Online counter gradient LC-FT-ICR-MS enables detection of highly polar natural organic matter fractions

Limei Han[1], Jan M. Kaesler[1], Chang Peng[2], Thorsten Reemtsma[1,2], and Oliver J. Lechtenfeld*[1,3]

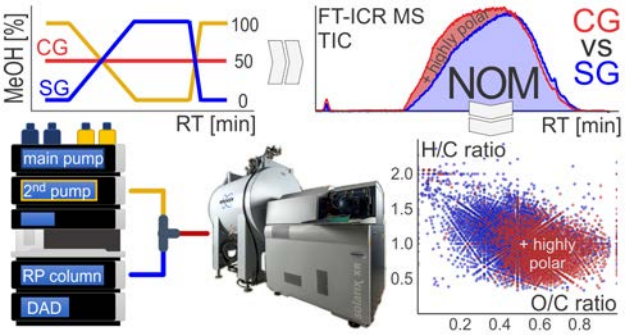
1 Department of Analytical Chemistry, Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany

2 Institute of Analytical Chemistry, University of Leipzig, 04103 Leipzig, Germany

3 ProVIS – Centre for Chemical Microscopy, Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany

* Corresponding Author: oliver.lechtenfeld@ufz.de.

ABSTRACT: Natural organic matter (NOM) is a highly complex mixture of natural organic molecules. The recent developments in NOM molecular characterization methods have shown that ESI-FT-ICR hyphenated with liquid chromatography (LC) is a promising approach to also obtain chemical information (such as polarity and molecular size) about NOM molecules. However, due to changing solvent composition during gradient elution in LC-FT-ICR-MS, ionization conditions also change throughout the chromatographic separation process. In this study, we applied a post-LC column counter gradient (CG) to ensure stable solvent conditions for transient ESI-MS signals. Suwanee River Fulvic Acid (SRFA) standard and a peat pore water were used as representative dissolved NOM samples for method development and validation. Our results show that in polar NOM fractions (which elute with < 50% methanol) the TIC intensity and number of assigned molecular formulas were increased by 48% and 20%, as compared to the standard gradient (SG) method. Further application of a Q-isolation and selective ion accumulation for low abundance fractions revealed over 3 times more molecular formulas (especially for CHNO, CHOS, CHNOS formula classes) than in full scan mode. The number of detected highly polar NOM compounds (with elemental ratios H/C < 1, O/C > 0.6) were more than 20 times larger for CG-LC mode as compared to direct infusion (DI) (5715 vs 266 MF). We conclude that the application of a post-column counter gradient in LC-FT-ICR-MS analyses of NOM offers novel insight into the most polar fractions of NOM which are inaccessible in conventional DI measurements.



INTRODUCTION

Natural organic matter (NOM) is a complex mixture of biologically derived organic molecules, which exists ubiquitously in the aquatic and terrestrial environment and plays a crucial role in carbon cycle [1,2]. Extensive turnover of biomolecules via biotic and abiotic processing constantly alters the molecular composition of NOM, resulting in an vast structural diversity [3]. The metabolism of plants and microorganisms produces polar metabolites and biopolymers which are comprised of highly polar subunits, often contain heteroatoms, and are released during degradation processes [3,4]. Primary degradation pathways of biopolymers typically include their transformation into small and polar compounds [5]. The polar fraction of NOM is water soluble, mobile in aquatic systems, and forms the dissolved organic matter (DOM) pool. In addition, polar DOM compounds can interact with anthropogenic chemicals in the environment or during water treatment processes [6,7]. Thereby formed molecules may be resistant to microbial degradation or contribute to toxicity potential [8,9]. While much is known about these processes on a bulk level, the molecular diversity and structural complexity of DOM typically hamper the quantitative assessment of molecular-level turnover in the environment or technical processes [10-12].

Ultrahigh resolution Fourier-transform ion cyclotron resonance (FT-ICR) or Orbitrap mass spectrometry (MS) reveal tens of thousands of peaks in DOM samples and are nowadays commonly used for the molecular characterization of DOM [2,13]. State-of-the-art method for mass spectrometric DOM analysis is the direct infusion electrospray ionization (DI-ESI) where the sample is continuously infused into the ion source of the mass spectrometer. This enables long acquisition times via spectra co-addition to reach highest sensitivity for low abundance ion species. Albeit its ease of use and high sensitivity, DI-ESI suffers from ionization artifacts such as charge competition and suppression [14,15]. As a consequence, DOM compounds with low abundances and/or low ionization efficiency like highly polar metabolites with rapid microbiological turnover, are hardly detected with DI measurements [14] . Another major drawback of DI-ESI MS for the advance of DOM process understanding in the environment is the lack of chemical information, i.e. via size or polarity separation, making it also impossible to pre-isolate individual isomers for structure elucidating MS/MS or IMS approaches [16,17].

