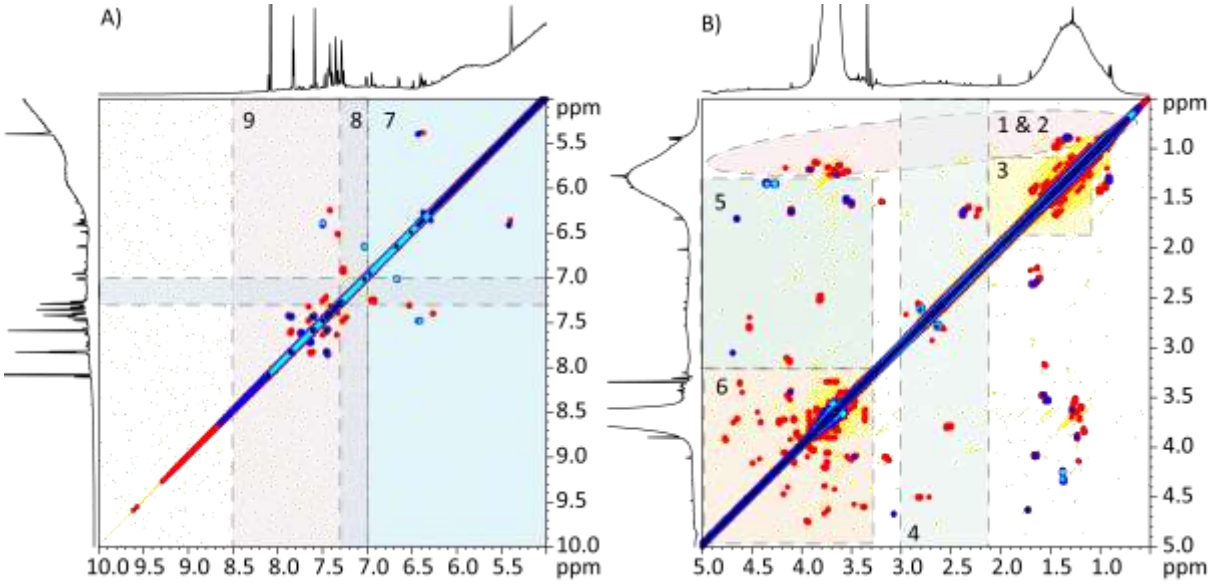

Figure S2: Two-dimensional symmetrized ^{1}H-^{1}H COSY spectrum of SPE-DOM (1 mg in 50 µL CD_3OD) from the deep Pacific Ocean (NEqPIW) recorded on a Bruker Avance Neo 800 MHz spectrometer equipped with a 1.7 mm MicroCryoProbe (total measurement time ~48 hours). The selected regions of the ^{1}H-^{1}H COSY spectrum of SPE-DOM highlight the aromatic/olefinic (**A**) and aliphatic correlations (**B**). Key structural features are: (1 & 2) chain-terminating methyl groups without/with oxygen, (3) intra-aliphatic cross peaks excluding methyl groups, (4) carbonyl derivatives including CRAM, (5) functionalized aliphatics connected with oxygenated carbon, (6) carbohydrates, esters, ethers, and alcohols, (7) olefinic and five-membered heterocycles, (8) aromatic, (9) polycyclic aromatic hydrocarbons (PAHs) and six-membered N-heterocycles. Structural categorization modified after Hertkorn, et al. [6].

