Available online at www.sciencedirect.com
SCIENCE DIRECT®
Geochimica et Cosmochimica Acta 70 (2006) 2990–3010
Geochimica
www.elsevier.com/locate/gca

Characterization of a major refractory component of marine dissolved organic matter

Norbert Hertkorn [a,*], Ronald Benner [b], Moritz Frommberger [a], Philippe Schmitt-Kopplin [a], Matthias Witt [c], Karl Kaiser [b], Antonius Kettrup [a], John I. Hedges ✠

[a] Institute of Ecological Chemistry, GSF Research Center for Environment and Health, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany
[b] Department of Biological Sciences and Marine Science Program, University of South Carolina, Columbia, SC 29208, USA
[c] Bruker Daltonics, Fahrenheitstrasse 4, D-28359 Bremen, Germany

Received 5 December 2005; accepted in revised form 14 March 2006

Abstract

Refractory carboxyl-rich alicyclic molecules (CRAM) are characterized in marine dissolved organic matter (DOM) using nuclear magnetic resonance spectroscopy and ultrahigh resolution mass spectrometry. CRAM are distributed throughout the water column and are the most abundant components of deep ocean DOM ever characterized. CRAM are comprised of a complex mixture of carboxylated and fused alicyclic structures with a carboxyl-C:aliphatic-C ratio of 1:2 to 1:7. CRAM are expected to constitute a strong ligand for metal binding, and multiple coordination across cations could promote aggregation and marine gel formation thereby affecting CRAM reactivity and the bioavailability of nutrients and trace metals. It appears CRAM are ultimately derived from biomolecules with structural similarities to sterols and hopanoids. The occurrence of CRAM in freshwater and terrestrial environments seems likely, considering the global distribution of biomolecules and the similarities of biogeochemical processes among environments.

1. Introduction

The oceans are a major global reservoir of reduced carbon (700 Pg), most of which occurs as dissolved organic matter (DOM) (Hedges, 1992). Much of the DOM in the ocean is refractory and has an average radiocarbon age of several millennia (Druffel et al., 1992). Little is known about the chemical composition of DOM and the reason for its refractory nature (Hedges et al., 2000; Benner, 2002). At present, in-depth knowledge about the origin, structure and function of DOM in the global carbon cycle remains elusive. Molecular analyses combined with degradative techniques have identified specific carbohydrates, amino acids and lipids in marine DOM, but these compounds only account for a small fraction of dissolved organic carbon (DOC) (Benner, 2002). The concentration and isolation of DOM from seawater by ultrafiltration facilitates its characterization by spectroscopic techniques. This approach, in combination with molecular analyses, has identified complex heteropolysaccharides (HPS) as major constituents of marine ultrafiltered DOM (UDOM) (Benner et al., 1992; Vernonclark et al., 1994; Aluwihare et al., 1997). A diverse suite of neutral, amino and acidic carbohydrates is found in HPS (McCarthy et al., 1996; Boon et al., 1998; Benner and Kaiser, 2003). The abundance of HPS is maximal in surface waters and declines sharply with depth, indicating HPS are a rapidly cycling component of DOM in the upper ocean. However, the most prominent contributors to the total ^{13}C and ^{1}H nuclear magnetic resonance (NMR) integral in UDOM from deep water remain undefined and have been recognized only as a humic-like background in ^{1}H NMR spectra (Aluwihare et al., 2002). Tentative assignments derived from cross polarization/magic angle spinning ^{13}C NMR spectra

* Corresponding author. Fax: +1149 89 3187 2713.
E-mail address: hertkorn@gsf.de (N. Hertkorn).
✠ Deceased: Inspiring us from heaven.

0016-7037/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.gca.2006.03.021

of marine dissolved humic substances indicated extensively branched, interlinked and possibly cyclic aliphatic carbon (Hedges et al., 1992).

Two previously characterized UDOM samples collected from surface (2 m) and deep (4000 m) waters of the mid Pacific Ocean were selected for this study (Benner et al., 1997). Additional samples from other locations were analyzed (data not shown), and all of the major structural features presented herein are generally representative of marine DOM and UDOM. The novel combination of ^{1}H NMR, ^{13}C NMR and Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) reveals the structural details of a major refractory constituent of marine UDOM. This newly characterized component is referred to as carboxyl-rich alicyclic molecules (CRAM). CRAM are slowly cycling components of marine UDOM, and they are the most abundant, identified components of DOM in the deep ocean.

2. Materials and methods

2.1. Sample collection

Surface (2 m) and deep (4000 m) water samples were collected from the Pacific Ocean (6°N, 140°W and 2°N, 140°W) using Niskin bottles. Samples were passed through a Nitex sieve (60 μm mesh size) and a 0.1 μm pore-size filter before tangential-flow ultrafiltration with a 1 kDa molecular-weight cutoff membrane to isolate ultrafiltered dissolved organic matter (UDOM; Benner et al., 1997). Ultrafiltered concentrates were diafiltered with deionized water to remove sea salts and freeze dried. The concentrations of dissolved organic carbon (DOC) in surface and deep waters were 72 and 44 μM, respectively. Ultrafiltration retained 26% of the DOC from surface water and 20% from deep water (Benner et al., 1997). The molar C:N ratios of the UDOM samples were 16.1 and 18.4, and the stable carbon isotopic compositions were −21.4‰ and −21.8‰ indicating a predominantly marine origin (Benner et al., 1997). Samples were also analyzed for hydrolysable amino acids (McCarthy et al., 1996), neutral sugars (McCarthy et al., 1996; Skoog and Benner, 1997) and amino sugars (Benner and Kaiser, 2003) by high performance liquid chromatography.

2.2. NMR spectroscopy

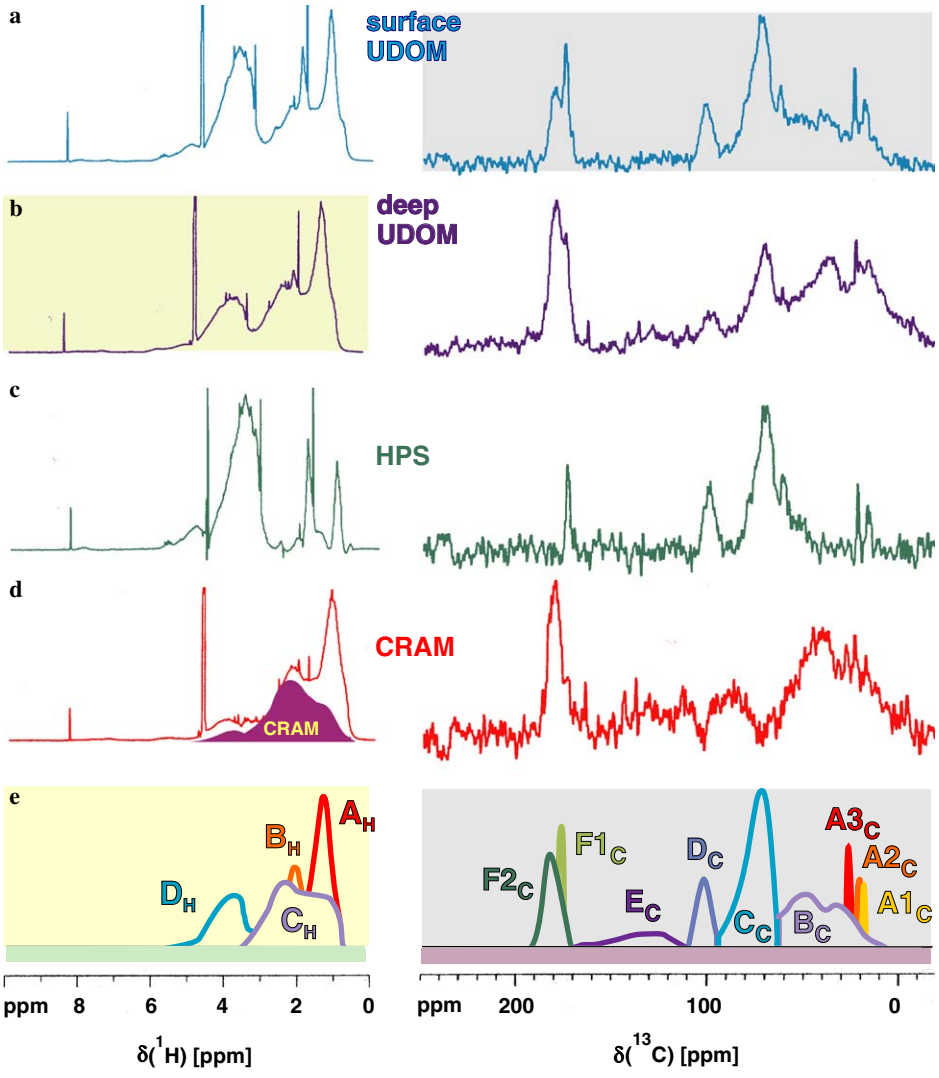
All experiments in this study were performed with a Bruker DMX 500 spectrometer at 303 K. The samples were dissolved in 0.1 N NaOD for ^{1}H NMR, ^{13}C NMR and 2D NMR spectroscopy. No fractionation was observed in the samples. The reference for ^{1}H NMR was $(H_3C)_3Si\text{-}CD_2\text{-}CD_2\text{-}COONa$ (−0.14 ppm), and for ^{13}C an external reference (CH_3OH in D_2O: 49.00 ppm) was used. All proton detected NMR spectra were acquired with a 5 mm *z*-gradient ^{1}H/^{13}C/^{15}N TXI cryogenic probe using 90° excitation pulses (90°(^{1}H) = 8.3 μs; 90° (^{13}C) = 19 μs). 1D ^{1}H NMR spectra were recorded using the first increment of the presat-NOESY sequence (solvent suppression with presaturation and spin-lock, 4.7 s acquisition time, 15.3 s relaxation delay, 320 scans, 1 ms mixing time, 1 Hz exponential line broadening). ^{13}C NMR spectra were acquired with a 5 mm broad-band probe (90°(^{13}C) = 9.2 μs), using inverse gated WALTZ-16 decoupling (8 s relaxation delay; 49,152 scans for surface UDOM, 68,542 scans for deep UDOM) with an acquisition time of 263 ms and an exponential line broadening of 35 Hz.

The one bond coupling constant ^{1}J(CH) used in 1D ^{13}C DEPT and 2D ^{1}H,^{13}C DEPT-HSQC spectra was set to 145 Hz [DEPT-HSQC NMR spectra were acquired with 320 scans and 312 increments (Willker et al., 1992)]. Carbon decoupled ^{1}H,^{13}C HSQC NMR spectra were acquired under the following conditions: ^{13}C-90-deg decoupling pulse, GARP (70 μs); F2 (^{1}H): acquisition time: 183 ms at spectral width of 5482 Hz, ^{1}J(CH) = 150 Hz, 1.82 s relaxation delay; F1 (^{13}C): SW = 22637 Hz (180 ppm); number of scans(F2)/F1-increments (^{13}C frequency) for surface UDOM (1920/800) and deep UDOM (238/191), respectively. HSQC and DEPT–HSQC spectra were calculated to a 2048×512 matrix with exponential line broadening of 30 Hz in F2 and a shifted sine bell (π/6) in F1 (Figs. 5 and 6). Gradient, but not sensitivity-enhanced, sequences (1 ms length, 450 μs recovery) were used for all proton detected spectra.

2.2.1. Analysis of NMR spectra

NMR integrals were measured manually from printed spectra. Difference NMR spectra were obtained using XWinNMR 3.0 software from spectra rather than from FIDs. In ^{13}C NMR, subtraction (d_UDOM − x_n · s_UDOM) was performed to attempt removal of the $F1_C$-component (i.e. the *N*-acetyl group, which is supposedly a substituent of HPS) to result in CRAM (CRAM: $x_1 = 0.76$). Removal of the $F2_C$-component (supposedly the carboxylic acid group of CRAM) resulted in HPS (with negative amplitude at $x_2 = 2.22$). Conversely, the manual subtraction (s_UDOM − $x_n \cdot$ d_UDOM) removed the $F2_C$ component to result in HPS ($x_3 = 0.44$) or, alternatively, the C_C component (carbohydrate methine carbon atoms and $HOCH_2$-groups) to result in CRAM ($x_4 = 1.17$; initial negative amplitude).

Bucket analysis (Brindle et al., 2002) was performed on the experimental ^{13}C NMR spectra of surface and deep UDOM, and of the computed difference NMR spectra, representing HPS and CRAM (Fig. 1); these were decomposed into 47 equidistant integral segments with 5 ppm bandwidth, ranging from 0 to 235 ppm (Fig. 1). Using the composition of a typical marine surface UDOM amino acid hydrolysate (with an average number of 2.83 side chain carbon atoms per CONH unit, which resulted, in conjunction with the atomic C/N ratio of 16.1, in 23.6% peptide carbon content in surface UDOM; Kaiser and Benner, 2006) and a typical marine deep UDOM amino



f

NMR resonance	D_H	C_H	B_H	A_H	$F1_C$	F_C	E_C	D_C	C_C	B_C	$A3_C$	$A2_C$	$A1_C$
δ(^{1}H/^{13}C) [ppm]	3.0-6.0	0.6-3.6	1.95	1.19	175	168-187	108-168	90-108	60-90	0-60	22.5	17.3	15.5
surface UDOM	56	30	6	8	4.2	17.7	2.0	8.5	36.1	35.7	2	0	0
deep UDOM	30	62	3	5	5.8	21.4	6.4	5.7	21.3	45.2	1.6	0	0
HPS	83	0	9	6	5.8	5.8	0	20.5	73.7	0	2.4	1.1	0.9
CRAM	0	100	0	0	0	19.9	11.8	5.2	9.6	53.5	0	0	0

Fig. 1. ^{1}H- and ^{13}C NMR data of marine UDOM and its major constituents: acquired ^{1}H (left) and ^{13}C (right) NMR spectra of surface (a) and deep (b) marine UDOM and, calculated from difference NMR spectra, of its two major constituents HPS (c) and CRAM (d). The ^{1}H NMR line spectrum (d) represents CRAM plus contributing genuine peptides; the purple area indicates the CRAM component as obtained from difference NMR spectra after subtraction of peptides (artificial mixture of polypeptides, which is representative of marine UDOM hydrolysates); (e) modeled spectra of deep UDOM (^{1}H NMR) and surface UDOM (^{13}C NMR) to indicate the NMR integral regions given in (f): ^{1}H NMR integrals (left) of four major resonances A_H–D_H A_H, methyl bound to purely aliphatic carbon; B_H, acetyl; C_H, functionalized CRAM and peptide aliphatics; D_H, heteroatom substituted protons (mostly carbohydrates and CH-α of peptides). $A_H + B_H + C_H + D_H$ are normalized to 100%, the total aromatic proton NMR integral for all materials is less than 1.5% (Fig. 2). ^{13}C NMR integrals of nine major resonances A_C–F_C $A1_C$, (poly)alanine-CH_3; $A2_C$, methylated carbohydrates; $A3_C$, *N*-acetyl carbohydrate; B_C, all aliphatics, aliphatic side chains of CRAM and peptides, including Cα; C_C, single oxygen substituted carbon; D_C, anomeric and other doubly oxygenated carbon; E_C, aromatics; F_C, carbonyl derivative (carboxylic, amide, ester) carbon, with $F1_C$, carboxylic acids and *N*-acetyl from HPS; $F2_C$, carboxylic acids and peptides from CRAM (with $F_C = F1_C + F2_C$). $B_C + C_C + D_C + E_C + F_C$ are normalized to 100%; grey underlined resonances $A1_C$–$A3_C$ and $F1_C$ (italics) are parts of aliphatic (B_C) and carbonyl derivative carbons (F_C), respectively.

acid hydrolysate (average number of 2.98 side chain carbon atoms per CONH unit, C/N atomic ratio: 18.4; resulting in 21.7% peptide carbon in deep UDOM; Kaiser and Benner, 2006), data reduced ^{13}C NMR spectra were computed via ACD CNMR predictor software, version 5.0 (Fig. 3).

