Characterisation of a newly available marine dissolved organic matter reference material (TRM-0522)

Stacey L. Felgate¹, Alexander Craig¹ ², Lindon Moodie² and Jeffrey Hawkes¹*

¹Analytical Chemistry, Department of Chemistry BMC, Uppsala University, Sweden
²Medicinal Chemistry, Department of Chemistry BMC, Uppsala University, Sweden

*Corresponding author: J. Hawkes (jeffrey.hawkes@kemi.uu.se)

Abstract

Recent methodological advances have greatly increased our ability to characterise aquatic dissolved organic matter (DOM) using high resolution instrumentation, including nuclear magnetic resonance (NMR) and mass spectrometry (HRMS). Reliable DOM reference materials are required for further method development and dataset alignment, but do not currently exist for the marine environment. This presents a major limitation for marine biogeochemistry and related fields, including natural product discovery. To fill this resource gap, we have prepared a marine DOM reference material (TRM-0522) from 45 m deep coastal seawater obtained ~ 1 km offshore of Sweden’s west coast. Over 3000 molecular formulas were assigned by direct infusion HRMS, confirming sample diversity, and the distribution of formulas in van Krevelen space was typical for a marine sample, with the majority of formulas in the region H/C 1-1.5 and O/C 0.3-0.7. The extracted DOM pool was more nitrogen (N) and sulfur (S) rich than a typical terrestrial reference material (SRFA). MZmine3 processing of UPLC-HRMS/MS data revealed 494 resolvable features (233 in negative mode; 261 in positive mode) over a wide range of retention times and masses. NMR data indicated low contributions of the aromatic protons, and generally speaking lignin, humic, and fulvic substances associated with terrestrial samples. Instead, carboxylic-rich aliphatic molecules (CRAM) were the most abundant components, followed by carbohydrates and aliphatic functionalities. This is consistent with a very low specific UV absorbance SUVA₂₅₄ value of 1.36 L mg C⁻¹ m⁻¹. When combined with comparisons with existing terrestrial reference materials (SRFA and PLFA), these results suggest that TRM-0522 is a useful and otherwise unavailable reference material for use in marine DOM biogeochemistry.

1. Introduction

Dissolved organic matter (DOM), a complex mixture of thousands if not millions of compounds, is ubiquitous in aquatic systems where it underpins food webs, plays a critical role in the global carbon cycle¹, and may hold countless natural products for use in biotechnology applications². From understanding global biogeochemical cycles to the complex ecological networks corresponding exo-metabolomes which underpin them, it is critical that researchers have the means and understanding with which to adequately
quantify and characterise DOM, but its highly dilute and variable nature means that it is challenging to study, and its chemical composition remains poorly characterised.

In recent years, several advanced methods have been developed which have allowed progress to be made, such as those using nuclear magnetic resonance (NMR) and high-resolution mass spectrometry (HRMS), including tandem mass spectrometry (HRMS/MS) and coupling to ultra high-performance liquid chromatography (UPLC). Complex but reliable reference materials are necessary in order to compare such methods, and critically to evaluate the scope and implications of the resulting datasets that are generated. Typically, materials from the International Humic Substances Society (IHSS) are used, which are long-established, internationally known and trusted, and can be purchased on the scale of hundreds of milligrams. The main downside of these materials is that they are all sourced from terrestrial, often very brown water environments, with the exception of Pony Lake Fulvic Acid (PLFA) which came from an Antarctic lake with no higher order plants, but which has now been depleted with no plans to restock. The introduction of a marine reference material for dissolved organic carbon (DOC) concentration analysis has been very successful, and demonstrates demand for marine-specific reference materials, but is not available at sufficient concentrations or volumes as to be useful as a DOM characterisation standard. Thus, a suitable reference material for marine DOM studies does not currently exist.

The lack of a marine DOM reference material is problematic and hinders method development and dataset alignment in marine biogeochemistry within the DOM characterisation field and in related subjects such as natural product discovery and trace metal speciation analysis. In order to fill this resource gap, we have prepared and characterised a marine DOM reference material from 45 m deep coastal seawater obtained from the Tjärnö Marine Laboratory (Gothenburg University) on the west coast of Sweden. The material was collected in this first effort in May 2022 from 1,500 L seawater using a commercially available aqueous-compatible C18 column (C18-AQ hereafter), and after sample clean-up, totalled 1.06 g in mass. We have given the sample a unique designation ‘Tjarno Reference Material - May 2022’ (TRM-0522), allowing for future efforts to collect and distribute more material. In this article, we characterise the material using HRMS, UPLC-HRMS/MS and NMR.