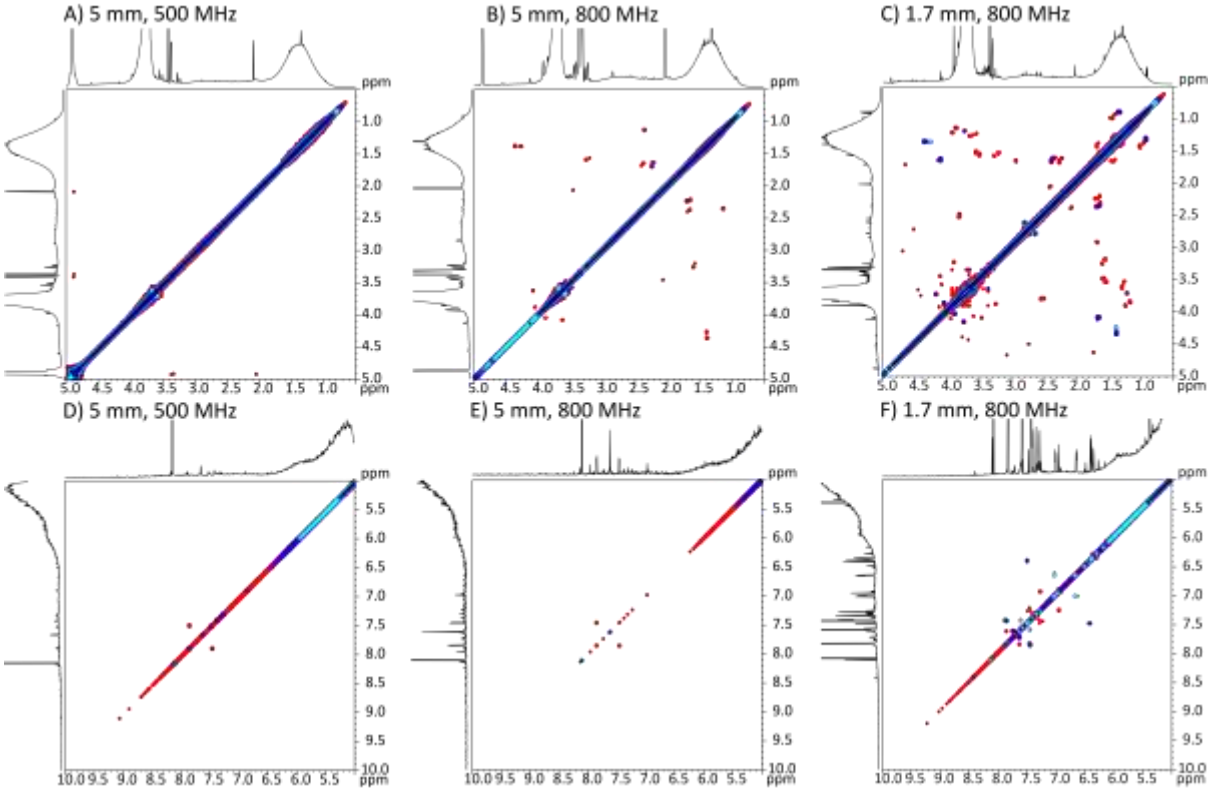

Figure S3: Comparison of two-dimensional symmetrized ^{1}H-^{1}H COSY spectra of DOM from the deep North Pacific Ocean (NEqPIW) recorded on a Bruker Avance Neo 500 MHz spectrometer equipped with a 5 mm TCI cryo probe (**A, D**), a Bruker Avance Neo 800 MHz spectrometer equipped with a 5 mm TCI cryo probe (**B, E**), and Bruker Avance Neo 800 MHz spectrometer equipped with a 1.7 mm TCI MicroCryoProbe (**C, F**). One mg of SPE-DOC (dissolved in 550 µL CD_3OD) was analyzed with helium-cooled 5 mm TCI cryoprobes (**A, D & B, E**) and a 1.7 mm MicroCryoProbe (dissolved in 50 µL CD_3OD) (**C, F**) with ~14 hours of measurement time for each spectrum). Because of the low sample concentrations and the unfavorably large size of the probes (5 mm in **A, D and B, E**), the 500 MHz and 800 MHz NMR spectra yielded very few interpretable cross signals. The sensitivity increased dramatically for this kind of small sample amounts which resulted in several hundred cross signals of abundant structural features in the spectrum at 800 MHz with the 1.7 mm MicroCryoProbe (**panels C, F**). Cross signals in the upfield (aliphatic features in **A, B, C**) and downfield (olefinic and aromatic features in **D, E, F**) spectra can only sufficiently be identified in the spectrum at 800 MHz with the 1.7 mm MicroCryoProbe with this small amount of sample material and in ~14 hours of measurement time.