From a combination of these data-reduced ^{13}C NMR spectra (Fig. 3), a rather conservative estimate for the minimum content of CRAM in UDOM was calculated, by use of a basic 3-component reverse mixing model (Nelson and Baldock, 2005) "UDOM = $n_1 \cdot$ HPS + $n_2 \cdot$ CRAM

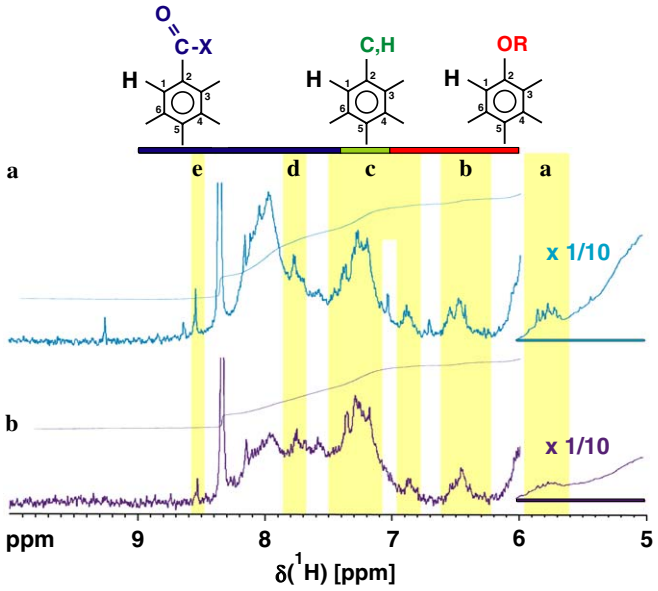


Fig. 2. Low field section of the ^{1}H NMR spectra of (a) surface (top) and (b) deep (bottom) UDOM expansion of Fig. 1a and b, indicating resonances of non-exchangeable protons (solvent: 0.1 N NaOD) occurring in both UDOM spectra (yellow boxes a–e), suggesting conformity of five-membered heterocyclic ring (a), phenolic (b), C-substituted (c), and carbonyl-substituted (d and e) aromatic protons in both environments. In contrast, low field (δ (^{1}H) >7.8 ppm) aromatic carbonyl derivatives are much more abundant in surface than in deep UDOM. Small amounts of formiate ($\delta(^1H) = 8.31$ ppm), methanol ($\delta(^1H) = 3.24$ ppm; Fig. 1) and acetate ($\delta(^1H) = 1.82$ ppm; Fig. 1) are formed by hydrolysis of UDOM in NaOD.

$+ n_3 \cdot$ peptides," based on the assumption that all nitrogen in UDOM is proteinaceous; linear regression analysis with equal weighting factors for all buckets provided the respective coefficients n_i for the computed 3-component surface and deep UDOM shown in Fig. 3. That procedure does not recognize inorganic nitrogen, nucleic acids and other heterocyclic nitrogen, amino- and *N*-acetyl sugars, which certainly are constituents of marine UDOM, and therefore provides an upper limit for the content of peptides.

2.2.2. Computer generation and computing of NMR chemical shift data of an aliphatic model polycarboxylic acid $C_{644}H_{1142}O_{146}$

The computer generation of an aliphatic model polycarboxylic acid $C_{644}H_{1142}O_{146}$ was initiated with a long straight chain aliphatic carbon backbone (polymethylene), carboxylic groups were then introduced, and the spacing in between those was varied from one (like in polyacrylic acid) to four carbon atoms (like in polyamylic acid) in increasing order. A systematic pattern of increasing methylation throughout the entire chain was then applied to account for a variable degree of branching, resulting in 83 aliphatic methylene carbon (with no methyl attached), 83 methine carbon (with a single methyl attached), and 83 quaternary carbon (doubly methylated) atoms; in total, 249 methyl groups have been placed between the terminal carboxylic groups. The chemical environment of the 73 carboxylic groups necessary to accomplish all feasible substitution patterns (cf. Fig. 7) was defined as H—C—COOH (resulting in another 73 methine carbon atoms), which act as spacers in between the methylated aliphatic chain units. The entire assembly results in a molecular formula $C_{644}H_{1142}O_{146}$ (excluding two terminal methyl groups), corresponding to a H/C atomic ratio of 1.773 and a O/C atomic ratio of 0.227 (cf. Figs. 7 and 11).

The individual proton and carbon chemical shifts of $C_{644}H_{1142}O_{146}$ were then computed [by means of Advanced Chemistry Development HNMR and CNMR predictor software (version 5.0)] from three sections ($C_{154}H_{310}O_{31}$; $C_{164}H_{330}O_{25}$; $C_{216}H_{434}O_{24}$, cf. Fig. 7) with sufficient overlap to avoid effects of chain termination on the proton and/or carbon NMR chemical shifts for any given atom position.

2.3. Capillary electrophoresis

Capillary zone electrophoresis (CZE) was performed using a PA/CE 5000 Beckman–Coulter instrument with a 57 cm fused silica capillary (50 cm length to the detector) and UV-detection (214 nm) as well as electrospray ionization negative mode MS detection (LCQ-Duo Finnigan ion trap mass spectrometer under conditions analogous to those in Schmitt-Kopplin and Kettrup, 2003). CZE separations were carried out at 30 °C in 25 mM ammonium carbonate buffers at pH 9.3 and 11.4 under 25 kV. For capillary gel electrophoresis (CGE), 0.3% methylcellulose was added to the pH 9.3 carbonate buffer.

2.4. FTICR mass spectrometry

High-resolution mass spectra for molecular formula assignment were performed with a Bruker APEX-Q III Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) equipped with a 7 Tesla superconducting magnet using the electrospray positive ion mode of the ESI/MALDI combine source. The sample solution (1 mg UDOM in 1 ml methanol) was diluted 1:1 with aqueous 0.2% formic acid and introduced into the electrospray source by infusion with a flow of 120 μL h^{-1}. The spectra were externally calibrated with the collision induced dissociation spectrum of the peptide luteinizing hormone-releasing hormone (Sigma, [35263-73-1]). The spectra were acquired with a time domain of 1 megaword with a mass range of 200–2000 m/z. The spectra were zero filled to a processing size of 2 megawords. Before Fourier transformation of the time-domain transient, a sine apodization was performed. The ions were mass selected between ion source and FTICR-MS analyzer with a quadrupole using a setting for ion selection of about 100 mass units. The ion accumulation time in the ion source was set to 3 s for each scan. Sixty scans were added for each mass spectrum.

Standard CH_2-based Kendrick mass defect analysis has been performed on 613 CRAM molecular compositions $C_nH_mO_q$ derived from positive ions $C_nH_{m+1}O_q^+$ as obtained

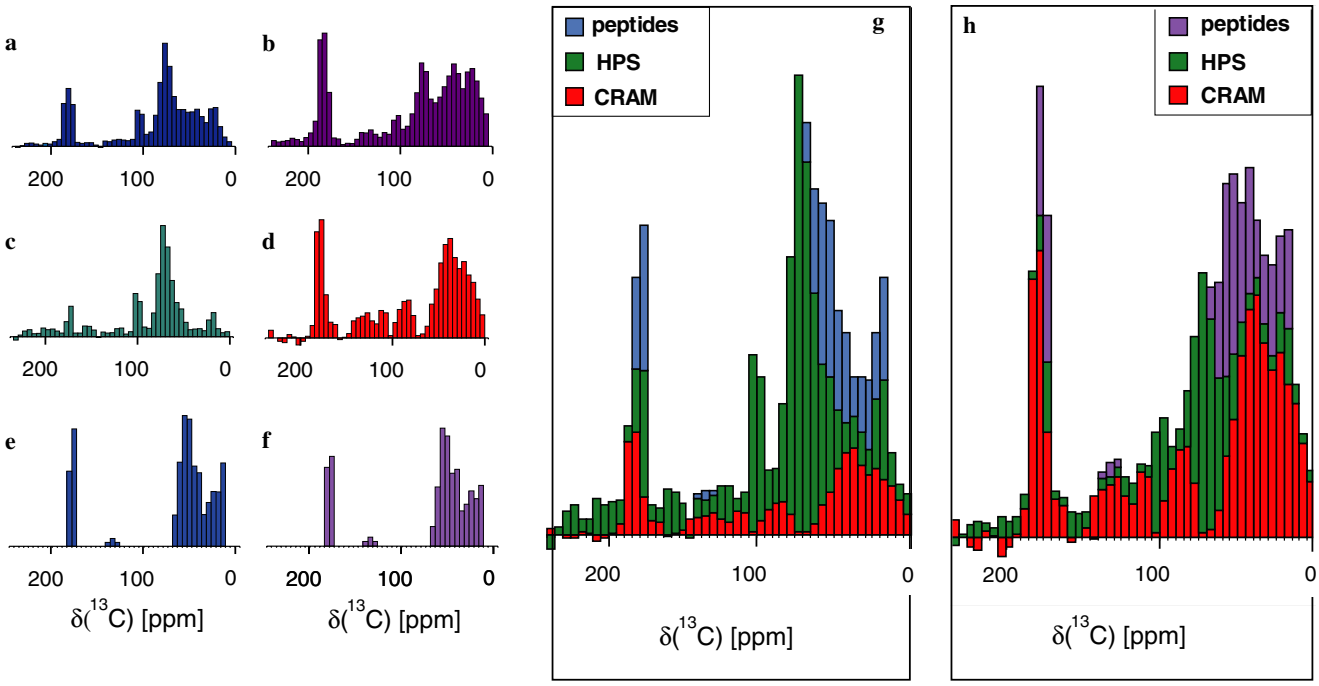


Fig. 3. Data reduced ^{13}C NMR spectra (47 buckets of integrated ^{13}C NMR intensity with 5 ppm width each, ranging from 0 to 230 ppm) of (a) surface and (b) deep UDOM, as computed from the experimental NMR spectra, and of the main UDOM constituents (c) HPS and (d) CRAM, as computed from the difference NMR spectra (cf. Fig. 1). The histogram spectra of the proteins in surface UDOM (e) and deep UDOM (f) were computed from chemical shifts of a protein with the composition of a surface and deep marine NOM hydrolysate via ACD software (cf. experimental). From these data-reduced ^{13}C NMR spectra, a very conservative estimate for the minimum content of CRAM in UDOM was computed, by use of a basic 3-component reverse mixing model "UDOM = $n_1 \cdot$ HPS + $n_2 \cdot$ CRAM + $n_3 \cdot$ peptides," based on the assumption that all nitrogen in UDOM is proteinaceous, resulting in a (g) data-reduced ^{13}C NMR spectrum of a 3-component surface UDOM ($n_1 = 0.601$, $n_2 = 0.229$, $n_3 = 0.236$) and a (h) data-reduced ^{13}C NMR spectrum of a 3-component deep UDOM ($n_1 = 0.271$, $n_2 = 0.509$, $n_3 = 0.217$). A better conformity of experimental and computed, data reduced, ^{13}C NMR spectra of surface and deep UDOM was obtained at smaller values of n_3, suggesting the occurrence of some nitrogen in UDOM in non-protein/peptide chemical environments.

in 7 Tesla FTICR mass spectra. The CO_2-based Kendrick mass has been computed by multiplying the IUPAC mass of CRAM by the factor (44/43.989828). In order to keep the nominal Kendrick masses for given molecular compositions identical for both CH_2- and CO_2-based Kendrick mass analyses (Fig. 10), we have subtracted one mass unit from the computed nominal CO_2-based Kendrick mass.

3. Results and discussion

3.1. Analysis of one-dimensional ^1H and ^{13}C NMR spectra

Four major functional forms of non-exchangeable hydrogen were recognizable in ^1H NMR spectra of surface and deep UDOM acquired in NaOD (Fig. 1a and b). Strong resonances were observed in the 0.6–1.6 ppm region, which is characteristic of hydrogens bound to saturated carbon with heteroatoms three or more bonds away. Resonances in the 1.8–2.8 ppm region, which are characteristic of hydrogens bound to saturated carbon with heteroatoms two bonds away, and those in the 3.0–5.5 ppm region, which are characteristic of H—C bound to oxygen, were also prevalent. In contrast, only weak resonances (<1.5% of total ^1H NMR integral) were found in the olefinic and aromatic regions (6–9 ppm; Fig. 2).

Four major functional forms of carbon were recognizable in the ^{13}C NMR spectra (Fig. 1a and b). In the 0–60 ppm region (A_C, B_C in Fig. 1) carbon is mainly bound to carbon and hydrogen atoms only; however, aliphatic carbon of amines, amides (C-α of peptides; cf. Fig. 5) and methoxyl groups also contribute. The 60–108 ppm region (C_C and D_C in Fig. 1) is characteristic of carbon bound to one or more heteroatoms, and the 108–168 ppm region (E_C in Fig. 1) is characteristic of olefinic and aromatic carbon. Carbonyl derivatives (esters, amides and carboxylic acids) resonate from 168 to 225 ppm (F_C in Fig. 1). Strong resonances were observed in all these regions except that of olefinic and aromatic C.

The NMR spectra of surface and deep UDOM closely resembled each other in the positions of their relative intensity minima and maxima, but differ vastly in their relative amplitudes (Fig. 1a and b). This indicated surface and deep UDOM were composed of variable amounts of preserved substructures, which can be visualized by difference NMR spectra (Fig. 1c and d). Two major UDOM constituents were observed in difference spectra. A complex mixture of heteropolysaccharides (HPS) was represented by surface minus deep UDOM (Fig. 1c), with major resonances in D_H and $C_C + D_C$ (Fig. 1e). A mixture, we refer to as carboxyl-rich alicyclic molecules (CRAM) was represented

by deep minus surface UDOM (Fig. 1d), with major resonances in C_H, B_C, and $F2_C$ (Fig. 1e).

The larger peak capacity of ^{13}C NMR spectroscopy (ratio of total bandwidth/individual linewidth) relative to that of ^{1}H NMR, in conjunction with the capability to directly observe quaternary carbon atoms, allows to detect finer resolved structural detail in ^{13}C NMR spectra of UDOM (Fig. 1e), albeit at a lower signal to noise ratio (at natural abundance, the NMR receptivity of ^{13}C nuclei is 1/5682 compared with that of proton nuclei). Alternatively, the very high sensitivity of high-field proton NMR spectra allows to determine rather tiny variations in UDOM composition in minute detail (Fig. 2), providing crucial (and quantitative) structural information not available by any other means at present.

Difference NMR spectra emphasize components that have varying concentrations. In the spectra (Fig. 1c and d), they describe HPS and CRAM, because peptides comprise a similar fraction of carbon in surface and deep UDOM (McCarthy et al., 1996; Hedges et al., 2001).

3.1.1. Computation of the CRAM content in UDOM from difference NMR spectra

The main UDOM constituents HPS and CRAM were visualized by difference NMR spectra (cf. Fig. 1), which were created manually from spectra rather than from FIDs (free induction decays). Poor resolution and S/N ratio (^{13}C NMR) of UDOM NMR spectra, even after extended acquisition time, made it advisable to avoid negative signal amplitudes in the difference NMR spectra (realized by manual adjustment of the relative multiplication factors; cf. experimental section).

Difference NMR spectra, derived from poorly resolved spectra of complex unknowns should be regarded as rather qualitative guidance tools. Hence, they emphasize UDOM constituents with variable concentration (like HPS and CRAM, respectively) and tend to attenuate or even ignore constituents with a more uniform distribution (like, e.g. proteins/peptides; McCarthy et al., 1996; Hedges et al., 2001). In addition, the S/N-ratio of difference spectra is intrinsically rather poor, and noisy sections are insufficiently defined. More importantly, global weighting factors employed across the entire spectral bandwidth do not recognize the specific distribution of NMR resonances of the individual molecular NOM constituents. Regarding the level of amino acid composition and protein content, surface and deep UDOM appear remarkably similar, if the amino acid composition of hydrolyzates (and its amino sugar content/composition), their C/N ratio, and NMR properties are considered.