2. Methods

2.1 Reagents

LC-MS grade methanol (Supelco LiChroSolv hypergrade for LC-MS) and fuming hydrochloric acid (37% HCl) were obtained from Merck (UK). Acetonitrile (ACN) was obtained from Sigma Aldrich (Supelco LiChroSolv hypergrade for LC-MS), formic acid (FA) was obtained from VWR (AmalaR Normapur, VWR Sweden), and 25% ammonia (NH₃) was obtained from Sigma Aldrich (Merck Life Science AB, Sweden, Supelco Suprapur grade). Hippuric acid, fuscidic acid, chloroform (CHCl₃) containing 100 - 200 ppm amylenes as stabiliser, deuterium oxide (D₂O), and 25 % ammonium hydroxide (NH₄OH) solution were obtained from Sigma Aldrich (Merck Life Science AB, Sweden). Suwannee River Fulvic Acid (SRFA; batch number 2S101F) and Pony Lake Fulvic Acid (PLFA; batch number 1R109F) were obtained from the International Humic Substances Society (IHSS, Saint-Paul, USA).
2.2 DOM Extraction

Seawater was obtained from the Tjärnö Marine Laboratory seawater system, which draws water from a 45 m deep intake located ~1.0 km off-shore of the nearest land mass on the west coast of Sweden at 58°52.843 N 11°06.378 E (Figure 1).

![Figure 1. Map showing location of Tjärnö Marine Laboratory (red pin) and seawater intake (yellow pin) on the west coast of Sweden. Images provided by Landsat / Copernicus via Google Earth, and © TerraMetrics (main image) and CNES / Airbus (lower insert).](image)

Water from this intake flows through a 10 mm coarse sieve before being pumped through a three-layer sand filtration system (6 - 8 mm; 3 - 4 mm; 1 - 2 mm) into a water reservoir. This reservoir feeds an extensive aquarium seawater system which includes a series of seawater taps (www.gu.se/en/tjarno/study-and-work/sea-water-systems). Water was drawn from this system on a daily basis between 16th and 20th May, 2022, during which time salinity (mean ± standard deviation = 33.3 ± 0.2 PSU; range = 32.9 - 33.7 PSU) and water temperature (mean ± standard deviation = 9.6 ± 0.5 °C; range = 9.0 - 10.4 °C) at the intake were relatively stable (Figure 2).
Figure 2. Five minute resolution salinity (blue) and water temperature (red) data from the Tjarnö Marine Laboratory deep water intake during the extraction period (16th - 20th May, 2022). Data provided by Gothenburg University ([www.weather.mi.gu.se/tjarno/data.shtml](http://www.weather.mi.gu.se/tjarno/data.shtml)).

A PuriFlash C18-AQ F1600 flash column (Interchim) was pre-conditioned with ~ 5 column volumes (12.5 L) of LCMS grade methanol (MeOH), starting at a rate of 20 mL min\(^{-1}\) and gradually increasing to 250 mL min\(^{-1}\) (4.5 bar pressure), followed by ~ 10 column volumes (25 L) MilliQ water acidified to pH 2.0 using ACS grade hydrochloric acid (HCl), at a rate of 400 mL min\(^{-1}\) (4 bar pressure). An operating pH of 2 was initially used based on standard protocols (e.g. Dittmar et al. 2008) but this appeared to strip a small amount of material from the column, and so the operating pH was raised to 3. A further 2.5 L acidified MilliQ and 2.5 L 100 % MeOH were passed through the column at a reduced flow rate to 250 mL min\(^{-1}\) (4.5 bar pressure), and then it was stored in 50 % MeOH for 4 days. Immediately prior to use, the column was flushed with 5 L of 50 % MeOH, 5 L of 20 % MeOH, and 5 L acidified MQ (pH 3) with the flow rate gradually increased from 20 mL min\(^{-1}\) to 180 mL min\(^{-1}\) (3 - 4 bar pressure).

The DOM extraction process is summarised in Figure 3. Each day, a fresh batch of seawater was collected from the tap into a clean 1000 L HDPE aquarium tank. From there, it was pumped sequentially through three pool filters (5 µm, 0.5 µm, and 0.5 µm pore size) into a second clean 1000 L HDPE tank, where it was acidified with hydrochloric acid (HCl) to a pH of ~3. This water was pumped through the column for ~ 22 hours each day. On the first day the flow rate was gradually increased from 180 mL min\(^{-1}\) (3.5 bar pressure) to 330 mL min\(^{-1}\) (4.5 bar pressure) over a period of 3 hours. A flow rate of between 320 and 360 mL min\(^{-1}\) and 4 bar pressure was maintained for the rest of the extraction. Accumulated salts were then removed by flushing the column with ~ 2 column volumes (4 L) MilliQ 2.5 L methanol (MeOH). The column was then flushed with a further 4 L acidified MilliQ water. During the elution process, a fresh batch of seawater was filtered and acidified, after which the process was repeated. Eluted MeOH was collected into a series of ashed (4 hours at 550°C) glass
Duran bottles and chilled overnight, along with a portion of acidified MilliQ which came through the column on either side of the MeOH (~ 3.5 L total each day). This was done to ensure no MeOH (and by extension, no retained DOM) was missed. A total of 1,500 L seawater was pumped through the column over five days.