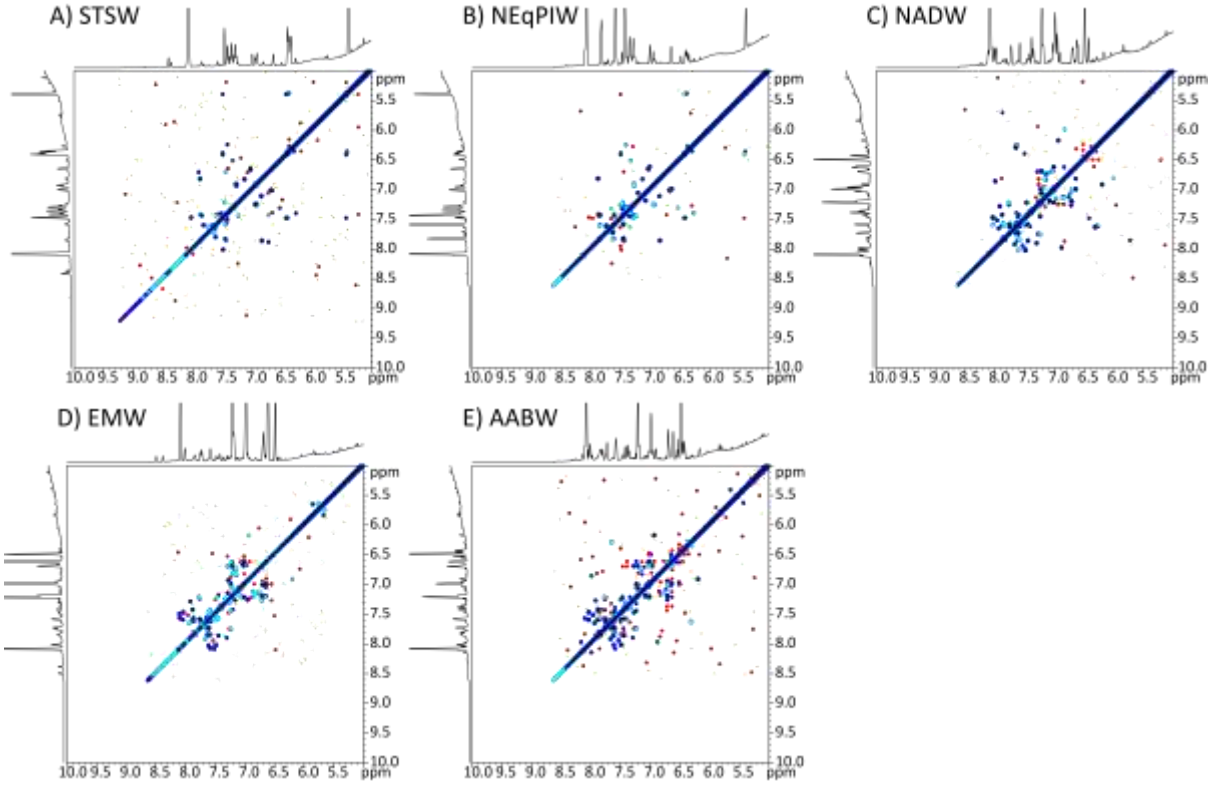

Figure S4: Comparison of the downfield region (^{1}H NMR δ_H = 5.0 – 10.0 ppm) with aromatic and olefinic cross peaks of two-dimensional symmetrized ^{1}H-^{1}H COSY NMR spectra of SPE-DOC (1 mg in 50 µL CD_3OD) recorded on a Bruker Avance Neo 800 MHz spectrometer equipped with a 1.7 mm MicroCryoProbe. Total experiment time was 48 hours per sample. The spectra display the proton-proton correlations related to the molecules containing aromatic and olefinic groups clearly highlighting the variation in the molecular composition of SPE-DOM from **A)** Subtropical Surface Water (STSW), **B)** North Equatorial Pacific Intermediate Water (NEqPIW), **C)** North Atlantic Deep Water (NADW), **D)** Eurafrican Mediterranean Water (EMW), and **E)** Antarctic Bottom Water (AABW).