Data reduction schemes decrease the dimensionality of spectral data to a more readily interpretable form, and they are also useful to illustrate the compositional relationships of a marine surface and deep UDOM with respect to its two major non-protein constituents, namely CRAM and HPS. Here, the ^{13}C NMR spectra of UDOM and its main constituents were first decomposed into 47 equidistant integral segments with 5 ppm bandwidth, ranging from 0 to 235 ppm; the resulting histograms are in essence very low resolution NMR spectra (Fig. 3).

From these data-reduced ^{13}C NMR spectra (Fig. 3), a conservative estimate for the minimum content of CRAM in UDOM was computed by use of a basic 3-component reverse mixing model (Nelson and Baldock, 2005) $\text{UDOM} = n_1 \cdot \text{HPS} + n_2 \cdot \text{CRAM} + n_3 \cdot \text{peptides}$, based on the assumption that all nitrogen in UDOM is proteinaceous. The resulting histogram ^{13}C NMR spectra of a 3-component surface UDOM ($n_1 = 0.601$, $n_2 = 0.229$, $n_3 = 0.236$) and a 3-component deep UDOM ($n_1 = 0.271$, $n_2 = 0.509$, $n_3 = 0.217$) demonstrated that CRAM is a major constituent of both surface and deep UDOM. A better conformity of experimental and computed, data reduced, ^{13}C NMR spectra of surface and deep UDOM was obtained at smaller values of n_3, suggesting the occurrence of some nitrogen in UDOM in non-protein/peptide chemical environments (Benner and Kaiser, 2003).

3.1.2. Key NMR data to establish the existence of CRAM in UDOM

Two key observations were critical for establishing the existence of CRAM in UDOM. The first observation was based on the nitrogen content of UDOM relative to the quantity of the carbonyl derivative (i.e. carboxylic acid, ester, amide) resonance F_C. The C/N atomic ratios of surface and deep UDOM were 16.1 and 18.4, respectively. Assuming all nitrogen is amide, the N/C atomic ratio defines the fraction of the total ^{13}C NMR integral in amide form. Calculations indicated 6.2% and 5.4% of the total ^{13}C NMR resonances of surface and deep UDOM were represented by amides; i.e. a C/N atomic ratio of (16.1 s_UDOM/18.4 d_UDOM) corresponds to a percent N/C ratio of 6.4 for s_UDOM and 5.7 for d_UDOM, respectively (cf. Fig. 1). This is considerable less carbon NMR resonance integral than found in $F2_C$ of s_UDOM (17.7%) and d_UDOM (21.4%), respectively. Amide only accounted for about one third of the $F2_C$ resonance in surface UDOM and one fourth of it in deep UDOM. Consequently, a substantial excess of carboxyl (ester or acid) resonance was present in the ^{13}C NMR spectra of both surface and deep UDOM.

The second key observation came from ^{1}H NMR spectra of deep UDOM acquired under conditions of slow chemical exchange in dry DMSO-d_6 under total exclusion of moisture, using vacuum-line techniques. Here both non-exchangeable (typically protons attached to carbon atoms) and exchangeable protons (typically protons attached to heteroatoms, like O, N, S) produce recognizable NMR resonances. Under these conditions, NMR peaks of exchangeable protons show appreciable line-broadening (10–200 Hz). Accordingly, the direct observation of a sizeable carboxylic acid resonance (Fig. 4) demonstrated the presence of carboxylic acids rather than esters in deep UDOM.

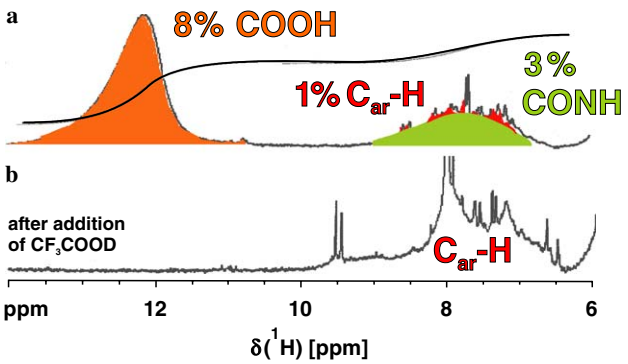


Fig. 4. The chemical exchange of labile protons in deep UDOM is slowed down sufficiently in DMSO-d_6 under total exclusion of moisture to allow the observation (a) of both exchangeable [carboxyl, 8% of total ^{1}H NMR integral (orange); peptide, 3% of total ^{1}H NMR integral (green)] and non-exchangeable [C_{ar} – 1% of total ^{1}H NMR integral (red)] protons. Consecutive addition of CF_3COOD combines all labile proton NMR resonances into a single far downfield signal (b; here: $\delta(^1H) = 14.7$ ppm). The chemical shift of non-exchangeable aromatic protons in deep UDOM is affected by pH variation, which causes alterations in charge, conformation and relative orientation of UDOM constituent molecules.

3.1.3. Comparative analysis of one- and two-dimensional NMR spectra of UDOM

Cross peaks in two-dimensional heteronuclear single quantum coherence (HSQC) NMR spectra relate directly bound carbon and hydrogen atoms. In analogy to procedures established in 1D NMR spectra of NOM and UDOM (Fig. 1), chemical shift ranges (areas) corresponding to fundamental UDOM substructures can be also defined in HSQC NMR spectra (Hertkorn et al., 2002). Seven areas of chemical shift were used to define major UDOM substructures (Fig. 5). In order of increasing chemical shift, area 1 defined methyl bound to carbon and sulfur (dotted circle), and in its lower left corner, branched purely aliphatic CH pairs and polymethylene ($\delta_{H/C} = 1.2/29$ ppm). Area 2 represented a complex set of methylene and methine cross peaks without direct bonds to heteroatoms (except purely aliphatic amines, which are considered as rather elusive constituents of UDOM). Low intensity methoxyl cross peaks occur in area 3, and cross peaks representative mainly of CH-α in proteins and vicinal dicarboxylic acids occurred in area 4. Very highly superimposed carbohydrate methylene cross peaks in area 5, carbohydrate methine cross peaks in area 6 and better resolved, but still strongly overlapping anomeric CH cross peaks in area 7 all indicated the complex structure of HPS.

This analysis is corroborated by one and two dimensional multiplicity edited ^{13}C NMR spectra of deep UDOM, in which the superposition of the ^{13}C DEPT NMR spectra provided a ratio of methyl:methylene:methine:quaternary C of 7:20:46:27 (Fig. 6). The large chemical shift ranges of methyl, methylene and methine in surface and deep UDOM provided further indications of the diverse chemical environments within CRAM.

Aliphatic branching imparts predictable chemical shifts in ^{13}C NMR spectra useful for structural interpretation. Methylation in α- and β-positions induces a downfield ^{13}C NMR chemical shift of ~8 ppm each, while γ-methylation induces an upfield chemical shift of ~4 ppm (Kalinowski et al., 1984). δ- and more remote alkylation has marginal effects and can be neglected (Kalinowski et al., 1984). In contrast, the effects of aliphatic branching on proton chemical shifts are considerably less pronounced and less regular. In general, steric crowding in aliphatic compounds induces downfield proton chemical shift. A substantial degree of branching can be deduced from the considerable range (>25 ppm) of methyl chemical shifts in the edited one dimensional ^{13}C NMR spectrum of deep marine UDOM (Fig. 6). A similar reasoning applies to methylene and methine chemical environments in both one and two dimensional NMR spectra. The near Gaussian distribution of the ^{13}C NMR chemical shifts within the HSQC cross peak area 2 (Fig. 5) indicated broad substitution patterns in CRAM.

Another indication of the high variability of chemical environments in CRAM was the considerable range of chemical shift (~20 ppm) and the near Gaussian distribution of intensity of the carboxyl resonance $F2_C$ in one-dimensional ^{13}C NMR spectra of UDOM (Fig. 1). Extensively α- and β-methylated chemical environments in aliphatic carboxylic acids were not common in CRAM, as indicated by the absence of extreme ^{13}C NMR downfield HSQC cross peaks which would fall outside the boundaries observed for HSQC area 2 in Fig. 5. It appeared that moderate branching and a fairly even distribution of carboxyl groups most accurately described the chemical environment in CRAM.

The positions and amplitudes of HSQC cross peaks allowed discrimination between peptide and aliphatic polycarboxylic acid chemical environments, both of which contributed to the considerable expansion of HSQC area 2 and methylene cross peaks (Figs. 5, 6, and 8a). The HSQC cross peaks of peptides occupied only the ^{13}C NMR upfield part of area 2 HSQC cross peaks (Figs. 5 and 8b; Hertkorn et al., 2002). Therefore, another component was required to complement the remaining downfield ^{13}C NMR area 2 HSQC cross peaks in UDOM. This component was consistent with an aliphatic, polycarboxylic-acid structure, such as CRAM.

3.1.4. Assessment of the proton and carbon NMR chemical shift space of a computer-generated aliphatic model polycarboxylic acid $C_{644}H_{1142}O_{146}$, a necessary prerequisite to evaluate the chemical environments of CRAM within UDOM

A trustworthy resonance assignment in NMR spectra of complex unknowns typically requires a comparative assessment of experimental and calculated NMR data, based on judiciously selected model structures. Nowadays, proton and carbon NMR spectra of extended spin systems can be computed, based on empirical correlations, with reasonable accuracy on desktop computers. An iterative analysis

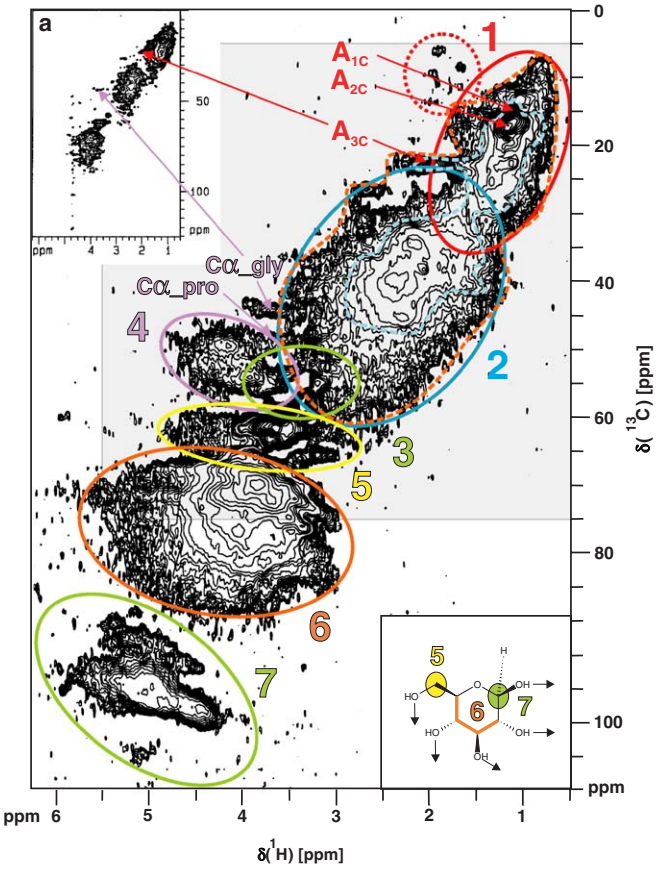


Fig. 5. Two dimensional NMR spectra of UDOM: ^{1}H, ^{13}C HSQC NMR spectra of surface UDOM with seven groups of major constituents (cf. text) and of deep UDOM (insert a) top left: the relative cross peak amplitudes within surface- and deep UDOM HSQC NMR spectra reflect both the variable relative fractions and NMR relaxation characteristics of UDOM constituents. A faster transverse relaxation of CRAM as compared to HPS is observed as well as a faster transverse relaxation of the HPS and peptide components in deep UDOM compared to that in surface UDOM (e.g. disappearance of areas 4 and 7 cross peaks in deep UDOM). That indicates a restricted flexibility and possibly an enhanced contribution of cross linkage at depth as observed with recent mass spectrometric results (Boon et al., 1998). The grey shaded section corresponds to the spectral range shown in Fig. 8.

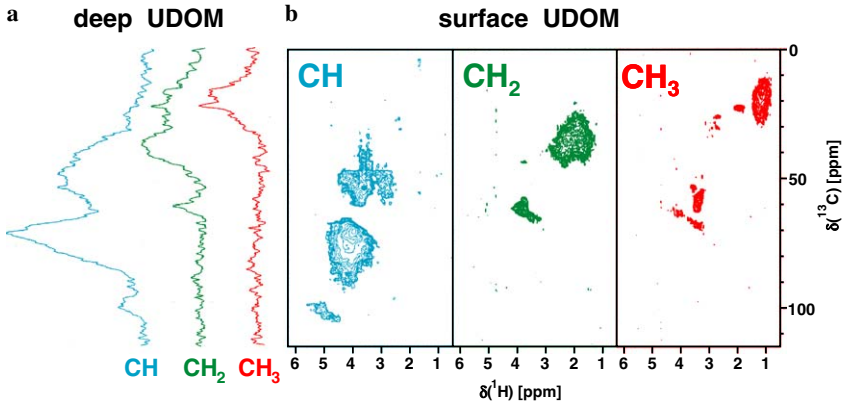


Fig. 6. Scaled multiplicity edited one dimensional (DEPT) ^{13}C NMR subspectra of deep UDOM with methyl (red; 7% of total carbon), methylene (green; 20% of total carbon) and methine (blue; 46% of total carbon) carbon shown. Two dimensional (^{1}H,^{13}C DEPT-HSQC) NMR spectra of surface UDOM (surface UDOM exhibits slower NMR transverse relaxation as compared with deep UDOM, Fig. 5): methyl (red), methylene (green) and methine (blue).

of model and experimental NMR spectra, corroborated by the current understanding of NOM bio- and geosynthetic pathways, can be used to propose model structures conforming to spectral data and to discriminate among several proposed NOM substructures (cf. Fig. 13). Although this initial procedure requires verification by independent and complementary analytical data, it nevertheless provides very significant structural detail at atomic resolution, which is not available by any other analytical technique at present.

The very low resolution NMR signature of CRAM, which covers a substantial chemical shift range in both ^{1}H and ^{13}C NMR spectra (Figs. 1, 5, and 6), already indicates a fairly complex aliphatic polycarboxylic acid structure with considerable intrinsic heterogeneity. For the evaluation of ^{1}H and ^{13}C NMR properties of CRAM, a model open chain aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$ with no unsaturation other than derived from carboxylic groups was computer-generated in a way to allow for systematic increase in aliphatic branching as well as increase in the spatial separation of carboxylic groups (Fig. 7).

That modeled molecule $C_{644}H_{1142}O_{146}$ defines a NMR chemical shift space for open chain branched polycarboxylic acids at the level of connectivities (i.e. definition of chemical bonds) and serves to evaluate the NMR features of UDOM, which are attributed to CRAM (cf. Fig. 7). The absence of *model* acid HSQC cross peaks in the HSQC NMR spectra of surface and deep UDOM indicates non-occurrence or non-significance of respective *model* acid partial structures in the respective NOM material (see caption of Fig. 5). Full methylation ($m = 2$) in $C_{644}H_{1142}O_{146}$ leads to the disappearance of HSQC cross peaks for the respective chain positions, but induces strong deshielding on $\delta(^{13}C)$ and minor effects on $\delta(^{1}H)$ to neighboring positions up to three bonds away.

Accordingly, multiply methylated α- and β-carbon atoms (representing extensively branched non-functionalized aliphatics) with characteristic downfield carbon chemical shift and highly shielded protons, can be discarded as prominent constituents of CRAM (cf. Figs. 5 and 8). Extensive branching is also not consistent with the low abundance of purely aliphatic (i.e. non-carboxylated) carbon atoms in CRAM and with the limited occurrence of methyl groups (Fig. 6).