Figure 3. Schematic diagram and method summary describing DOM extraction method, from collection of seawater from the tap to daily column elution.

2.3 Sample processing

Each eluted sample was dried down into a round-bottom flask using a rotary evaporator, and the dried material transferred into an 8 mL amber glass vial using small volumes of acetonitrile (ACN) and ultrapure water (MilliQ). A residue which could not be dissolved in ACN or ultrapure water was removed using 50 % MeOH, which was then removed using a rotary evaporator. We suspected that this residue was leached column material, and so each daily fraction was redissolved in 50 % MeOH and centrifuged for 15 mins at 6500 rpm. A white precipitate was pelleted out allowing the supernatant sample to be pipetted from the top. These samples were then freeze-dried and weighed. The total mass of these freeze-dried fractions was 1.8 g. Each fraction was redissolved in 50 % MeOH and combined to produce a single 115 mL sample.

Triplicate 10 µl sub-samples were dried down in a Speedvac drying centrifuge and re-diluted in 11 mL MilliQ water for measurement of dissolved organic carbon (DOC; see below for method). DOC concentrations in these samples were 4.74 ± 0.44 mg L⁻¹ (mean ± standard deviation), or 5297 ± 48.6 mg L⁻¹ in the concentrated sample. This was equivalent to just 34 % of the total sample mass, with DOC typically representing > 40 % DOM in aquatic samples.

A further processing step was applied to ensure complete removal of any residual column material which might be adding to the total sample mass. The mixture of material was dissolved in 500 mL of a 0.1% solution of 25% ammonium hydroxide in LCMS grade water.
This mixture was washed with 500 mL of CHCl₃ three times. This led to three separate liquid components: an aqueous phase, an organic phase, and an emulsion layer. The aqueous and organic phases were concentrated under reduced pressure using a rotary evaporator, then freeze dried. The emulsion layer (~ 100 mL) was mixed with 500 mL of a 0.1% solution of 25 % NH₄OH in LCMS grade water and stirred for 30 minutes to break down the majority of the emulsion. The mixture was then washed with 500 mL of CHCl₃ another three times. Again, this led to an aqueous phase, an organic phase, and an emulsion layer. These liquid components were concentrated under reduced pressure, before being freeze dried. Five separate dried components were obtained: two aqueous, two organic, and one from the emulsion. ¹H NMR and LC-MS analysis of these five fractions showed that the material presumed to be from the column had been transferred to the organic phases, while the vast majority of the DOM had remained in the aqueous phases. While both the column material and the DOM were observed in the emulsion fraction, this accounted for just ~ 2% of the total dried mass and was discarded, along with the organic portions. The two aqueous portions were re-dissolved in 50 mL of a 0.1% solution of 25% NH₄OH in LCMS grade water, combined, dried under reduced pressure, and freeze dried to produce a dried mass of 1058 mg. This was re-dissolved in 100 mL of 10 % MeOH in MilliQ water with 0.1 % NH₃, aliquoted for distribution, and dried down a final time using a Speedvac (1 mg and 10 mg aliquots) or freeze drier (larger aliquots). A 1 mg sub-sample was re-diluted in 41 mL MilliQ water, giving a DOM concentration of 24.7 mg L⁻¹. DOC concentration in this sample was 10.1 mg L⁻¹, equivalent to 41 % DOM, which is in line with expectations and, along with detailed characterisation data (see later), suggests that removal of column material was successful.

Aliquots were stored in a - 20 °C freezer in 1.5 mL Eppendorf tubes (1 mg and 10 mg aliquots), 8 mL amber glass vials (20 mg aliquots), and 40 mL clear glass vials (50 mg and 100 mg aliquots).