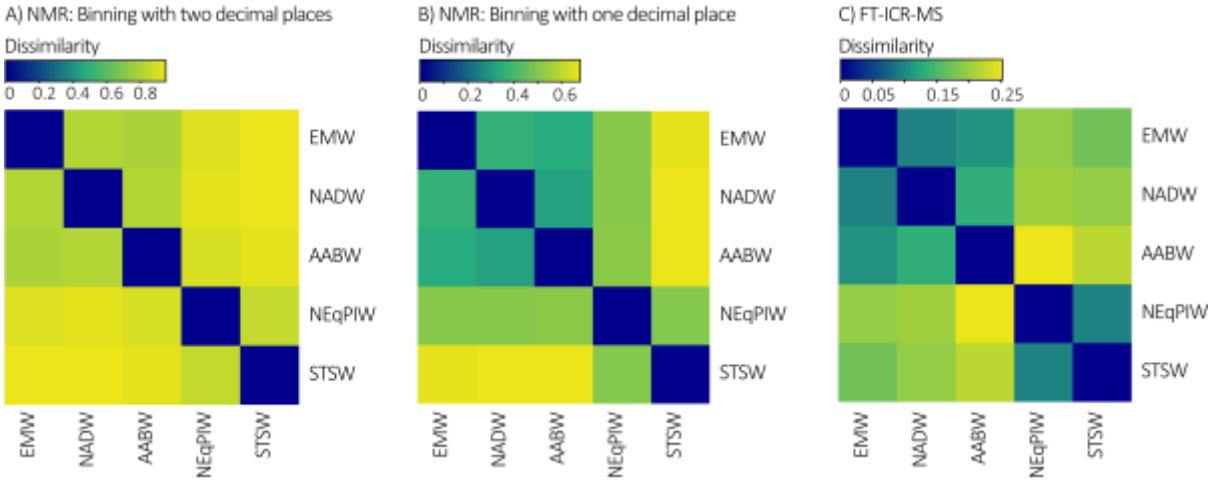

Figure S5: Structural and molecular formulae level dissimilarity of SPE-DOM from different oceanic water masses as deciphered by ^{1}H-^{1}H COSY NMR and FT-ICR-MS. Bray-Curtis dissimilarity is based on the relative intensities of spectral features (peaks) identified in the ^{1}H-^{1}H COSY NMR spectra when binning with two decimal places (precision of 0.01 ppm) (**A**) or binning with one decimal place (precision of 0.1 ppm) (**B**) of the F1 and F2 frequencies. (**C**) Bray–Curtis dissimilarity based on relative abundances of FT-ICR-MS derived molecular formulae.