Instead, carboxylic groups appear to be not more than 2–3 positions distant from aliphatic carbon (i.e. in α-, β-, γ-positions, respectively). A near random occurrence of

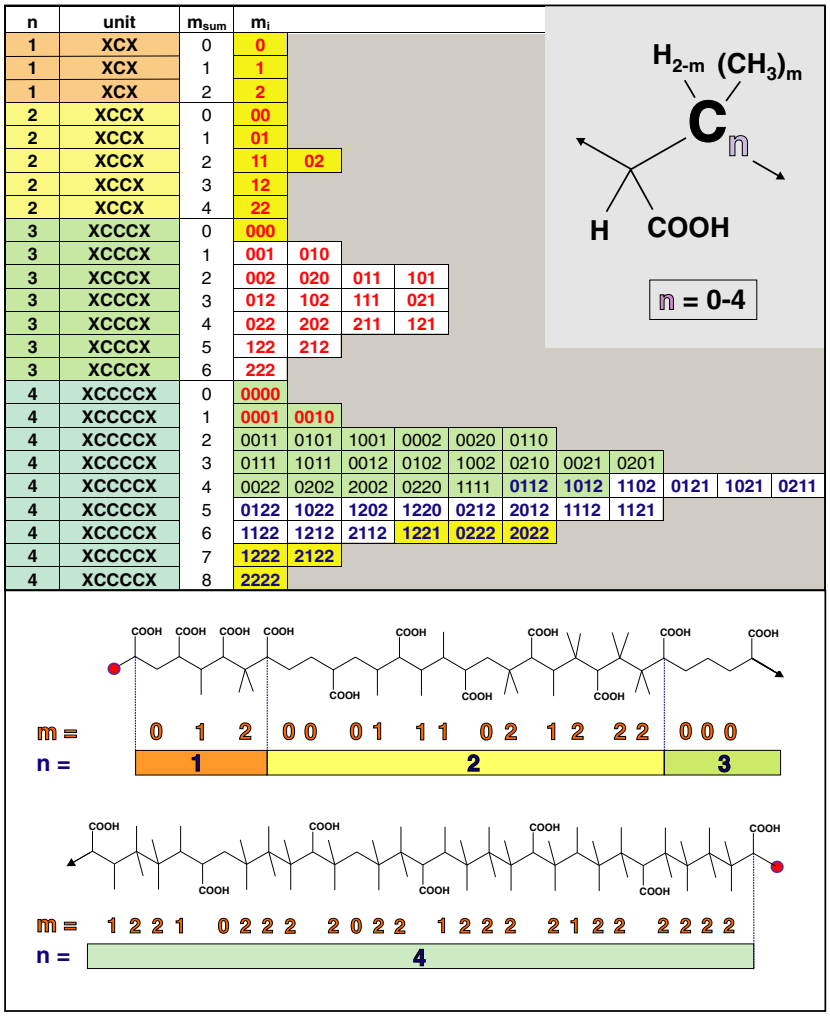


Fig. 7. Model open chain aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$, with a variable spacing between H—C—COOH groups, XC_nX, X = H—C—COOH, $n = 1$–4 (left column) and variable patterns of methylation ($m_i = 0$, 1, 2; m_{sum} equals sum of methyl groups between adjacent X). ^{1}H and ^{13}C chemical shifts were computed from three overlapping sections (red, $C_{214}H_{366}O_{62}$; green, $C_{212}H_{374}O_{50}$; blue, $C_{262}H_{476}O_{48}$); since the entire molecule was too large for direct calculation. The two terminal sections of the molecule, which have been highlighted in yellow, are shown in the bottom part for better visualization and understanding.

aliphatic branching in CRAM accounts most appropriately for the observed ~25 ppm carbon chemical shift range within the methylene trace of the multiplicity-edited 1D and 2D NMR spectra (Fig. 6) and a similarly large spread of $\delta(^{13}C)$ across the entire area 2 HSQC cross peaks in surface and deep UDOM; respectively (Fig. 5). The relatively minor content of methyl carbon in deep UDOM (7%, Fig. 6) relative to the substantial fractions of methylene and methine carbon (20% and 40%, respectively) already suggests the presence of cyclic structures, because alicyclic ring topology stipulates removal of methyl groups compared with open chain analogues. However, the overall complexity of CRAM chemical structures, and the presence of other key structures in UDOM (peptides and carbohydrates in particular) precludes overly quantitative statements at this stage of NMR analysis. However, peptides are in average characterized by a rather large proportion of methyl groups. According to the amino acid composition of a marine UDOM hydrolysate, the ratio of Cq/CH/CH_2/CH_3 in the peptide fraction is 30/32/22/16 in surface UDOM, and 30/33/23/14 in deep UDOM, respectively (see also Section 3.4).

3.2. Capillary electrophoretic separation of UDOM

NMR provides quantitative information about the composition of CRAM, indicating the near absence of double bond and aromatic hydrogen atoms. However, it cannot unequivocally discriminate between alicyclic and open chain aliphatic environments (cf. Fig. 8). The NMR-derived structural information about short range molecular order in CRAM is complemented by electrospray ionization (ESI) mass spectrometry data, which provide ultra-high resolution and mass accuracy, leading to the assignment of molecular formulae for hundreds to thousands of UDOM constituent molecules. However, the variable ionization efficiency of UDOM molecules, which is a crucial and structure-dependent parameter in the generation of UDOM ions in a mass spectrometer, has the potential to severely bias mass spectra-derived data about UDOM composition and structure towards efficiently ionisable UDOM constituents.

Here, highly complementary data from capillary electrophoretic separation of UDOM offer genuine options to validate mass spectral data of UDOM ion composition and structure against NMR derived molecular-level structural information. The detection of a molecule after chromatographic or electrophoretic separation offers intrinsic structural information, since the electrophoretic mobility as well as the chromatographic retention time depend on molecular properties.

The surface and deep UDOM samples showed structure-specific bimodal capillary zone electrophoretic (CZE) mobility profiles in both CZE/UV hyphenation with short wavelength (214 nm) detection, which is considered as rather unselective to structural details, and in CZE/ESI-MS with mass spectrometric detection (Fig. 9). A larger fraction of highly charged aliphatic molecules was observed in deep UDOM (Schmitt-Kopplin and Junkers, 2003; Schmitt-Kopplin and Kettrup, 2003). The effective mobilities for deep UDOM fell in the range

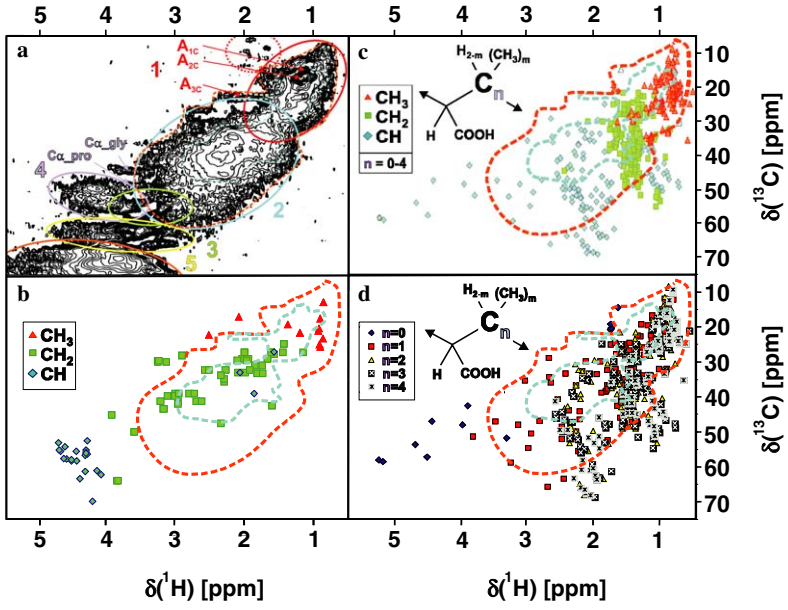


Fig. 8. (a) Upfield section of the $^{1}H,^{13}C$ HSQC NMR spectrum of surface UDOM (partial plot of Fig. 5) (b)∗, multiplicity edited chemical shifts of $^{1}H,^{13}C$ HSQC cross peaks of all 20 proteinaceous amino acids in peptides, following alanine (Hertkorn et al., 2002); (c)∗, HSQC cross peaks of a randomly branched model aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$ ({—[$(CH_3)_l$—C—COOH]—$C_n(CH_3)_m$—} with $l = 0, 1$; $m = 0$–2, and $n = 0$–4, cf. text below), displayed according to (c) carbon multiplicity and (d) according to the separation of carboxylic substituents. (∗ contour lines indicate aliphatic HSQC cross peaks of surface UDOM at unit and at fivefold intensity; Fig. 5).

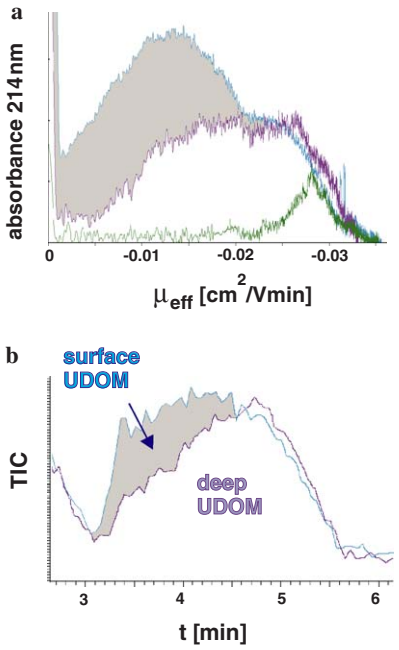


Fig. 9. Electropherograms of surface and deep marine UDOM: (a) Capillary zone electropherogram (CZE; UV detection at 214 nm; effective mobility scale in $cm^2\ V^{-1}\ min^{-1}$) of surface and deep UDOM and a polyacrylic acid standard (2 kDa); (b) superimposed CE-ESI/MS electropherograms of surface and deep UDOM (migration time in min). The congruence of (a) and (b) indicates near uniform ionization efficiency of NOM constituents in marine UDOM.

observed for aliphatic polycarboxylic acids and were consistent with its higher content of CRAM, as indicated by NMR data. The close similarities of the electropherograms obtained at pH 9.4 and 11.4 confirmed the near absence of phenols in UDOM (cf. Fig. 2) and other weakly acidic compounds in both surface and deep UDOM (data not shown). The congruence of electrophoretic mobility profiles in non-selective CZE/UV and mass selective CZE/MS hyphenation indicated only slight variance of electrospray ionization efficiencies within UDOM constituents (Fig. 9). Low mobility constituents, like carbohydrates, showed somewhat lower ionization efficiencies than carboxylic acids, as indicated by the lesser relative intensity of these constituents in surface UDOM compared with deep UDOM. Nevertheless, owing to the overall correspondence of CZE/UV and CZE/MS electropherograms, the mass spectra obtained from marine UDOM were considered representative of UDOM composition.

3.3. FTICR mass spectrometry of UDOM

Electrospray ionization (ESI) mass spectra of UDOM indicated, depending on resolution, hundreds to thousands of individual signals and a conspicuous series of mass spacing patterns separated by 2 and 14 Da (Fig. 10). These patterns are most pronounced in deep UDOM at negative ion ESI with both ion trap (data not shown) and Fourier transformation ion cyclotron resonance (FTICR) mass spectrometry systems. Based on accurate mass measurements obtained from ultra-high-resolution FTICR mass spectra, the most prominent recurring mass differences between these ions were defined as 14.0156 Da [variation in methylene count $(CH_2)_n$], 2.0157 Da (variation in double bond equivalents/H_2), 1.0034 Da (mass difference between ^{13}C and ^{12}C), and 0.0364 Da (formal exchange of CH_4 versus oxygen) (Stenson et al., 2003). A representative subset of the most intense peaks, which were observable as prominent members of these mass spacing patterns in both positive and negative ionization modi, was attributed to CRAM, based on the finding (cf. above), that prominent NMR-detectable features of CRAM would produce strong mass spectroscopic signatures as well. For an in-depth analysis; 613 positive ions XH^+ were selected and the molecular formulae X were determined. These were found to be most accurately represented by the composition $C_nH_mO_q$, indicating the absence of nitrogen in CRAM and a substantial degree of unsaturation in excess of carboxyl groups. The CH_2-specific Kendrick mass defect analysis (Hughey et al., 2001; Hsu et al., 1992) of CRAM showed classes of compounds varying only in the number of methylene groups and, almost exclusively, relative Kendrick mass defects (Δm) corresponding to variance in unsaturation (double bond equivalents (DBE), $\Delta m = -0.013$ Da) and exchange of oxygen vs. CH_4 ($\Delta m = -0.0364$ Da; Fig. 12).

In a van Krevelen diagram (Visser, 1983; Kim et al., 2003), the area of CRAM-derived elemental compositions was centered along the line connecting the virtual end members CH_2 and CO, respectively, and was clearly distanced from the region of peptides (Fig. 11). CRAM exhibited an O/C ratio in excess of 0.25 and was devoid of extended aromatic systems and carboxylated, saturated hydrocarbons. Lower-molecular-weight CRAM exhibited a higher degree of oxidation than higher-molecular-weight molecules (Fig. 11), suggesting a size-reactivity continuum similar to that observed previously in marine DOM (Amon and Benner, 1996; Benner, 2002). All molecular formulas derived from these ions exhibited a higher degree of unsaturation than could be introduced by carboxyl groups alone, even if all oxygen was assumed in carboxyl groups (cf. purple line in Fig. 11). NMR data confirm the very low abundance of hydrogen in aromatics and double bonds in both surface and deep UDOM (<1.5% of total ^1H NMR integral; Figs. 1 and 2).

3.3.1. CH_2 and CO_2 based Kendrick mass analysis of CRAM

Kendrick mass defect analyses of complex mixtures are not restricted to CH_2 units, but are also applicable to any other fragment (Kim et al., 2003). Based on the large amplitude of the $F2_C$ resonance in ^{13}C NMR spectra of UDOM and CRAM (which has been established to represent mainly carboxylic acids, cf. NMR discussion, Section 3.1, Figs. 1 and 2), CO_2 is another key building block of CRAM. It is therefore reasonable to expand the Kendrick

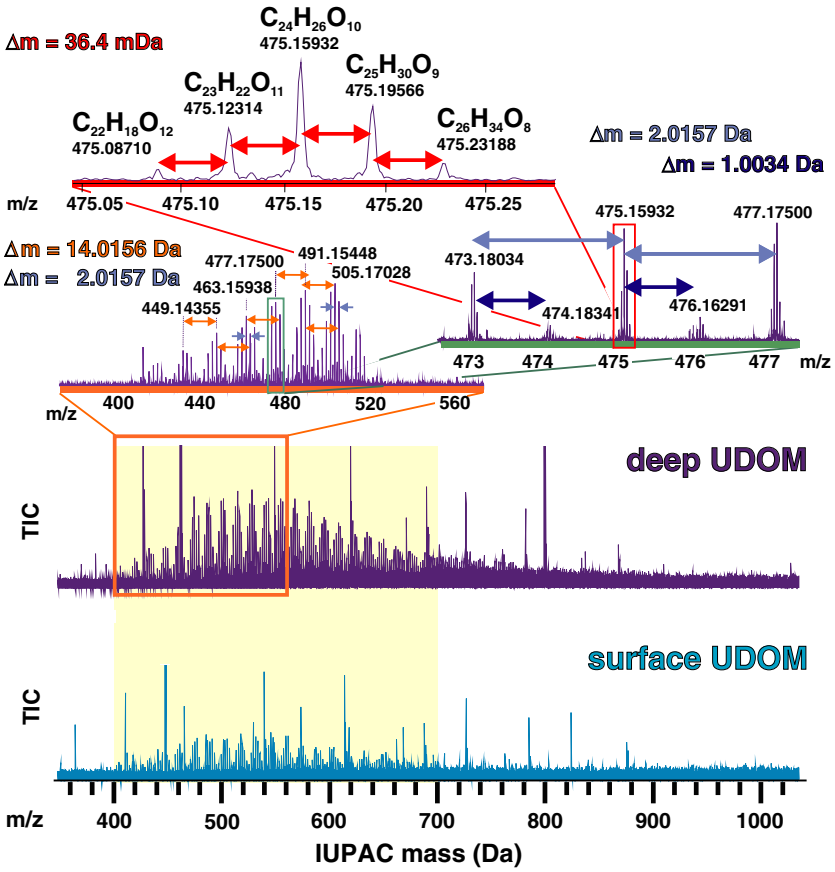


Fig. 10. ESI positive ion 7 Tesla FTICR mass spectra of (bottom, left) surface and (middle left) deep UDOM with selected expanded sections (top left), given to demonstrate the very high resolution in these spectra; the most significant recurring mass differences (Δm = 14.0156, 2.0157, 1.0034 Da, 36.4 mDa) have been explained in the text. In the uppermost expansion, molecular formulae $C_nH_mO_q$ are given for prominent peaks. The yellow shaded mass section equals the mass range used for the Kendrick mass analysis (Fig. 11).

mass defect sorting scheme (Fig. 12) to both key fragments of CRAM, namely CH_2 and CO_2. Since the IUPAC mass scale is based on ^{12}C (IUPAC mass: 12.00000), the presence of hydrogen (^{1}H) induces positive mass defects and the presence of ^{16}O causes negative mass defects in both IUPAC and Kendrick systems, respectively.