2.4 Analysis

2.4.1 DOC and absorbance

DOC was measured using a Sievers M9 TOC analyser and quantified as Non-Purgeable Organic Carbon (NPOC) against MilliQ blanks and a Potassium hydrogen phthalate (KPH) standard. Replicate measurements had a coefficient of variation of ≤ 2 %. The instrument was calibrated between 1 and 50 mg L⁻¹ using KPH, and subsequent KPH standards were all within 5 % of their calibrated value. Absorbance was measured between 200 and 400 nm at 2 nm intervals on a Shimadzu 1800 UV-Vis Spectrophotometer, using a 1 cm pathlength quartz cuvette. The instrument was baseline corrected us, and the sample was referenced against a MilliQ blank. DOC specific absorbance at 254 nm (SUVA₂₅₄), a commonly used metric for aromaticity where a higher number is associated with a greater aromatic content, was calculated by dividing absorbance at 254 nm (in cm⁻¹) by DOC concentration (in mg L⁻¹) and multiplying by 100¹².

2.4.2 HRMS

Dried samples were accurately weighed and dissolved/diluted to 50 mg DOM L⁻¹ with 50% methanol (LCMS grade, LiChroSolv, VWR) in ultrapure water (Milli-Q, Millipore). These
solutions were infused at 10 µL min\(^{-1}\) into a high-resolution mass spectrometer (Q-Exactive Orbitrap, Thermo Fisher) equipped with a heated electrospray ionisation source (HESI) operating in negative ion mode, at 100 °C with a setting of 25 units of sheath gas and 3 kV. The S-lens was set to 60 and capillary temperature was 250 degrees °C. The MS resolution was set to 140,000 at 200 Da, and transients were measured between m/z 150 - 1000, with 369.11911 and 525.19775 set as lock masses. 500 transients were collected and averaged, and a wash solution of 50% methanol:water (LCMS grade) was infused between samples until the blank signal stabilised.

Formulas were assigned allowing 4 - 50 C, 4 - 100 H, 0 - 2 N, 0 - 1 S and 0 - 1 \(^{13}\)C. Combinations of N, S and \(^{13}\)C were not allowed, nor were double bond equivalence minus O greater than 10. A 5\(^{th}\) order polynomial calibration was performed using a series of expected ions, as in a previous paper\(^{13}\), and formulas were assigned if the mass offset was less than 1.5 ppm.

### 2.4.4 UPLC-HRMS/MS

A 1 mg dried aliquot was dissolved in 200 µL 10% CMS grade LCMS grafe methanol in LCMS grade water, sonicated for 5 minutes, and vortexed to give a concentration of 5000 mg L\(^{-1}\). 10 µl sample containing 50 µg DOM was injected onto a UPLC C18 column (2 x 150 mm, 1.7 µm, Phenomenex Kinetix Core Shell) at a flow rate of 0.4 mL min\(^{-1}\), using a Vanquish UPLC (Thermo Fisher). Three mobile phases were prepared: LCMS grade water with 0.1 % formic acid (A), LCMS grade acetonitrile with 0.1 % formic acid (B), and 90:10 LCMS grade MeOH in LCMS grade water with 0.05 % formic acid and 0.25 % NH\(_3\). The gradient was run as follows: 0 - 1 min = 1% B, 0 % C; 1 - 9 min = 1 - 99 % B, 0 % C; 9 - 10 min = 99 % B, 0 % C; 10 - 10.1 min = 99 - 1% B, 0 - 99 % C; 10.1 - 11.4 min = 1 % B, 99 % C; 11.4 - 11.5 min = 1% B, 99 - 0 % C; 11.5 - 15 min = 1% B, 0 % C. An Orbitrap Q Exactive mass spectrometer was used, operating in negative mode for one analysis (triplicates plus blank) and positive mode for another (triplicates plus blank). Data were collected in data dependent mode, with the top 5 peaks sent for HCD fragmentation, and are available at the MassIVE data repository (https://massive.ucsd.edu; MSV000090282). MZmine3 was used to align MS1 level LC-MS features\(^{14}\) and find chromatographically resolved peaks\(^{6}\) (MZmine3 project available in supplementary information), and Global Natural Product Social Molecular Networking (GNPS) jobs were run on positive and negative mode fragmentation data\(^{15}\) to cluster and library match fragmentation patterns at the MS2 level\(^{16}\).

### 2.4.5 NMR

NMR spectra were acquired at 298 K on a Bruker 500 MHz spectrometer using a TXO CRPHe TR-\(^{13}\)C/\(^{15}\)N/\(^{1}\)H 5 mm-Z CryoProbe. The samples were dissolved in D\(_2\)O, and referenced to the residual solvent peak at 4.87 ppm in the \(^{1}\)H NMR spectrum, or to MeOH at 49.86 ppm which was added as an external reference for the \(^{13}\)C NMR spectrum. \(^{1}\)H NMR spectra were gathered over 330 scans, with a 3.28 s acquisition time, and a 15.3 s relaxation delay. \(^{13}\)C NMR spectra were gathered over 9600 scans, with a 0.92 s acquisition time, and an 8 s relaxation delay.