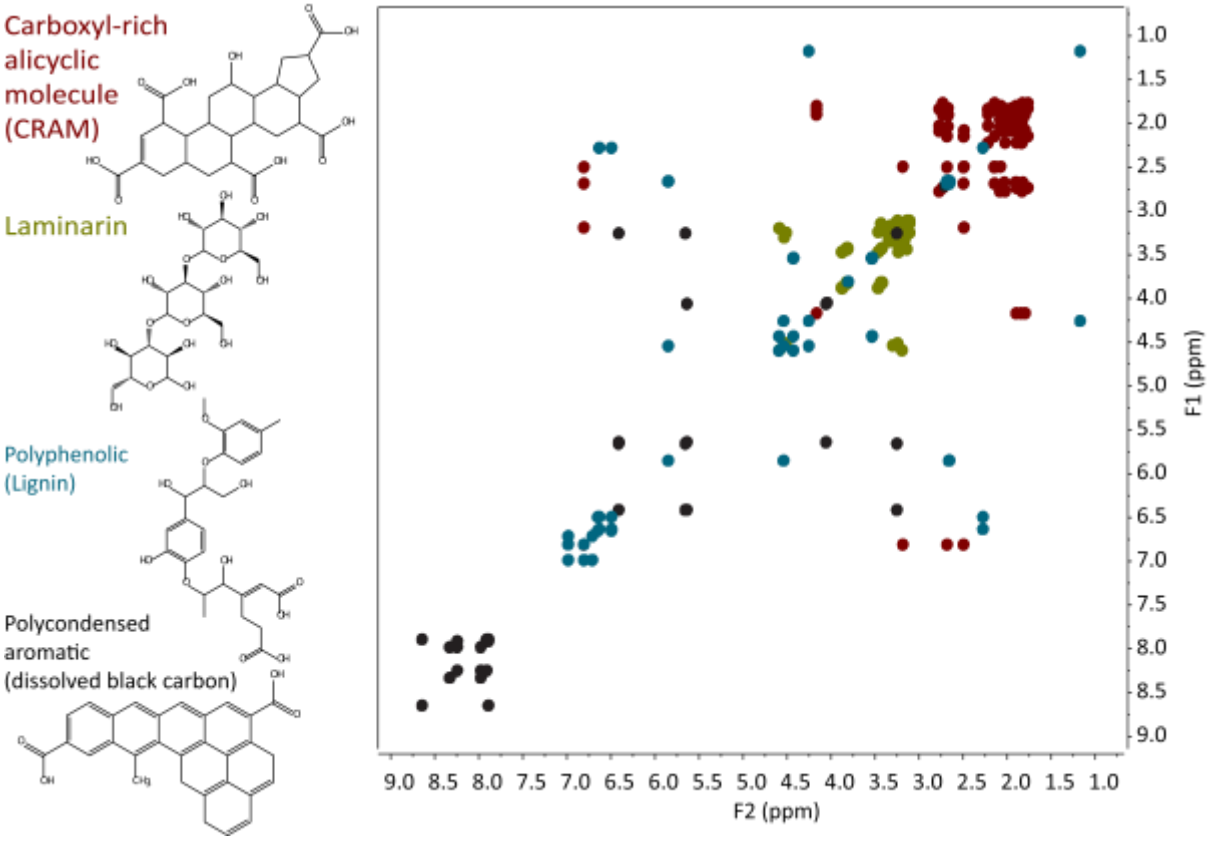

Figure S6: Simulated ^{1}H-^{1}H COSY NMR spectrum containing spectral features of selected organic molecules that can be present in natural DOM. Even these relatively simple natural molecules can produce a whole range of signals in a ^{1}H-^{1}H COSY spectrum based on the number of chemically distinct protons and their through-bond coupling interactions. This illustrates that operationally defined structural groups do not necessarily reflect an entire molecule but also parts of the same molecule. For example, the polyphenolic lignin structure will have signals (blue circles) in the functionalized aliphatics section, in the carboxyl-rich alicyclic molecule (CRAM, red circles) and aromatics sections. Note that each of these compounds will have only one molecular formula in FT-ICR-MS analysis. The COSY spectra of the depicted structures were simulated using the NMRdb online prediction tool [7] and the spectra were superimposed using the MestReNova software (Version 14.2.0). Chemical structures of the CRAM (red signals), polyphenolic (lignin, blue signals), and polycondensed aromatic compounds (black signals) are from Dittmar, et al. [8]. The laminarin structure (green signals) was retrieved the from ChemSpider chemical database (CSID:388438, http://www.chemspider.com/Chemical-Structure.388438.html (accessed 14:25, Jul 2, 2021).

References

1. Riedel, T.; Dittmar, T., A method detection limit for the analysis of natural organic matter via Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2014,** *86*, (16), 8376-8382.

2. Merder, J.; Freund, J. A.; Feudel, U.; Hansen, C. T.; Hawkes, J. A.; Jacob, B.; Klaproth, K.; Niggemann, J.; Noriega-Ortega, B. E.; Osterholz, H.; Rossel, P. E.; Seidel, M.; Singer, G.; Stubbins, A.; Waska, H.; Dittmar, T., ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Anal. Chem.* **2020,** *92*, (10), 6832-6838.

3. Koch, B. P.; Dittmar, T., From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **2006,** *20*, (5), 926-932.

4. Koch, B. P.; Dittmar, T., From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **2016,** *30*, (1), 250-250.

5. Seidel, M.; Beck, M.; Riedel, T.; Waska, H.; Suryaputra, I. G. N. A.; Schnetger, B.; Niggemann, J.; Simon, M.; Dittmar, T., Biogeochemistry of dissolved organic matter in an anoxic intertidal creek bank. *Geochim. Cosmochim. Acta* **2014,** *140*, 418-434.

6. 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.

7. Banfi, D.; Patiny, L., www. nmrdb. org: Resurrecting and processing NMR spectra on-line. *Chimia International Journal for Chemistry* **2008,** *62*, (4), 280-281.

8. Dittmar, T.; Stubbins, A., 12.6—Dissolved organic matter in aquatic systems. In *Treatise on Geochemistry, 2nd edn. Elsevier: Oxford*, Birrer B; Falkowski P; K, F., Eds. 2014; pp 125-156.