CRAM, that belong to identical CH_2-based Kendrick mass defect classes, produce linear patterns of CO_2-based Kendrick mass defects. Conversely, CRAM with identical CO_2-based Kendrick mass defects, which represent molecules belonging to the same CO_2-based class, produce analogous linear patterns in the CH_2-based Kendrick mass analysis (Fig. 12). However, the significantly larger mass spacing (44 Da for CO_2 instead of 14 Da for CH_2) precludes the occurrence of very extended CO_2-based series within the observed mass range of CRAM. Nevertheless, the importance of CO_2 as a fundamental building block of CRAM is well established by this CO_2-based Kendrick mass analysis, which specifically indicates four series of common CO_2-based Kendrick mass defects (KMD_CO_2: 0.73303, 0.69663, 0.67774, 0.67219 Da, respectively; cf. Fig. 12, and experimental section) with five members (Fig. 12) each.

However, while the occurrence of certain molecular backbones with variable degrees of carboxylation is certainly a realistic option for CRAM, it has to be emphasized that many isomers may contribute to a given molecular composition; these are not resolved by any Kendrick mass sorting scheme.

The Kendrick mass analysis solely relies on molecular formulae (i.e. it is insensitive to isomeric composition) and accordingly, the Kendrick mass defect analysis is governed by the nature and abundance of different atoms present in a molecule. A considerable regularity of CRAM can be deduced as more than 95% of the CH_2-based Kendrick mass defect differences are defined by 13.6 mDa (variation in double bond equivalents DBE) and 36.4 mDa (formal exchange of oxygen and methane in the molecular formula) mass differences.

3.4. NMR properties and prominent structures of CRAM

For any given $C_nH_mO_q$ molecular formula, isomers will produce different NMR signatures, thereby providing extremely useful and indispensable guidance for a molecular-level structural analysis of CRAM. For illustration of this point, the molecular formula $C_{28}H_{32}O_{13}$ has been

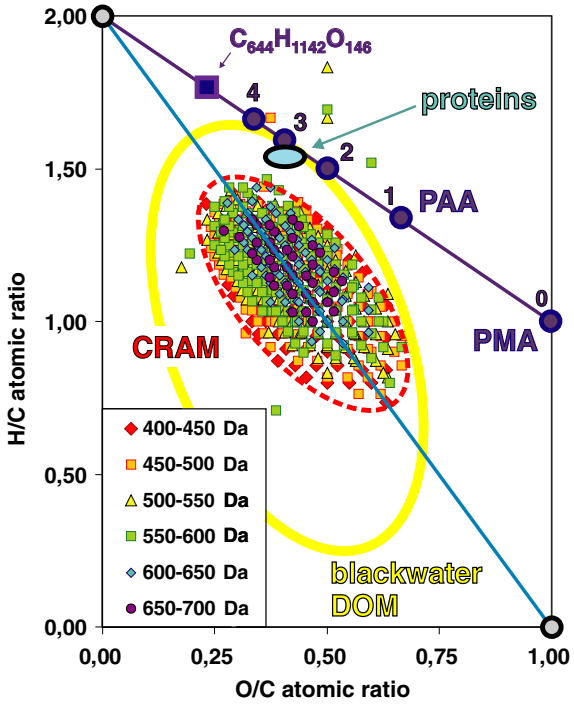


Fig. 11. Mass sorted van Krevelen diagram of 613 ions attributed to CRAM (red ellipsoid, broken line) with region of proteins indicated in blue. The purple line refers to *any* fully saturated open chain aliphatic carboxylic acid, with aliphatic polycarboxylic acids indicated according to spacing between carboxyl groups (cf. Fig. 7); e.g. polymaleic acid (PMA; $n = 0$) polyacrylic acid (PAA; $n = 1$), higher polycarboxylic acids ($n = 2$–4) and model aliphatic polycarboxylic acid $C_{644}H_{1142}O_{146}$ (Fig. 7). The proportion of unsaturation and oxidation in CRAM increases slightly with decreasing mass of the molecular ions. The total area occupied by the 50 Da mass increments shown decreases with increasing mass, indicating an increased diversity of molecular compositions at lower masses. The blue line connects the virtual endmembers CH_2 and CO and bisects the compositional space of CRAM nearly in half. That positioning of CRAM within the van Krevelen diagram is the consequence of a complex relationship between the number of feasible isomers, which can be assembled from molecular formulae $C_nH_mO_q$, the occurrence of these isomers in UDOM, and the ionization efficiency of any single isomer under the given experimental conditions (Hertkorn and Frommberger, 2006). CRAM occupies a minor section of elemental compositions as compared to a blackwater DOM (Hughey et al., 2001).

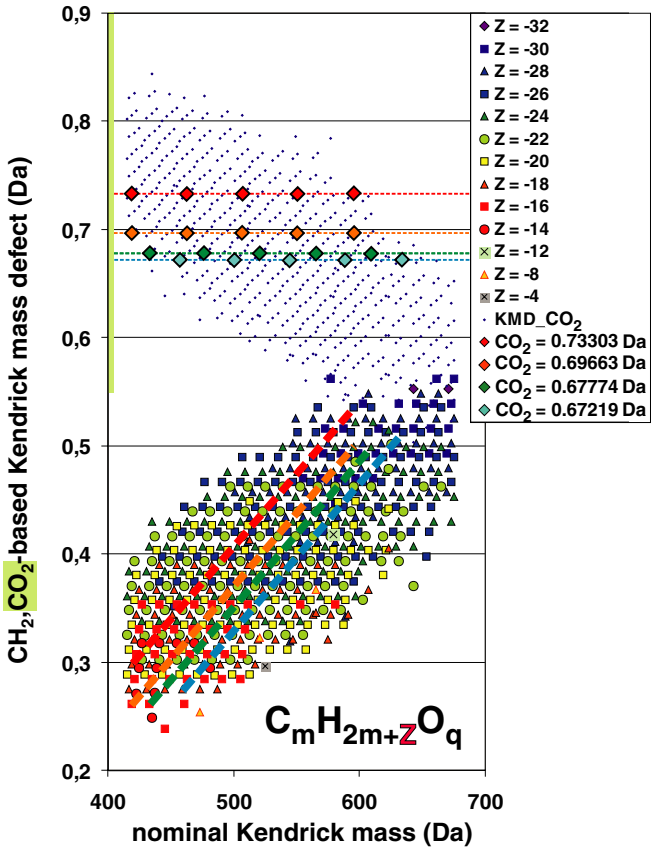


Fig. 12. Combined CH_2-based (bottom) and CO_2-based (top) Kendrick mass defect analysis of 613 CRAM ions in deep UDOM in the mass range of 415–675 Da. The CH_2-based analysis is displayed according to Z-values of the molecular formulae $C_mH_{2m+Z}O_q$, resulting in 13 series of Z-values ($Z = -32$ to 4: Z (type): $C_cH_{2c+Z}O_o$) in 156 homologous series $C_cH_hO_o$ ($c = 19$–37, $h = 14$–50, $o = 6$–16, DBE = 3–21). In the CO_2-based Kendrick mass effect analysis, several series of related CRAM with identity in CO_2-based Kendrick mass defects are recognizable, each showing a mass difference of the key unit CO_2 (44 Da). Caused by the larger mass spacing compared with CH_2 (14 Da), the CO_2-based series in CRAM are more truncated than the CH_2-based series. CRAM are composed of four five-membered series of molecular compositions $C_nH_mO_q$ with identical CO_2-based Kendrick mass defects, indicated by coloured diamonds (cf. text), 27 series with four members each, 70 series with three members each, and shorter series, respectively, which all are represented by dots. Molecules with identical CO_2-based Kendrick mass defects differ in CH_2-based mass defects, and vice versa; this has been indicated by the four pairs of dotted lines, that represent the four five-membered CO_2-based KMD-series indicated.

selected to represent CRAM. This molecule exhibits a H/C and O/C elemental ratio (H/C = 1.15; O/C = 0.42) close to the locus of all 613 CRAM molecules identified (cf. Fig. 11). Furthermore, $C_{28}H_{32}O_{13}$ represents the center molecule of the singular seven membered CH_4/O isobaric series identified within the FTICR mass spectrum of CRAM.

Analogous considerations as described below would apply for all 613 CRAM molecular compositions identified. These have been found to occupy a certain, however restricted area of compositional space (DBE/C = 0.30–0.68; DBE/H = 0.20–0.95; DBE/O = 0.77–1.75) within the van Krevelen diagram (Fig. 11).

Based on the proton NMR spectra of surface and deep marine UDOM (Figs. 1 and 2), the selected three isomeric molecules of $C_{28}H_{32}O_{13}$ with a IUPAC mass of 576.546 do not contain aromatic or olefinic protons.

Isomer I represents a terpenoid-derived assortment of condensed alicyclic six- and five-membered rings with rather evenly distributed carboxylic acid units (to fulfil the restraints defined by Figs. 6–8). Such a topology corresponds to a minimum occurrence of methyl groups which would terminate any branched alkane unit.

Isomer II contains a fully substituted aromatic ring with no hydrogen atom attached. This feature fulfils the key NMR requirement of the absence of aromatic protons in marine UDOM (Fig. 2), while allowing for some aromatic carbon NMR integral. A characteristic NMR feature of *isomer II* is the considerable amount of aromatic carbon

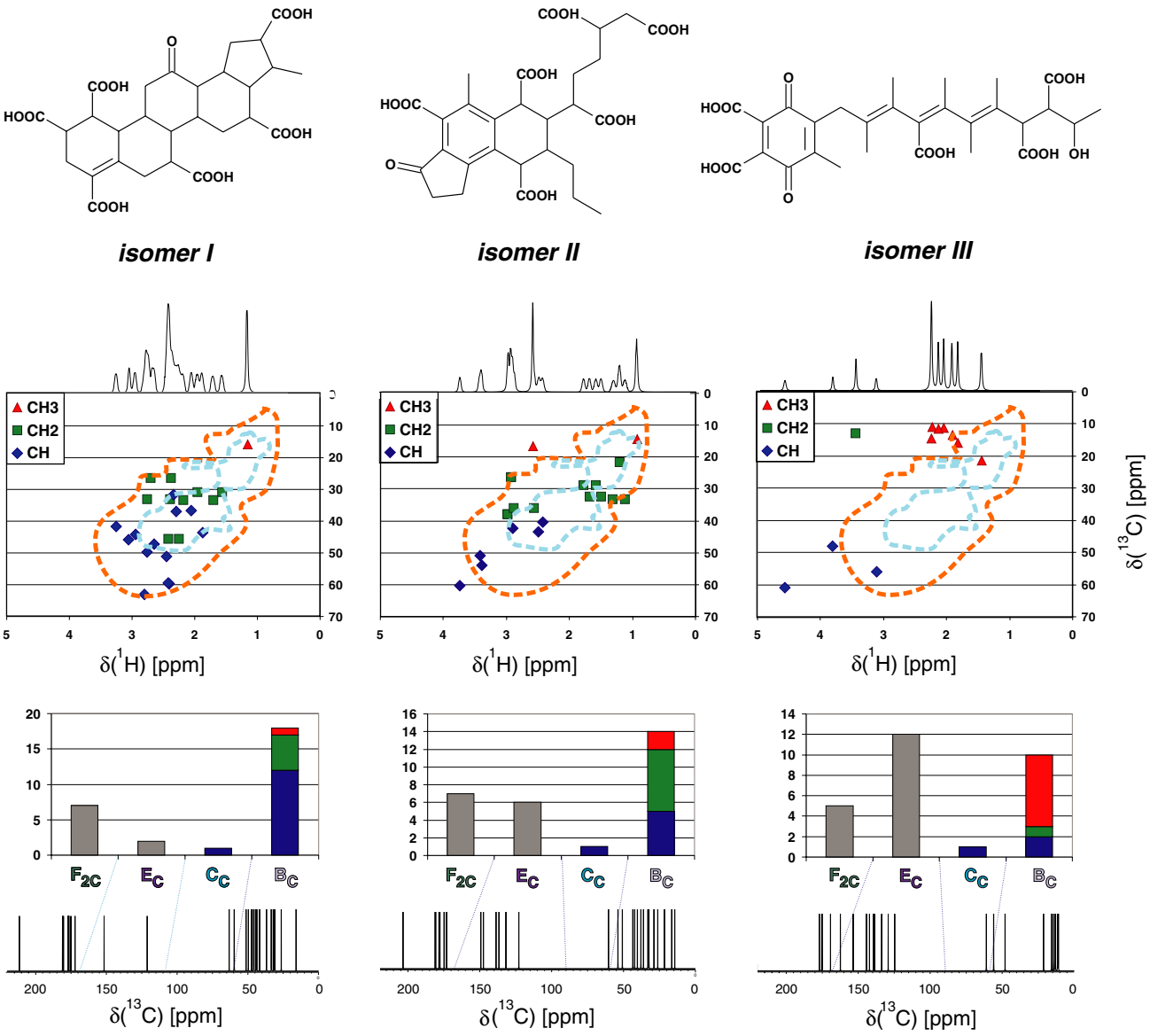


Fig. 13. Three $C_{28}H_{32}O_{13}$ isomers (top row; IUPAC mass 576.546) together with computed ^{1}H NMR (linewidth 10 Hz), ^{1}H, ^{13}C HSQC-spectra (middle row with HSQC contour lines as given in Fig. 5: cross peaks are labelled according to carbon multiplicity) and computed carbon NMR spectra (bottom) with numbers of methyl (red), methylene (green), methane (blue) and quaternary (grey) carbon atoms provided, according to chemical shift ranges given in Fig. 1 (cf. text). *Isomer I* and *isomer II* conform much better to the NMR characteristics of UDOM than *isomer III* [in one-dimensional ^{1}H (Fig. 1), non-edited (Fig. 1) and multiplicity-edited ^{13}C NMR spectra (Fig. 6) and in the ^{1}H, ^{13}C HSQC NMR spectra (Fig. 5)].

NMR resonances (section E_C, Fig. 1) as opposed to *isomer I*. The fraction of methyl groups can be adjusted by a variable degree of branching more easily than in molecules like *isomer I*. However, in molecules resembling *isomer II*, the maximum feasible fraction of methyl carbon still remains rather limited.

Isomer III, which is assembled with partial structures that resembe carotenoids, is almost exclusively composed of methyl and quaternary carbon atoms and exhibits NMR properties, which all are very divergent from those observed in UDOM.