### 3. Results and Discussion
3.1 [DOM], [DOC] and absorbance

The final dry DOM mass was 1.06 g, equivalent to 0.71 g DOM L$^{-1}$ seawater extracted. DOC accounted for 41% of the extracted DOM, equivalent to 0.29 g DOC L$^{-1}$ seawater extracted (24.2 µM SPE-DOC). DOC typically comprises ~ 50% DOM in humics which are likely to be highly phenolic and aromatic in character, and ~ 45% in freshwaters, where humics are mixed with the rest of the aquatic DOM pool. The aged, recalcitrant DOM typically found in seawater is thought to be dominated by alicyclic and linear compounds which are rich in carboxylic acids, and the increased abundance of oxygen from carboxyl groups likely decreases the percentage of carbon in marine DOM relative to the terrestrial systems.

For a DOC concentration of 10.1 mg L$^{-1}$, absorbance at 254 nm was 0.137 cm$^{-1}$, giving a SUVA$_{254}$ of 1.36 L mg C$^{-1}$ m$^{-1}$. SUVA$_{254}$ typically ranges from 1.0 to 6.0 L mg C$^{-1}$ m$^{-1}$ in surface waters, although values > 6.0 L mg C$^{-1}$ m$^{-1}$ have been reported for samples with a strong terrestrial signal and values < 1 L mg C$^{-1}$ m$^{-1}$ have been reported for samples dominated by fresh DOM production (e.g. algal leachate samples). Coastal ocean samples typically exhibit values of around 1.5 - 3.5 L mg C$^{-1}$ m$^{-1}$, whilst open ocean samples tend to be closer to 1.0 L mg C$^{-1}$ m$^{-1}$. Thus, the SUVA$_{254}$ value of TRM-0522 (1.36 L mg C$^{-1}$ m$^{-1}$) is within the expected range and indicates a considerably lower degree of aromaticity relative to other available reference materials e.g. Suwannee River Natural Organic Matter (SRNOM; 5.34 L mg C$^{-1}$ m$^{-1}$), SRFA (4.20 mg C$^{-1}$ m$^{-1}$), and PLFA (2.0 - 3.2 L mg C$^{-1}$ m$^{-1}$).

3.2 HRMS

High resolution mass spectrometry analysis yielded > 3000 peaks (Table 1). In terms of distribution between CHO, CHON and CHOS containing formulas, the sample was somewhere between the IHSS reference materials SRFA and PLFA. TRM-0522 contained less N and S containing peaks than PLFA (which are actually the most numerous in that sample), but was much more N and S rich than SRFA (Figure 4). However, it should be noted that multiple N and S containing peaks, as well as CHONS peaks, could not be confidently assigned following Orbitrap analysis, and so the true diversity of the TRM-0522 sample could not be completely assessed.

The distribution of formulas in van Krevelen space was typical for a marine sample, with the majority of formulas in the region H/C 1 - 1.5 and O/C 0.3 - 0.7 (Figure 5). This is where the most recalcitrant DOM is typically found, namely 'Island of Stability' molecules and CRAM formulas. These data are consistent with the NMR findings, and suggest that the TRM sample is a useful and otherwise unavailable reference mixture for relatively aged, recalcitrant marine DOM.

Table 1. High resolution mass spectrometry peak metrics for TRM-0522, SRFA, and PLFA showing number of peaks (Peaks) and the intensity weighted average of oxygen to carbon (O/C$_{wa}$), hydrogen to carbon (H/C$_{wa}$), mass to charge (m/z$_{wa}$) ratios, and the number of common peaks detected relative to a recent inter-laboratory study.
<table>
<thead>
<tr>
<th>Peaks</th>
<th>O/C&lt;sub&gt;wa&lt;/sub&gt;</th>
<th>H/C&lt;sub&gt;wa&lt;/sub&gt;</th>
<th>m/z&lt;sub&gt;wa&lt;/sub&gt;</th>
<th>Interlab common peaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRM-0522</td>
<td>3012</td>
<td>0.46</td>
<td>1.28</td>
<td>385.8</td>
</tr>
<tr>
<td>SRFA (2S101F)</td>
<td>2323</td>
<td>0.50</td>
<td>1.10</td>
<td>359.3</td>
</tr>
<tr>
<td>PLFA (1R109F)</td>
<td>4158</td>
<td>0.43</td>
<td>1.33</td>
<td>322.7</td>
</tr>
</tbody>
</table>

Figure 4: Reconstructed mass spectra (top) and van Krevelen diagrams (bottom) of formulas assigned in the TRM-0522 sample, along with two other reference mixtures from the IHSS - Suwannee River Fulvic Acid (SRFA) and Pony Lake Fulvic Acid (PLFA). CHO-containing formulas without other heteroatoms are shown in black, those containing nitrogen (CHON) are shown in red, and those containing sulfur (CHOS) are shown in blue. The total number of each is indicated in the associated van Krevelen diagram.