Polarity or size based liquid-chromatographic separation methods for DOM have been previously applied [18-20], but are often lacking separation power for DOM molecules [21]. Only the combination with high resolution mass spectrometry (HRMS) yields molecular-level based information of DOM, such as capillary electrophoresis or reversed phase liquid chromatography (RPLC) coupled with FT-ICR [19,22], size exclusion chromatography (SEC) coupled with Orbitrap [23,24], or FT-ICR [25], and 2D-LC coupled with FT-ICR [26]. The coupling of LC to HRMS generally improves sensitivity to detect compounds hidden in DI measurements. Most recent applications involve RPLC and SEC online coupled with Orbitrap [27-29] and FT-ICR [30].

Compared with offline methods, online method avoids fraction collection and additional injection steps, which makes it a time-saving, low contamination risk, and automated method.

However, the LC gradient determines the solvent composition at the ESI source, which is a crucial parameter affecting ionization efficiency [31]. In DI-ESI, varying solvent compositions, especially very low and high water content result in major changes of molecular weight distribution [32] and reduced measurement reproducibility [33]. In contrast to DI-ESI, matrix effects caused by salts and simultaneously eluting DOM also vary in LC-MS applications, differently affecting ionization of compounds with the same m/z value eluting at different retention times. Thus, the apparent DOM m/z peak intensity in LC-MS is influenced by a) solvent composition, b) charge competition with other solution constituents and c) the inherent ionization efficiencies of individual isomers eluting at different retention times in addition to the actual concentration of individual isomers.

Furthermore, transient signals are challenging for Fourier-transform based mass spectrometric detectors due to ion-cloud interactions in the analyzer cell [34]. Hence optimizing MS methods for spectral quality (i.e. at the maximum TIC) results in limited sensitivity for low abundance ion species at retention times when the overall elution is low or when many high abundance compounds co-elute. Alternatively, ion pre-selection (e.g. in an isolation cell) prior detection, which is also applied in DI measurements to achieve a higher dynamic range and hence sensitivity, offers the potential to selectively improve m/z ranges with lower ion populations [34-36]. This may prove advantageous for the detection of hetero-atom containing compounds in NOM, which have generally lower abundances in ESI-MS, and in particular in terrestrial samples with high C:N ratio.

In this study, we developed a novel online reversed-phase LC-FT-ICR MS method for the chemical separation and characterization of DOM. To enhance the ionization efficiency of water soluble polar compounds, a pure water elution step and a post-column counter gradient (CG) to stabilize the solvent ratio before the sample entered the ESI source of the FT-ICR-MS was applied. In addition, Q-isolation and mass spectral stitching mode was used to enhance the detection of low intensity highly polar species, which were not detectable in full scan mode.

MATERIAL AND METHODS

Chemicals and samples

Analytical grade chemicals (D-Glucuronic acid, Fraxin, Isoferulic Acid, 3-O-ß-D-Glucuronide and 2-(4-(2,2-dicarboxy-ethyl)-2,5-dimethoxy-benzyl)-malonic acid) were selected as model compounds for LC column performance and reproducibility control [27]. Detailed information about the model compounds can be found in the supplementary information (SI) to this article (Table S1).

Suwannee River Fulvic Acid (SRFA) was purchased from the International Humic Substances Society (SRFA II; 2S101H). A peat pore water (PPW) sample was collected from the Neustädter Moor (52.594527 °N 8.672156 °E, Lower Saxony, Germany). Detailed information can be found in SI.

Reversed phase liquid chromatography (RP-LC)

Chromatographic separation of DOM was performed on a UHPLC system (UltiMate 3000RS, Thermo Fischer Scientific, Waltham, MA, U.S.A.) equipped with binary pump (HPG-3200RS), column oven (TCC-3000RS), auto sampler (WPS-3000TRS), and diode array detector (DAD-3000RS).

A reversed phase polar end-capped C18 column (ACQUITY HSS T3, 1.8 µm, 100 Å, 150x3mm, Waters, USA) equipped with guard column (ACQUITY UPLC HSS T3 VanGuard Pre-column, 100Å, 1.8 µm, 2.1 mm X 5 mm) was used to separate the DOM. This column showed superior separation of DOM over other RP columns types, especially for the most polar compounds (Figure S1).