An interesting consequence of the considerable degree of unsaturation and oxygenation in $C_{28}H_{32}O_{13}$ molecules is a relative downfield proton NMR shift of the methyl resonances, especially in structures related to *isomer III*, compared to methyl resonances found in peptides and proteins or other alkane environments. This observation, which refers also to other conceivable structures derived from the 613 CRAM ions, implies a sizable proportion of the methyl resonances in UDOM being derived from peptides rather than from CRAM. Thus, CRAM structures related to *isomers I* and *II* are likely to be more common than those derived from *isomers III*. Both *isomers I* and *II* also conform better to the ^{1}H, ^{13}C HSQC NMR cross peak pattern observed in marine UDOM than does *isomer III* (areas 1 and 2; Fig. 5).

4. Conclusions

Based on all these data, it appears the overwhelming majority of CRAM-ions is represented by carboxylated and fused alicyclic rings with very few hydrogen atoms in double bonds. The biochemical origins and mechanisms of formation of CRAM are unclear, but CRAM shares some structural characteristics found in terpenoids, a diverse class of biochemicals (e.g. isoprenoids, sterols, hopanoids, crenarchaeols) occurring as membrane constituents and secondary metabolites in a wide range of prokaryotic and eukaryotic organisms (Ourisson et al., 1987). Several studies have noted the occurrence of prokaryotic cell wall and membrane components in marine DOM (Tanoue et al., 1995; McCarthy et al., 1998; Benner and Kaiser, 2003; Wakeham et al., 2003), so they are likely sources of CRAM as well. The molecular formulae determined from the FTICR mass spectra data indicated the occurrence of polycarboxylated fused-ring systems rather than open chain isomers. These polycarboxylated fused-ring systems form the general structures of sterols and hopanoids with oxidized side chains.

Similar structures could also be derived from other biochemical precursors that have undergone more extensive alterations. Acrylic acid is a decomposition product of dimethylsulphoproprionate, which is produced by several abundant marine phytoplankton (Dacey and Wakeham, 1986). Acrylic acid is susceptible to addition reactions and photo-polymerization in seawater and could also be a precursor of CRAM. Furthermore, radical mediated photo-oxidation of any functionalized organic molecule in oxygenated surface ocean waters is expected to produce oxidation products with non-repetitive structures (Schmitt-Kopplin et al., 1998).

CRAM are the most abundant identified component of DOM in the deep ocean. Previous studies characterized only ~3% of DOC as specific biochemicals in the deep ocean (Benner, 2002) whereas CRAM accounts for ~8% of the DOC. The structural diversity found within CRAM and its substantial content of alicyclic rings and branching contribute to its resistance to biodegradation and refractory nature. It appears CRAM are largely comprised of the decomposition products of biomolecules, as indicated by its prevalence of carboxyl groups and pattern of increasing oxidation with decreasing molecular size. CRAM are expected to constitute a strong ligand for metal binding, and multiple coordination across calcium cations could promote aggregation and gel formation (Chin et al., 1998), thereby affecting the reactivity of organic matter and the bioavailability of associated nutrients and trace metals.

The occurrence of CRAM in freshwater and terrestrial environments seems likely, considering the global distribution of biomolecules and the similarities of biogeochemical processes among environments. In addition to its abundance in the ultrafiltered fraction of DOM in the ocean, CRAM also appears to be an important component of the humic fraction of marine (Hedges et al., 1992) and freshwater DOM (Leenheer et al., 2003).

The very low resolution signature of CRAM in both separation and spectroscopic techniques has precluded so far its recognition as a relevant constituent of natural organic matter, although the occurrence of sizable amounts of branched aliphatic structures in marine (Hatcher et al., 1979; Hatcher et al., 1981; Hedges et al., 1992; Sardessai and Wahidullah, 1998) and freshwater (Wilson et al., 1981; Buddrus and Lambert, 1995; Lambert and Buddrus, 1996) NOM materials has been proposed in numerous earlier NMR studies. Terpenoid structures have been characterized in fossil organic matter, such as kerogen and petroleum, indicating their ubiquitous and refractory nature (Ourisson et al., 1987; Summons, 1993). Unraveling the mysteries about its formation and cycling will provide critical insights into the global carbon cycle.

Acknowledgments

The authors gratefully acknowledge a Humboldt research award for J.I.H., funding from the U.S. NSF for R.B. and K.K., and valuable discussions with K. Whitehead, Th. Dittmar, B. Koch, E. M. Perdue and Y. Gelinas. We thank H. Neumeir, E. Holzmann and S. Thaller (GSF) for skillful technical assistance during measurements and preparation of the manuscript.

Associate editor: H. Rodger Harvey

Appendix A

Experimentally determined and theoretical mass numbers from positive ion 7 Tesla Fourier transform ion cyclotron mass spectra of deep marine UDOM together with molecular formula $C_cH_hO_o$, representing the minimal deviation from the theoretical mass values with the options $C_{0–50}H_{0–100}O_{0–50}N_{0–50}S_{0–10}$ provided prior to calculation

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
1	415,10141	415,10236	2,29	21	18	9
2	415,13797	415,13874	1,85	22	22	8
3	415,17405	415,17513	2,6	23	26	7
4	417,08070	417,08162	2,21	20	16	10
5	417,11720	417,11801	1,94	21	20	9
6	417,15349	417,15439	2,16	22	24	8
7	417,19014	417,19078	1,53	23	28	7
8	419,09650	419,09727	1,84	20	18	10
9	419,13277	419,13366	2,12	21	22	9
10	419,16910	419,17004	2,24	22	26	8
11	419,20525	419,20643	2,81	23	30	7
12	421,07604	421,07654	1,19	19	16	11
13	421,11206	421,11292	2,04	20	20	10
14	421,14848	421,14931	1,97	21	24	9
15	421,18488	421,18569	1,92	22	28	8
16	423,09139	423,09219	1,89	19	18	11
17	423,12766	423,12857	2,15	20	22	10

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
18	423,16415	423,16496	1,91	21	26	9
19	423,20011	423,20194	4,32	22	30	8
20	425,10680	425,10784	2,45	19	20	11
21	425,14331	425,14422	2,14	20	24	10
22	425,17971	425,18061	2,12	21	28	9
23	427,12243	427,12349	2,48	19	22	11
24	427,13816	427,13874	1,36	23	22	8
25	427,15946	427,15987	0,96	20	26	10
26	427,17417	427,17513	2,25	24	26	7
27	429,11749	429,11801	1,21	22	20	9
28	429,15372	429,15493	2,82	23	24	8
29	429,18993	429,19078	1,98	24	28	7
30	431,09658	431,09727	1,6	21	18	10
31	431,13293	431,13366	1,69	22	22	9
32	431,16917	431,17004	2,02	23	26	8
33	431,20541	431,20643	2,37	24	30	7
34	433,07524	433,07654	3	20	16	11
35	433,11211	433,11292	1,87	21	20	10
36	433,14847	433,14931	1,94	22	24	9
37	433,18493	433,18569	1,75	23	28	8
38	433,22145	433,22208	1,45	24	32	7
39	435,05420	435,05580	3,68	19	14	12
40	435,09141	435,09219	1,79	20	18	11
41	435,12783	435,12857	1,7	21	22	10
42	435,16413	435,16496	1,91	22	26	9
43	435,20050	435,20134	1,93	23	30	8
44	435,23671	435,23773	2,34	24	34	7
45	437,07072	437,07145	1,67	19	16	12
46	437,10700	437,10784	1,92	20	20	11
47	437,14342	437,14422	1,83	21	24	10
48	437,17964	437,18061	2,22	22	28	9
49	437,21682	437,21699	0,39	23	32	8
50	439,08647	439,08710	1,43	19	18	12
51	439,12268	439,12349	1,84	20	22	11
52	439,15890	439,15987	2,21	21	26	10
53	439,19500	439,19626	2,87	22	30	9
54	441,10196	441,10275	1,79	19	20	12
55	441,11771	441,11801	0,68	23	20	9
56	441,13824	441,13914	2,04	20	24	11
57	441,15386	441,15439	1,2	24	24	8
58	441,17399	441,17552	3,47	21	28	10
59	441,18987	441,19078	2,06	25	28	7
60	443,09775	443,09727	−1,08	22	18	10
61	443,13321	443,13366	1,02	23	22	9
62	443,16940	443,17004	1,44	24	26	8
63	443,20580	443,20643	1,42	25	30	7
64	445,11250	445,11292	0,94	22	20	10
65	445,14867	445,14931	1,44	23	24	9
66	445,18507	445,18569	1,39	24	28	8
67	445,22147	445,22208	1,37	25	32	7
68	445,25768	445,25846	1,75	26	36	6
69	447,09165	447,09219	1,21	21	18	11
70	447,12791	447,12857	1,48	22	22	10
71	447,16426	447,16496	1,57	23	26	9
72	447,20056	447,20134	1,74	24	30	8
73	449,07084	449,07145	1,36	20	16	12
74	449,10727	449,10784	1,27	21	20	11
75	449,14355	449,14422	1,49	22	24	10
76	449,17993	449,18061	1,51	23	28	9
77	449,21598	449,21699	2,25	24	32	8
78	451,08662	451,08710	1,06	20	18	12
79	451,12281	451,12349	1,51	21	22	11
80	451,15921	451,15987	1,46	22	26	10
81	451,19554	451,19626	1,6	23	30	9
82	453,10227	453,10275	1,06	20	20	12
83	453,13847	453,13914	1,48	21	24	11
84	453,15372	453,15439	1,48	25	24	8
85	453,17438	453,17552	2,52	22	28	10
86	455,08099	455,08217	2,59	19	18	13
87	455,11765	455,11840	1,65	20	22	12
88	455,13374	455,13366	−0,18	24	22	9
89	455,15374	455,15479	2,31	21	26	11
90	455,16949	455,17004	1,21	25	26	8
91	455,19091	455,19117	0,57	22	30	10
92	455,20547	455,20643	2,11	26	30	7
93	457,11283	457,11292	0,2	23	20	10
94	457,14887	457,14931	0,96	24	24	9
95	457,18514	457,18569	1,2	25	28	8
96	457,22123	457,22208	1,86	26	32	7
97	459,09207	459,09219	0,26	22	18	11
98	459,12820	459,12857	0,81	23	22	10
99	459,16443	459,16496	1,15	24	26	9
100	459,20082	459,20134	1,13	25	30	8
101	459,23688	459,23773	1,85	26	34	7
102	461,07133	461,07145	0,26	21	16	12
103	461,10744	461,10784	0,87	22	20	11
104	461,14364	461,14422	1,26	23	24	10
105	461,18001	461,18061	1,3	24	28	9
106	461,21627	461,21699	1,56	25	32	8
107	461,25225	461,25338	2,45	26	36	7
108	463,08673	463,08710	0,8	21	18	12
109	463,12301	463,12349	1,04	22	22	11
110	463,15938	463,15987	1,06	23	26	10
111	463,19570	463,19626	1,21	24	30	9
112	463,23195	463,23264	1,49	25	34	8
113	465,10212	465,10275	1,35	21	20	12
114	465,13850	465,13914	1,38	22	24	11
115	465,17489	465,17552	1,35	23	28	10
116	465,21126	465,21191	1,4	24	32	9
117	467,08175	467,08202	0,58	20	18	13
118	467,11797	467,11840	0,92	21	22	12
119	467,15428	467,15479	1,09	22	26	11
120	467,19052	467,19117	1,39	23	30	10
121	469,09644	469,09767	2,62	20	20	13
122	469,13354	469,13405	1,09	21	24	12
123	469,14928	469,14931	0,06	25	24	9
124	469,16973	469,17044	1,51	22	28	11
125	469,18540	469,18569	0,62	26	28	8
126	469,20560	469,20682	2,6	23	32	10
127	469,22126	469,22208	1,75	27	32	7
128	471,12867	471,12857	−0,21	24	22	10
129	471,16454	471,16496	0,89	25	26	9
130	471,20077	471,20134	1,21	26	30	8
131	471,23708	471,23773	1,38	27	34	7
132	473,10800	473,10784	−0,34	23	20	11
133	473,14410	473,14422	0,25	24	24	10
134	473,18034	473,18061	0,57	25	28	9
135	473,21658	473,21699	0,87	26	32	8
136	473,25286	473,25338	1,1	27	36	7
137	473,27433	473,27451	0,38	24	40	9
138	475,08710	475,08712	0,04	22	18	12
139	475,12314	475,12349	0,74	23	22	11
140	475,15932	475,15984	1,09	24	26	10
141	475,19566	475,19626	1,26	25	30	9
142	475,23188	475,23265	1,62	26	34	8

(*continued on next page*)

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
143	477,06688	477,06637	−1,07	21	16	13
144	477,10250	477,10275	0,52	22	20	12
145	477,13876	477,13914	0,8	23	24	11
146	477,17500	477,17552	1,09	24	28	10
147	477,21150	477,21191	0,86	25	32	9
148	477,24772	477,24829	1,19	26	36	8
149	479,08156	479,08202	0,96	21	18	13
150	479,11793	479,11840	0,98	22	22	12
151	479,15439	479,15479	0,83	23	26	11
152	479,19070	479,19117	0,98	24	30	10
153	479,22700	479,22756	1,17	25	34	9
154	481,09746	481,09767	0,44	21	20	13
155	481,13374	481,13405	0,64	22	24	12
156	481,17007	481,17044	0,77	23	28	11
157	481,20644	481,20682	0,79	24	32	10
158	481,24250	481,24321	1,48	25	36	9
159	483,11270	483,11332	1,28	21	22	13
160	483,14943	483,14970	0,56	22	26	12
161	483,16505	483,16496	−0,19	26	26	9
162	483,18572	483,18609	0,77	23	30	11
163	483,20090	483,20134	0,91	27	30	8
164	483,23722	483,23773	1,06	28	34	7
165	485,14441	485,14422	−0,39	25	24	10
166	485,18052	485,18061	0,19	26	28	9
167	485,21664	485,21699	0,72	27	32	8
168	485,25280	485,25388	2,23	28	36	7
169	487,12357	487,12349	−0,16	24	22	11
170	487,15954	487,15987	0,68	25	26	10
171	487,19590	487,19626	0,74	26	30	9
172	487,23223	487,23264	0,84	27	34	8
173	487,26770	487,26903	2,73	28	38	7
174	489,10270	489,10275	0,1	23	20	12
175	489,13898	489,13914	0,33	24	24	11
176	489,17533	489,17552	0,39	25	28	10
177	489,21167	489,21191	0,49	26	32	9
178	489,24807	489,24830	0,47	27	36	8
179	491,08234	491,08202	−0,65	22	18	13
180	491,11814	491,11840	0,53	23	22	12
181	491,15448	491,15479	0,63	24	26	11
182	491,19085	491,19117	0,65	25	30	10
183	491,22711	491,22756	0,92	26	34	9
184	491,26377	491,26394	0,35	27	38	8
185	493,09760	493,09767	0,14	22	20	13
186	493,13379	493,13405	0,53	23	24	12
187	493,17006	493,17044	0,77	24	28	11
188	493,20635	493,20682	0,95	25	32	10
189	493,24234	493,24321	1,76	26	36	9
190	495,07700	495,07693	−0,14	21	18	14
191	495,11324	495,11332	0,16	22	22	13
192	495,14951	495,14970	0,38	23	26	12
193	495,18578	495,18609	0,63	24	30	11
194	495,22194	495,22247	1,07	25	34	10
195	497,09269	497,09258	−0,22	21	20	14
196	497,12865	497,12897	0,64	22	24	13
197	497,16495	497,16535	0,8	23	28	12
198	497,18045	497,18061	0,32	27	28	9
199	497,20120	497,20174	1,09	24	32	11
200	497,21671	497,21699	0,56	28	32	8
201	497,25254	497,25338	1,69	29	36	7
202	499,16013	499,15984	−0,58	26	26	10
203	499,19612	499,19626	0,28	27	30	9
204	499,23217	499,23265	0,96	28	34	8
205	499,26879	499,26903	0,48	29	38	7