### 3.3 UPLC-HRMS/MS

UPLC-HRMS/MS analysis allows a greater level of molecular detail to be inspected by separating the geochemical DOM background across a wide elution period, albeit without finely detailed chromatographic resolution<sup>6,8</sup>, and enabling determination and annotation of truly resolvable metabolite features<sup>9,16</sup>. In Table 2, the number of features found in TRM-0522 in positive and negative mode UPLC-HRMS and UPLC-HRMS/MS are presented. A variety of resolvable features were found over a wide range of retention times and masses, numbering 233 (negative) and 261 (positive). Unlike the typical broadly eluting features
found in the geochemically degraded DOM background, these features are well resolved and can be used as reference peaks for method development, validation, and drift correction if TRM-0522 is included in metabolomics datasets. The number of library hits found using GNPS molecular networking was surprisingly low (only 5 per ionisation mode, and most of these known laboratory contaminants). This may be due to the fact that so much of DOM is not characterised and annotated\textsuperscript{31}, and may also be due to the extraction method used here being poor at retaining particularly hydrophilic compounds. Possibly, the pipeline and gravel filtration method used to collect coastal water for the aquarium activities at the research site also filters out many known metabolites. However, hundreds of resolvable features were found at the MS1 and MS2 level (Figure 5), they simply were not matched to known compounds.

Table 2. Resolved peak metrics from LC-MS/MS

<table>
<thead>
<tr>
<th>Mode</th>
<th>MS1 features (with MS2 trigger)\textsuperscript{*}</th>
<th>MS2 clusters\textsuperscript{\textdagger}</th>
<th>Library matches\textsuperscript{\textdagger}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>261 (200)</td>
<td>130</td>
<td>5</td>
</tr>
<tr>
<td>Negative</td>
<td>233 (163)</td>
<td>299</td>
<td>5</td>
</tr>
</tbody>
</table>

\textsuperscript{*}From MZmine3, \textsuperscript{\textdagger}From GNPS, see data sharing.

Figure 5: UPLC-HRMS features determined by MZmine3, shown as m/z vs. retention time, and with relative intensity in triplicate analysis shown as point size. Only points with an average more than 10 x blank are shown, and their number is indicated in text at the top left.

3.4 NMR

Examining the $^1$H NMR spectrum, 6 distinct regions can be identified (Figure 6), with ranges for these regions defined by previous work\textsuperscript{32}. Region A spans from 0.6 to 1.3 ppm, and consists of aliphatic polymethylene and methyl functionalities, while region B, spanning 1.3 - 2.9 ppm, includes $N$- and $O$- substituted aliphatic signals. Further downfield, region C
includes predominantly O-alkyl signals, and covers 2.9 - 4.1 ppm. Region D is composed predominantly of the α-proton of peptides, comprising signals from 4.1 - 4.8 ppm. The presence of the residual water peak at 4.79 ppm obscures any accurate integration of the anomeric protons of carbohydrates. Aromatic and phenolic hydrogen atoms can be observed between 6.2 and 7.8 ppm in region E. Finally, amide protons can typically be observed in region F, from 7.8 - 8.4 ppm. However, exchange with deuterium from the solvent means this signal is non-quantitative. The peak observed at 3.34 ppm is attributable to residual methanol. Finally, no peak at 6.57 ppm, typically attributed to lignin, is observed. The relative amounts of each accurately definable area (regions A - E) are presented in Table 3.

Figure 6: $^1$H NMR spectrum of TRM-0522 at 500 MHz in D$_2$O. Coloured boxes A - F indicate defined regions corresponding to different groups.

Table 3: Functionalities related to, and percentage contribution of integrals for each region in the $^1$H NMR spectrum of TRM-0522.

<table>
<thead>
<tr>
<th>Region</th>
<th>ppm range</th>
<th>Functionalities</th>
<th>% Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.6-1.3</td>
<td>Polymethylene, methyl</td>
<td>18</td>
</tr>
<tr>
<td>B</td>
<td>1.3-2.9</td>
<td>N- and O- substituted aliphatics</td>
<td>45</td>
</tr>
<tr>
<td>C</td>
<td>2.9-4.1</td>
<td>O- alkyl</td>
<td>27</td>
</tr>
<tr>
<td>D</td>
<td>4.1-4.8</td>
<td>Peptide α-proton</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>6.2-7.8</td>
<td>Aromatic and phenolic</td>
<td>1</td>
</tr>
</tbody>
</table>
The $^1$H NMR spectra of TRM-0522 retains the key features that have previously been described for marine and freshwater DOM$^{19,33,34}$. Specifically, regions have been attributed to carbohydrates, CRAM, and aliphatic compounds$^{19}$. The aliphatic region has been described as including both peptide derived aliphatics, but also material derived from linear terpenoids (MDLT)$^{35}$. Roughly, the ranges displayed on Figure 6 and Table 3 correlate to these constituent molecules, with range A including aliphatics, range B including CRAM, and range C including carbohydrates. While the percentage contributions obtained by region integration can only approximately estimate the actual compound composition, the integrated ranges are in alignment with previously reported data that show CRAM as the most abundant component, followed by carbohydrates, and finally aliphatics$^{33}$. Of note is a pronounced feature at 3.52 - 3.76 ppm, which in the majority of previously reported samples is not observed as a distinct peak$^{19,33,34}$. However, this peak is seen in spectra of samples isolated at several depths in the Atlantic ocean$^{19}$, and is possibly derived from methoxy-ether functionalities$^4$. Consistent with other marine water samples, TRM-0522 lacks strong signals attributable to alkenes or aromatics, indicating a low contribution from lignin, humic, and fulvic substances. This also corresponds well to our finding of very low SUVA$_{254}$ (1.36 L$^{-1}$ g cm$^{-1}$).

In the $^{13}$C NMR spectrum, seven separate regions are presented, as according to definitions used in previous reports$^{36–40}$. Region A covers 0 to 45 ppm, and contains signals corresponding to alkyl chains that are not functionalized with heteroatoms. From 45 - 65 ppm, region B presents substituted alkyl carbons attached predominantly to nitrogen atoms, but also some oxygen atoms (e.g. methoxy carbons), as well as some highly branched aliphatic signals that neighbour, but are not directly attached to, heteroatoms. Region C, from 65 - 95 ppm, contains predominantly signals attributable to oxygen substituted carbon atoms, such as ethers and ring carbons in carbohydrate molecules. Region D, from 95 - 110 ppm contains di-oxygenated carbons, such as those from the anomeric position of carbohydrates. Next, region E consists of aromatic carbon atom signals, and covers the range of 110 - 145 ppm, while region F consists of phenolic carbon atom signals, and spans 145-160 ppm. Finally, signals for carboxylic functionalities, including acids, amides, esters, and ketones, are found in region G, covering 160 - 220 ppm. A peak corresponding to residual methanol can be observed at 49.86 ppm.
Figure 7: $^{13}$C NMR spectrum of TRM-0522 at 125 MHz in D$_2$O. Coloured boxes A-G indicate defined regions corresponding to different groups.

Generally, information extrapolated from the $^{13}$C NMR spectrum of TRM-0522 agrees with the analogous $^1$H NMR spectrum. Furthermore, the general regions and rough peak shapes correspond to previously reported spectra$^{4,19,35,41}$. While there are observable peaks in the aromatic/olefinic region, they are relatively minor when compared to those derived from saturated carbon functionalities. This is consistent with the relatively low content of lignin, humic, and fulvic substances in oceanic samples$^{4,19,35,41}$. A carbonyl signal is observed in the region typically associated with carboxylic acid and amide functionalities (~ 190 - 170 ppm), while no detectable signal is associated with other carbonyl functionalities (ketones, aldehydes, carbamates, carbonates). In agreement with the $^1$H NMR spectrum analysis, a strong band corresponding to the alcoholic carbons in sugar compounds (95 - 75 ppm) is observed. While the anomeric signals are obscured by the residual water peak in the $^1$H NMR spectrum, a peak attributable to the anomeric carbons of carbohydrates is observed in the $^{13}$C NMR spectrum (~ 105 - 95 ppm). The remaining two regions contain strong aliphatic signals, with the signal reducing in intensity outside the 55 - 15 ppm range. This is consistent with the presence of cyclic and linear terpenoid derived material, with cyclic material tending towards the higher chemical shift range, and linear material tending towards the lower chemical shift range$^{19,35}$.