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
206	501,10332	501,10275	−1,14	24	20	12
207	501,13913	501,13918	0,1	25	24	11
208	501,17535	501,17552	0,34	26	28	10
209	501,21164	501,21191	0,54	27	32	9
210	501,24796	501,24830	0,68	28	36	8
211	503,11856	503,11840	−0,32	24	22	12
212	503,15462	503,15479	0,34	25	26	11
213	503,19086	503,19117	0,62	26	30	10
214	503,22729	503,22756	0,54	27	34	9
215	503,26311	503,26395	1,67	28	38	8
216	505,09796	505,09757	−0,77	23	20	13
217	505,13399	505,13405	0,12	24	24	12
218	505,17028	505,17044	0,32	25	28	11
219	505,20659	505,20682	0,46	26	32	10
220	505,24287	505,24321	0,67	27	36	9
221	505,27948	505,27960	0,24	28	40	8
222	507,07665	507,07693	0,55	22	18	14
223	507,11334	507,11332	−0,04	23	22	13
224	507,14965	507,14970	0,1	24	26	12
225	507,18589	507,18609	0,39	25	30	11
226	507,22222	507,22247	0,49	26	34	10
227	507,25851	507,25886	0,69	27	38	9
228	509,09263	509,09258	−0,1	22	20	14
229	509,12912	509,12897	−0,29	23	24	13
230	509,16541	509,16535	−0,12	24	28	12
231	509,20169	509,20174	0,1	25	32	11
232	509,23797	509,23812	0,29	26	36	10
233	511,10839	511,10823	−0,31	22	22	14
234	511,14460	511,14462	0,04	23	26	13
235	511,16061	511,15987	−1,45	27	26	10
236	511,18096	511,18100	0,08	24	30	12
237	511,19629	511,19626	−0,06	28	30	9
238	511,21697	511,21739	0,82	25	34	11
239	511,23274	511,23265	−0,18	29	34	8
240	511,26817	511,26903	1,68	30	38	7
241	513,12388	513,12388	0	22	24	14
242	513,13960	513,13914	−0,9	26	24	11
243	513,16024	513,16027	0,06	23	28	13
244	513,17546	513,17552	0,12	27	28	10
245	513,21171	513,21191	0,39	28	32	9
246	513,24792	513,24830	0,74	29	36	8
247	513,28415	513,28468	1,03	30	40	7
248	515,11859	515,11840	−0,37	25	22	12
249	515,15491	515,15479	−0,23	26	26	11
250	515,19116	515,19117	0,02	27	30	10
251	515,22746	515,22756	0,19	28	34	9
252	515,26413	515,26395	−0,35	29	38	8
253	517,13427	517,13405	−0,43	25	24	12
254	517,17047	517,17044	−0,06	26	28	11
255	517,20683	517,20682	−0,02	27	32	10
256	517,24318	517,24321	0,06	28	36	9
257	517,27953	517,27960	0,14	29	40	8
258	519,11368	519,11332	−0,69	24	22	13
259	519,14982	519,14970	−0,23	25	26	12
260	519,18606	519,18609	0,06	26	30	11
261	519,22248	519,22247	−0,02	27	34	10
262	519,25884	519,25886	0,04	28	38	9
263	521,09248	521,09258	0,19	23	20	14
264	521,12896	521,12897	0,02	24	24	13
265	521,16586	521,16535	−0,98	25	28	12
266	521,20209	521,20174	−0,67	26	32	11
267	521,23841	521,23812	−0,56	27	36	10
268	521,25915	521,25925	0,19	24	40	12

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
269	523,10951	523,10823	−2,45	23	22	14
270	523,14510	523,14462	−0,92	24	26	13
271	523,18113	523,18100	−0,25	25	30	12
272	523,21757	523,21739	−0,34	26	34	11
273	525,12430	525,12388	−0,8	23	24	14
274	525,16078	525,16027	−0,97	24	28	13
275	525,19827	525,19665	−3,08	25	32	12
276	525,21333	525,21191	−2,7	29	32	9
277	525,24989	525,24830	−3,03	30	36	8
278	525,29167	525,29055	−2,13	24	44	12
279	527,15534	527,15479	−1,04	27	26	11
280	527,19211	527,19117	−1,78	28	30	10
281	527,22854	527,22756	−1,86	29	34	9
282	527,26300	527,26395	1,8	30	38	8
283	529,13376	529,13405	0,55	26	24	12
284	529,17101	529,17044	−1,08	27	28	11
285	529,20730	529,20682	−0,91	28	32	10
286	529,24323	529,24321	−0,04	29	36	9
287	529,27869	529,27960	1,72	30	40	8
288	531,11442	531,11332	−2,07	25	22	13
289	531,15010	531,14970	−0,75	26	26	12
290	531,18664	531,18609	−1,04	27	30	11
291	531,22296	531,22247	−0,92	28	34	10
292	531,26012	531,25886	−2,37	29	38	9
293	533,13024	533,12897	−2,38	25	24	13
294	533,16628	533,16535	−1,74	26	28	12
295	533,20258	533,20174	−1,58	27	32	11
296	533,23918	533,23812	−1,99	28	36	10
297	533,27459	533,27451	−0,15	29	40	9
298	535,10793	535,10823	0,56	24	22	14
299	535,14548	535,14462	−1,61	25	26	13
300	535,18169	535,18100	−1,29	26	30	12
301	535,21832	535,21739	−1,74	27	34	11
302	535,25532	535,25377	−2,9	28	38	10
303	537,12430	537,12388	−0,78	24	24	14
304	537,16085	537,16027	−1,08	25	28	13
305	537,19693	537,19665	−0,52	26	32	12
306	537,23294	537,23304	0,19	27	36	11
307	539,10272	539,10315	0,8	23	22	15
308	539,13962	539,13953	−0,17	24	26	14
309	539,17695	539,17592	−1,91	25	30	13
310	539,21414	539,21230	−3,41	26	34	12
311	539,22775	539,22756	−0,35	30	34	9
312	539,26374	539,26395	0,39	31	38	8
313	541,17032	541,17044	0,22	28	28	11
314	541,20747	541,20682	−1,2	29	32	10
315	541,24339	541,24321	−0,33	30	36	9
316	541,28011	541,27960	−0,94	31	40	8
317	543,15041	543,14970	−1,31	27	26	12
318	543,18662	543,18609	−0,98	28	30	11
319	543,22312	543,22247	−1,2	29	34	10
320	543,25982	543,25886	−1,77	30	38	9
321	543,29657	543,29524	−2,45	31	42	8
322	545,13033	545,12897	−2,49	26	24	13
323	545,16683	545,16535	−2,71	27	28	12
324	545,20296	545,20174	−2,24	28	32	11
325	545,23951	545,23812	−2,55	29	36	10
326	545,27471	545,27451	−0,37	30	40	9
327	545,29011	545,28976	−0,64	34	40	6
328	547,11005	547,10823	−3,33	25	22	14
329	547,14564	547,14462	−1,86	26	26	13
330	547,18189	547,18100	−1,63	27	30	12
331	547,21781	547,21739	−0,77	28	34	11
332	547,25477	547,25377	−1,83	29	38	10
333	547,29101	547,29016	−1,55	30	42	9
334	549,08862	549,08750	−2,04	24	20	15
335	549,12464	549,12388	−1,38	25	24	14
336	549,16067	549,16027	−0,73	26	28	13
337	549,19750	549,19665	−1,55	27	32	12
338	549,23387	549,23304	−1,51	28	36	11
339	549,26958	549,26942	−0,29	29	40	10
340	551,10194	551,10315	2,2	24	22	15
341	551,13953	551,13953	0	25	26	14
342	551,17638	551,17592	−0,83	26	30	13
343	551,21312	551,21230	−1,49	27	34	12
344	551,25021	551,24869	−2,76	28	38	11
345	553,11897	553,11880	−0,31	24	24	15
346	553,15550	553,15518	−0,58	25	28	14
347	553,19217	553,19157	−1,08	26	32	13
348	553,20730	553,20682	−0,87	30	32	10
349	553,22778	553,22795	0,31	27	36	12
350	553,24354	553,24321	−0,6	31	36	9
351	553,26390	553,26434	0,8	28	40	11
352	553,28027	553,27960	−1,21	32	40	8
353	555,14861	555,14970	1,96	28	26	12
354	555,18670	555,18609	−1,1	29	30	11
355	555,22287	555,22247	−0,72	30	34	10
356	555,25934	555,25886	−0,86	31	38	9
357	555,29546	555,29524	−0,4	32	42	8
358	557,16619	557,16535	−1,51	28	28	12
359	557,20252	557,20174	−1,4	29	32	11
360	557,23876	557,23812	−1,15	30	36	10
361	557,27509	557,27451	−1,04	31	40	9
362	557,31059	557,31090	0,56	32	44	8
363	559,14624	559,14462	−2,9	27	26	13
364	559,18204	559,18100	−1,86	28	30	12
365	559,21890	559,21739	−2,7	29	34	11
366	559,25526	559,25377	−2,66	30	38	10
367	559,29190	559,29016	−3,11	31	42	9
368	561,12579	561,12388	−3,4	26	24	14
369	561,16193	561,16027	−2,96	27	28	13
370	561,19840	561,19665	−3,12	28	32	12
371	561,23516	561,23304	−3,78	29	36	11
372	561,27149	561,26942	−3,69	30	40	10
373	563,10396	563,10315	−1,44	25	22	15
374	563,14031	563,13953	−1,39	26	26	14
375	563,17699	563,17592	−1,9	27	30	13
376	563,21347	563,21230	−2,08	28	34	12
377	563,25054	563,24869	−3,28	29	38	11
378	563,28464	563,28507	0,76	30	42	10
379	565,11782	565,11880	1,73	25	24	15
380	565,15630	565,15516	−2,02	26	28	14
381	565,19280	565,19157	−2,18	27	32	13
382	565,22903	565,22795	−1,91	28	36	12
383	565,26533	565,26434	−1,75	26	44	13
384	565,28679	565,28547	−2,34	29	40	11
385	567,09784	567,09806	0,39	24	22	16
386	567,13473	567,13445	−0,49	25	26	15
387	567,17112	567,17083	−0,51	26	30	14
388	567,20802	567,20722	−1,41	27	34	13
389	567,22236	567,22247	0,19	31	34	10
390	567,24435	567,24360	−1,32	28	38	12
391	567,25812	567,25886	1,3	32	38	9
392	569,16570	569,16535	−0,61	29	28	12
393	569,20197	569,20174	−0,4	30	32	11

(*continued on next page*)

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
394	569,23903	569,23812	−1,6	31	36	10
395	569,27504	569,27451	−0,93	32	40	9
396	571,14362	571,14462	1,75	28	26	13
397	571,18162	571,18100	−1,09	29	30	12
398	571,21806	571,21739	−1,17	30	34	11
399	571,25458	571,25377	−1,42	31	38	10
400	571,29089	571,29016	−1,28	32	42	9
401	573,12262	573,12388	2,2	27	24	14
402	573,16101	573,16027	−1,29	28	28	13
403	573,19766	573,19665	−1,76	29	32	12
404	573,23382	573,23304	−1,36	30	36	11
405	573,26983	573,26942	−0,72	31	40	10
406	573,30635	573,30581	−0,94	32	44	9
407	575,13984	575,13953	−0,54	27	26	14
408	575,17730	575,17592	−2,4	28	30	13
409	575,21358	575,21230	−2,23	29	34	12
410	575,25001	575,24869	−2,29	30	38	11
411	575,28617	575,28507	−1,91	31	42	10
412	575,32184	575,32146	−0,66	32	46	9
413	577,08381	577,08241	−2,43	25	20	16
414	577,11871	577,11880	0,16	26	24	15
415	577,15671	577,15518	−2,65	27	28	14
416	577,19275	577,19157	−2,04	28	32	13
417	577,22946	577,22795	−2,62	29	36	12
418	577,26537	577,26434	−1,78	30	40	11
419	577,30022	577,30072	0,87	31	44	10
420	579,13479	579,13445	−0,59	26	26	15
421	579,17217	579,17083	−2,31	27	30	14
422	579,20901	579,20722	−3,09	28	34	13
423	579,22957	579,22835	−2,11	25	38	15
424	579,24544	579,24360	−3,18	29	38	12
425	581,14974	581,15010	0,62	26	28	15
426	581,18616	581,18648	0,55	27	32	14
427	581,20139	581,20174	0,6	31	32	11
428	581,22408	581,22287	−2,08	28	36	13
429	581,23839	581,23812	−0,46	32	36	10
430	581,25875	581,25925	0,86	29	40	12
431	581,27367	581,27451	1,45	33	40	9
432	581,29416	581,29564	2,55	30	44	11
433	583,14535	583,14462	−1,25	29	26	13
434	583,18254	583,18100	−2,64	30	30	12
435	583,20355	583,20313	−0,72	27	34	14
436	583,21814	583,21739	−1,29	31	34	11
437	583,25485	583,25377	−1,85	32	38	10
438	583,28889	583,29016	2,18	33	42	9
439	585,16116	585,16027	−1,52	29	28	13
440	585,19810	585,19665	−2,48	30	32	12
441	585,23446	585,23304	−2,43	31	36	11
442	585,27116	585,26942	−2,97	32	40	10
443	585,30574	585,30581	0,12	33	44	9
444	587,11861	587,11840	−0,36	31	22	12
445	587,14058	587,13952	−1,81	28	26	14
446	587,17671	587,17592	−1,35	29	30	13
447	587,21343	587,21230	−1,92	30	34	12
448	587,24992	587,24869	−2,09	31	38	11
449	587,28570	587,28507	−1,07	32	42	10
450	587,32168	587,32146	−0,37	33	46	9
451	589,15645	589,15518	−2,16	28	28	14
452	589,19300	589,19157	−2,43	29	32	13
453	589,22949	589,22795	−2,61	30	36	12
454	589,26553	589,26434	−2,02	31	40	11
455	589,30020	589,30072	0,88	32	44	10
456	589,31508	589,31598	1,53	36	44	7

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
457	591,13625	591,13445	−3,04	27	26	15
458	591,17221	591,17083	−2,33	28	30	14
459	591,20852	591,20722	−2,2	29	34	13
460	591,24503	591,24360	−2,42	30	38	12
461	591,28141	591,27999	−2,4	31	42	11
462	591,31638	591,31637	−0,02	32	46	10
463	593,15127	593,15010	−1,97	27	28	15
464	593,18780	593,18648	−2,23	28	32	14
465	593,22372	593,22287	−1,43	29	36	13
466	593,26065	593,25925	−2,36	30	40	12
467	595,12838	595,12936	1,65	26	26	16
468	595,16610	595,16575	−0,59	27	30	15
469	595,20296	595,20213	−1,39	28	34	14
470	595,24011	595,23852	−2,67	29	38	13
471	595,25475	595,25377	−1,65	33	38	10
472	595,29106	595,29016	−1,51	34	42	9
473	597,14508	597,14501	−0,12	26	28	16
474	597,18310	597,18140	−2,85	27	32	15
475	597,19761	597,19665	−1,61	31	32	12
476	597,21819	597,21778	−0,69	28	36	14
477	597,23416	597,23304	−1,88	32	36	11
478	597,26880	597,26942	1,04	33	40	10
479	597,30616	597,30581	−0,59	34	44	9
480	599,17644	599,17592	−0,87	30	30	13
481	599,21351	599,21230	−2,02	31	34	12
482	599,24976	599,24869	−1,79	32	38	11
483	599,28603	599,28507	−1,6	33	42	10
484	601,19331	601,19157	−2,89	30	32	13
485	601,22963	601,22795	−2,79	31	36	12
486	601,26624	601,26434	−3,16	32	40	11
487	603,13578	603,13445	−2,21	28	26	15
488	603,17236	603,17083	−2,54	29	30	14
489	603,20866	603,20722	−2,39	30	34	13
490	603,24512	603,24360	−2,52	31	38	12
491	603,28114	603,27999	−1,91	32	40	11
492	603,31841	603,31637	−3,38	33	44	10
493	605,15160	605,15010	−2,48	28	28	15
494	605,18814	605,18648	−2,74	29	32	14
495	605,22446	605,22287	−2,63	30	36	13
496	605,26105	605,25925	−2,97	31	40	12
497	607,12946	607,12936	−0,16	27	26	16
498	607,16680	607,16575	−1,73	28	30	15
499	607,20291	607,20213	−1,28	29	34	14
500	607,23938	607,23852	−1,42	30	38	13
501	607,27474	607,27490	0,26	31	42	12
502	607,31231	607,31129	−1,68	32	46	11
503	609,14328	609,14501	2,84	27	28	16
504	609,18237	609,18140	−1,59	28	32	15
505	609,21871	609,21778	−1,53	29	36	14
506	611,16107	611,16066	−0,67	27	30	16
507	611,21373	611,21230	−2,34	32	34	12
508	611,25062	611,24869	−3,16	33	38	11
509	611,28708	611,28507	−3,29	34	42	10
510	613,15710	613,15518	−3,13	30	28	14
511	613,19369	613,19157	−3,46	31	32	13
512	613,22972	613,22795	−2,89	32	36	12
513	615,17093	615,17083	−0,16	30	30	14
514	615,20900	615,20722	−2,89	31	34	13
515	615,24479	615,24360	−1,93	32	38	12
516	615,28079	615,27999	−1,3	33	42	11
517	617,18711	617,18648	−1,02	30	32	14
518	617,22303	617,22287	−0,26	31	36	13
519	617,25840	617,25925	1,38	32	40	12

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
520	617,29684	617,29564	−1,94	33	44	11
521	619,16392	619,16575	2,96	29	30	15
522	619,20248	619,20213	−0,57	30	34	14
523	619,23954	619,23852	−1,65	31	38	13
524	619,27599	619,27490	−1,76	32	42	12
525	619,31316	619,31129	−3,02	33	46	11
526	621,18238	621,18140	−1,58	29	32	15
527	621,21901	621,21778	−1,98	30	36	14
528	621,25546	621,25417	−2,08	31	40	13
529	621,29127	621,29055	−1,16	32	44	12
530	623,18234	623,18140	−1,51	29	34	15
531	623,21901	623,21778	−1,97	30	38	14
532	623,25546	623,25417	−2,07	31	42	13
533	623,29127	623,29055	−1,16	32	46	12
534	625,19679	625,19705	0,42	29	36	15
535	627,20697	627,20722	0,4	32	34	13
536	627,24355	627,24360	0,08	33	38	12
537	627,27981	627,27999	0,29	34	42	11
538	629,18569	629,18648	1,26	31	32	14
539	629,22295	629,22287	−0,13	32	36	13
540	629,25954	629,25925	−0,46	33	40	12
541	629,29493	629,29564	1,13	34	44	11
542	631,16507	631,16575	1,08	30	30	15
543	631,20210	631,20213	0,05	31	34	14
544	631,23845	631,23852	0,11	32	38	13
545	631,27504	631,27490	−0,22	33	42	12
546	631,31125	631,31129	0,06	34	46	11
547	633,18151	633,18140	−0,17	30	32	15
548	633,21794	633,21778	−0,25	31	36	14
549	633,25463	633,25417	−0,73	32	40	13
550	633,28964	633,29055	1,44	33	44	12
551	635,19707	635,19705	−0,03	30	34	15
552	635,23357	635,23343	−0,22	31	38	14
553	635,26882	635,26981	1,56	32	42	13
554	637,21097	637,21270	2,71	30	36	15
555	641,18505	641,18648	2,23	32	32	14
556	641,22181	641,22287	1,65	33	36	13
557	641,25860	641,25925	1,01	34	40	12
558	641,29380	641,29564	2,87	35	44	11
559	643,16556	643,16575	0,3	31	30	15
560	643,20264	643,20213	−0,79	32	34	14
561	643,23783	643,23852	1,07	33	38	13
562	643,27428	643,27490	0,96	34	42	12
563	643,31246	643,31129	−1,82	35	46	11
564	643,34644	643,34767	1,91	36	50	10
565	645,18303	645,18140	−2,53	31	32	15
566	645,21829	645,21778	−0,79	32	36	14
567	645,25538	645,25417	−1,88	33	40	13
568	645,29055	645,29055	0	34	44	12
569	647,19806	647,19704	−1,58	31	34	15
570	647,23523	647,23343	−2,78	32	38	14
571	647,27220	647,26982	−3,68	33	42	13
572	649,17570	649,17631	0,94	30	32	16
573	649,21404	649,21270	−2,06	31	36	15
574	649,25128	649,24908	−3,39	32	40	14
575	651,19075	651,19196	1,86	30	34	16
576	651,22930	651,22835	−1,46	31	38	15
577	651,26566	651,26473	−1,43	32	42	14
578	653,33046	653,33202	2,39	37	48	10
579	655,20084	655,20213	1,97	33	34	14
580	655,23730	655,23851	1,85	34	38	13
581	655,27478	655,27490	0,18	35	42	12
582	655,31179	655,31129	−0,76	36	46	11

Appendix A (*continued*)

Count	A Exp	B Theory	C Error (ppm)	D C	E H (M)	F O
583	657,18263	657,18140	−1,87	32	32	15
584	657,22030	657,21778	−3,83	33	36	14
585	657,25644	657,25417	−3,45	34	40	13
586	657,29314	657,29055	−3,94	35	44	12
587	659,19792	659,19705	−1,32	32	34	15
588	659,23479	659,23343	−2,06	33	38	14
589	659,27151	659,26982	−2,56	34	42	13
590	659,30782	659,30620	−2,46	35	46	12
591	661,17492	661,17631	2,1	31	32	16
592	661,21209	661,21270	0,92	32	36	15
593	661,24844	661,24908	0,97	33	40	14
594	661,28545	661,28547	0,03	34	44	13
595	663,19194	663,19196	0,03	31	34	16
596	663,22682	663,22835	2,31	32	38	15
597	663,26385	663,26473	1,33	33	42	14
598	665,20809	665,20761	−0,72	31	36	16
599	665,24396	665,24400	0,06	32	40	15
600	665,28262	665,28098	−2,47	33	44	14
601	669,25322	669,25417	1,42	35	40	13
602	669,29176	669,29055	−1,81	36	44	12
603	671,19723	671,19705	−0,27	33	34	15
604	671,23379	671,23343	−0,54	34	38	14
605	671,27049	671,26982	−1	35	42	13
606	673,21289	673,21270	−0,28	33	36	15
607	673,24986	673,24908	−1,16	34	40	14
608	673,28672	673,28547	−1,86	35	44	13
609	673,32329	673,32185	−2,14	36	48	12
610	675,19279	675,19196	−1,23	32	34	16
611	675,22942	675,22835	−1,58	33	38	15
612	675,26504	675,26473	−0,46	34	42	14
613	675,30185	675,30112	−1,08	35	46	13

(A) Experimentally determined mass numbers (based on IUPAC nominal mass; $^{12}C = 12.000000$) from the positive ion in atomic mass units; these ions are protonated adducts. (B) Theoretical mass number (based on IUPAC nominal mass) of the respective best hit (range allowed: C: 0–50, H: 0–100, O: 0–50, N: 0–50; other elements allowed) given in atomic mass units. (C) Difference between A and B. (D) Number of carbon atoms in the most likely molecular formula (calculated hit with the least deviation from theoretical values). (E) Number of hydrogen atoms in the most likely molecular formula (calculated hit with the least deviation from theoretical values; the original ion is a protonated adduct). (F) Number of oxygen atoms in the most likely molecular formula (calculated hit with the least deviation from theoretical values).

References

Aluwihare, L.I., Repeta, D.J., Chen, R.F., 1997. A major biopolymeric component to dissolved organic carbon in surface sea water. *Nature* **387**, 166–169.

Aluwihare, L.I., Repeta, D.J., Chen, R.F., 2002. Chemical composition and cycling of dissolved organic matter in the Mid-Atlantic Bight. *Deep-Sea Res. II.* **49**, 4421–4437.

Amon, R.M.W., Benner, R., 1996. Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **41**, 41–51.

Benner, R., 2002. In: Hansell, D.A., Carlson, C.A. (Eds.), *Biogeochemistry of Marine Dissolved Organic Matter*. Academic Press, Amsterdam, pp. 59–151.

Benner, R., Kaiser, K., 2003. Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter. *Limnol. Oceanog.* **48**, 118–128.

Benner, R., Pakulski, J.D., McCarthy, M., Hedges, J.I., Hatcher, P.G., 1992. Bulk chemical characterization of dissolved organic matter in the ocean. *Science* **255**, 1561–1564.

Benner, R., Biddanda, B., Black, B., McCarthy, M., 1997. Abundance, size distribution, and stable carbon and nitrogen isotopic compositions of marine organic matter isolated by tangential-flow ultrafiltration. *Mar. Chem.* **57**, 243–263.

Boon, J.J., Klap, V.A., Eglinton, T.I., 1998. Molecular characterization of microgram amounts of oceanic colloidal organic matter by direct temperature-resolved ammonia chemical ionization mass spectrometry. *Org. Geochem.* **29**, 1051–1061.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W.L., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E., Grainger, D.J., 2002. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using H-1-NMR-based metabonomics. *Nat. Med.* **8**, 1439–1445.

Buddrus, J., Lambert, J., 1995. Isolated paraffinic methyl groups in humic substances. *Org. Geochem.* **23**, 269–271.

Chin, W.C., Orellana, M.V., Verdugo, P., 1998. Spontaneous assembly of marine dissolved organic matter into polymer gels. *Nature* **391**, 568–572.

Dacey, J.W.H., Wakeham, S.G., 1986. Oceanic dimethylsulfide: production during zooplankton grazing on phytoplankton. *Science* **233**, 1314–1316.

Druffel, E.R.M., Williams, P.M., Bauer, J.E., Ertel, J.R., 1992. Cycling of dissolved and particulate organic matter in the open ocean. *J. Geophys. Res.* **97**, 15639–15659.

Hatcher, P.G., Rowan, R., Mattingly, M.A., 1979. ^{1}H and ^{13}C NMR of marine humic acids. *Org. Geochem.* **2**, 77–85.

Hatcher, P.G., Maciel, G.E., Dennis, L.W., 1981. Aliphatic structure of humic acids; a clue to their origin. *Org. Geochem.* **3**, 43–48.

Hedges, I.J., Hatcher, P.G., Ertel, J.R., Meyers-Schulte, K.J., 1992. A comparison of dissolved humic substances from seawater with Amazon River counterparts by ^{13}C NMR spectrometry. *Geochim. Cosmochim. Acta* **56**, 1753–1757.

Hedges, J.I., Eglinton, G., Hatcher, P.G., Kirchman, D.L., Arnosti, C., Derenne, S., Evershed, R.P., Kögel-Knabner, I., de Leeuw, J.W., Littke, R., 2000. The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Org. Geochem.* **31**, 945–958.

Hedges, J.I., 1992. Global biogeochemical cycles: progress and problems. *Mar. Chem.* **39**, 67–93.

Hedges, J.I., Baldock, J.A., Gélinas, Y., Lee, C., Peterson, M., Wakeham, S.G., 2001. Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* **409**, 801–804.

Hertkorn, N., Frommberger, M. (2006). unpublished data.

Hertkorn, N., Permin, A., Perminova, I.V., Kovalevskii, D., Yudov, M., Petrosyan, V., Kettrup, A., 2002. Comparative analysis of partial structures of a peat humic and fulvic acid using one- and two-dimensional nuclear magnetic resonance spectroscopy. *J. Environ. Qual.* **31**, 375–387.

Hsu, C.S., Qian, K., Chen, Y.C., 1992. An innovative approach to data analysis in hydrocarbon characterization by on-line liquid chromatography–mass spectrometry. *Anal. Chim. Acta* **264**, 79–89.

Hughey, C.A., Hendrickson, C.L., Rodgers, P.R., Marshall, A.G., 2001. Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **73**, 4676–4681.

Kaiser, K., Benner, R., (2006). Bacterial contribution to marine organic matter. *Limnol. Oceanogr., in preparation*.

Kalinowski, H.-O., Berger, S., Braun, S., 1984. *^{13}C-NMR-Spektroskopie*. Thieme, Stuttgart, pp. 88–96.

Kim, S., Kramer, R.W., Hatcher, P.G., 2003. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. *Anal. Chem.* **75**, 5336–5344.

Lambert, J., Buddrus, J., 1996. Quantification of isolated methyl groups in aquatic humic substances by means of ^{1}H and ^{13}C NMR spectroscopy. *Magn. Reson. Chem.* **34**, 276–282.

Leenheer, J.A., Nanny, M.A., McIntyre, C., 2003. Terpenoids as major precursors of dissolved organic matter in landfill leachates, surface water, and groundwater. *Environ. Sci. Technol.* **37**, 2323–2331.

McCarthy, M., Hedges, J.I., Benner, R., 1996. Major biochemical composition of dissolved high molecular weight organic matter in seawater. *Mar. Chem.* **55**, 281–297.

McCarthy, M., Hedges, J.I., Benner, R., 1998. Major bacterial contribution to marine dissolved organic nitrogen. *Science* **281**, 231–234.

Nelson, P.N., Baldock, J.A., 2005. Estimating the molecular composition of a diverse range of natural organic materials from solid-state ^{13}C NMR and elemental analyses. *Biogeochemistry* **72**, 1–34.

Ourisson, G., Rohmer, M., Poralla, K., 1987. Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annu. Rev. Microbiol.* **41**, 301–333.

Sardessai, S., Wahidullah, S., 1998. Structural characteristics of marine sedimentary humic acids by CP/MAS ^{13}C NMR spectroscopy. *Oceanol. Acta* **21**, 543–550.

Schmitt-Kopplin, P., Junkers, J., 2003. Capillary zone electrophoresis of natural organic matter. *J. Chromatogr. A* **998**, 1–20.

Schmitt-Kopplin, P., Kettrup, A., 2003. Capillary electrophoresis–electrospray ionization–mass spectrometry for the characterization of natural organic matter: an evaluation with free flow electrophoresis-off-line flow injection electrospray ionization-mass spectrometry. *Electrophoresis* **24**, 3057–3066.

Schmitt-Kopplin, P., Hertkorn, N., Schulten, H.R., Kettrup, A., 1998. Structural changes in a dissolved soil humic acid during photochemical degradation processes under O_2 and N_2 atmosphere. *Environ. Sci. Technol.* **32**, 2531–2541.

Skoog, A., Benner, R., 1997. Aldoses in various size fractions of marine organic matter: implications for carbon cycling. *Limnol. Oceanogr.* **42**, 1803–1813.

Stenson, A.C., Marshall, A.G., Cooper, W.T., 2003. Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **75**, 1275–1284.

Summons, R.E., 1993. In: Engel, M.H., Macko, S.A. (Eds.), *Organic Geochemistry: Principles and Applications Biogeochemical cycles: A Review of Fundamental Aspects of Organic Matter Formation, Preservation, and Composition*. Plenum Press, New York, pp. 3–21.

Tanoue, E., Nishiyama, S., Kamo, M., Tsugita, A., 1995. Bacterial membranes: possible source of a major dissolved protein in seawater. *Geochim. Cosmochim. Acta* **59**, 2643–2650.

Vernonclark, R.N., Goldberg, E.D., Bertine, K.K., 1994. Organic and inorganic characterization of marine colloids. *Chem. Ecol.* **11**, 69–83.

Visser, S.A., 1983. Application of Van Krevelen's graphical–statistical method for the study of aquatic humic material. *Environ. Sci. Technol.* **17**, 412–417.

Wakeham, S.G., Pease, T.K., Benner, R., 2003. Hydroxy fatty acids in marine dissolved organic matter as indicators of bacterial membrane material. *Org. Geochem.* **34**, 857–868.

Willker, W., Leibfritz, D., Kerssebaum, R., Bermel, W., 1992. Gradient selection in inverse heteronuclear correlation spectroscopy. *Magn. Reson. Chem.* **31**, 287–292.

Wilson, M.A., Barron, P.F., Gillam, A.H., 1981. The structure of freshwater humic substances as revealed by ^{13}C NMR spectroscopy. *Geochiom. Cosmochim. Acta* **45**, 1743–1750.