A useful point of comparison for TRM-0522 is the common reference material SRFA. An integrated $^1$H NMR spectrum of SRFA was obtained (Figure 8), run using the same parameters and instrument as was used for TRM-0522. Regions were defined in the same way as for TRM-0522, (see earlier), but a broadening of the residual water peak (4.79 ppm), led to overlap with region D (typically containing information regarding the α-proton of peptides), thus preventing accurate integration of this area. To account for this in comparison to TRM-0522, adjusted contributions of TRM-0522 are shown that exclude
region D, alongside the relative contributions of SRFA (Table 4). Two key pieces of information can be observed from this comparison. Firstly, there is significantly more material that contains aromatic hydrogens in SRFA than in TRM-0522 (10 vs 1 % respectively). This is typically associated with humic and fulvic substances, as well as lignin. Of course, in SRFA this should predominantly come from fulvic acid. However, the presence of lignin cannot be ruled out as the aromatic window is broad and covers the signature peak for lignin at 6.57 ppm. The second important piece of information is that the ratios of the A, B and C regions remain almost identical between spectra of SRFA and TRM-0522 samples, suggesting that the relative amounts of MDLT, CRAM, and complex polysaccharides remains consistent between the two samples. For $^{13}$C NMR comparison, data was used from the U.S. geological survey. The most obvious difference between the $^{13}$C NMR spectra of the materials is a significant aromatic peak spanning 110-145 ppm in the SRFA spectra, a region in which only relatively minor resonances can be observed in the TRM-0522 spectra. This reinforces the lack of the relatively photolabile humic, fulvic, and lignin substances in TRM-0522, and suggests the material is unlikely to be photoreactive. While other regions in the $^{13}$C NMR spectra of the two samples are relatively similar, differences are noted with the increased peak height in regions attributable to sugars in TRM-0522 (when compared to aliphatic signals), and an increased peak height indicative of substances attributable to carbonyl functionalities in SRFA (when compared to aliphatic signals).

**Figure 8:** $^1$H NMR spectrum of SRFA at 500 MHz in D$_2$O. Coloured boxes A - F indicate defined regions corresponding to different groups.
Table 4: Functionalities related to, and percentage contribution of integrals for each region in the $^1$H NMR spectrum of SRFA.

<table>
<thead>
<tr>
<th>Region</th>
<th>ppm range</th>
<th>Functionalities</th>
<th>% Contribution</th>
<th>Adjusted TRM-0522 Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.6-1.3</td>
<td>Polymethylene, methyl</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>B</td>
<td>1.3-2.9</td>
<td>N- and O- substituted aliphatics</td>
<td>46</td>
<td>49</td>
</tr>
<tr>
<td>C</td>
<td>2.9-4.1</td>
<td>O- alkyl</td>
<td>28</td>
<td>31</td>
</tr>
<tr>
<td>E</td>
<td>6.2-7.8</td>
<td>Aromatic and phenolic</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

5. Conclusion

The results presented in this paper demonstrate that TRM-0522 is a useful and reasonably representative reference material for use in marine DOM biogeochemistry. We note that use of the existing marine DOC CRM\textsuperscript{11} as a DOM standard is technically feasible, but consider this to be an inefficient option: assuming a typical DOC concentration of 0.5 mg L\textsuperscript{-1}, a C content of 50 %, and an extraction efficiency of 50 %, a 1 mg DOM extraction would require 100 x 20 mL vials at a cost of $140 USD. TRM-022 is currently available at a cost of approximately $25 USD per 1 mg, and has been well characterised using several common high-resolution methods (NMR; HRMS: UPLC-HRMS/MS). Thus, TRM-022 fills a resource gap for a DOM-specific marine reference material.

This resource is now available to the research community, and we encourage its use in marine biogeochemical studies including DOM characterization, natural products research, and metabolomics. During this first effort, we have made 700 mg marine DOM reference material available, with future, larger scale extractions planned. 50 mg of the current batch is available free of charge for researchers with financial barriers to purchase. A fixed extraction point associated with an established university research laboratory (Tjärnö Marine Laboratory, University of Gothenburg) was selected to ensure ongoing access, consistency of source, and a long-term supply. Thus, we intend to make TRM widely available and encourage its day-to-day analytical use. Enquiries can be sent to the corresponding author (JAH).

Acknowledgements and data sharing

The authors wish to thank Tze Yung (Patrick) Wong, Teresa Catalá, and Alex Villarreal for assistance during field work, and Ahmed Alrifaiy, Gunilla Johansson, and the research and support staff at Tjärnö Research Laboratory. This work was funded by FORMAS (JH Grant number 2021-00543). SLF received funding from the Carl Tryggers Foundation Postdoctoral Fellowship (CTS20:170). The authors report no conflicts of interest.
MZmine projects and direct infusion mass spectrometry datasets and data processing code are available as supplementary files. GNPS MS2 level clustering and library matching are available at the following links.

https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=84f7a8209cd6444f85027aaf8357ad4&view=view_all_clusters_withID_beta%7B%22LibraryID_asc%22%7D

https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=e2f9d825b55e49cf913e3fe3ed5a8&view=view_all_clusters_withID_beta#%7B%22table_sort_history%22%3A%22%22%22%22%22%22%7D

References