Ultrapure water (MQW; Merck, Darmstadt, Germany) and LC-MS-grade methanol (MeOH; Biosolve, Valkenswaard, Netherlands) were used as mobile phases A and B. 0.05% formic acid (FA) was added to both eluents. The pH in the aqueous eluent was adjusted to 3.00±0.02 with ammonium hydroxide (NH_4OH) and the same volume of NH_4OH was added to eluent B. Fresh eluents were prepared for each sequence.

The eluent gradient started at 0.5 min and linearly increased to 100 % MeOH within 14 min (Table S2 shows details of the gradient program). The applied gradient resulted in i) an isocratic separation step with pH3 water [22,37,38] and ii) a continuous elution of DOM reflecting its continuous nature without artificially induced separations from step or spiked gradients.

A second HPLC pump (LPG-3400SD) was used to add a post column flow (0.2 ml/min). A gradient was applied to reverse the gradient of the first pump at the column outlet using the same solvents (A: MeOH, B: MQW) but without buffer addition (i.e. no pH adjustment, cf. Table S3). This operation mode is referred to as counter gradient (CG). The combined flow (50% MQW, 50% MeOH) was split again for the mass spectrometer (0.1 ml/min) and DAD (0.3 ml/min) using an adjustable flow splitter (ERC, Germany). The DAD (210 – 280 nm) was used as second detector to monitor the performance of the HPLC system and to provide additional information to the compounds which absorbs light but unable to be ionized with negative mode [39]. In the discussion, only UV trace at 254 nm (UV_{254}) is shown.

As control experiments, the second pump was operated to match the gradient of the first pump (addition of buffer free solvents) ensuring same flow rates and concentration of DOM and buffer for MS measurements. This LC mode is termed standard gradient (SG) in the following.

FT-ICR-MS

An FT-ICR mass spectrometer equipped with a dynamically harmonized analyzer cell (solariX XR, Bruker Daltonics Inc., Billerica, MA, USA) and a 12 T refrigerated actively shielded superconducting magnet (Bruker Biospin, Wissembourg, France) was coupled to the LC system. An electrospray ionization source (Apollo II) was used in negative mode (capillary voltage: 4.3 kV, nebulizer gas pressure: 1.0 bar, dry gas temperature: 250 °C, dry gas flow rate: 8.0 L/min).

Mass spectra for LC-MS measurements were acquired in broadband mode (147.41 to 1000 m/z) with a transient size of 2M (~0.84s FID) and full profile mode. Lock-masses (from 150 m/z to 600 m/z) were used to compensate mass shifts during LC acquisition. The resolving power ($m/\Delta m_{50\%}$) at m/z 400 was approx. 265,000 which is sufficient to resolve all major DOM species in the considered mass range [22,30]. The ion accumulation time (IAT) was set between 45 ms for SRFA and 40 ms for PPW corresponding to a maximum TIC intensity for single scans between $1 \cdot 10^9$ and $1.5 \cdot 10^9$. These settings resulted in a scan rate of 0.87 s and mass spectral parameters similar to a 7 T ICR instrument equipped with a quadrupolar detection cell [30]. A total of 834 scans were recorded between 5 and 21 min retention time for each LC-MS measurement.

Next to full mass range acquisition, Q-isolation for selected mass ranges (CASI, continuous accumulation of selected ions) was used to detect low-abundant ions. Q-isolation with subsequent mass window stitching is suitable for LC-MS runs to accumulate ions of selected mass spectral windows in the hexapole concomitant to ion detection in the cell [40]. Ten mass windows with window sizes of 50 to 100 Da were used along the mass range of DOM peaks (150-1000 mz) by multiple injection on the LC system and sequential acquisition of mass windows by the FT-ICR MS. A higher IAT of 400 ms was used to increase the number of ions in the cell as compared to the full scan mode. Detailed information about the CASI setup can be found in the SI.

Data analysis

LC-MS data were split and averaged into 1 min wide segments between 5 and 21 min. Peak picking S/N threshold was set to 4 for both DI and LC-MS runs. Each resulting mass spectrum was internally recalibrated with a list of masses commonly found in DOM (150-1000 m/z) and the mass accuracy after linear calibration was better than 0.184 ppm (n = 251) (Figure S2). Mass spectral averaging and calibration of segments was done with DataAnalysis 5.0 (Bruker).

Molecular formulas (MF) were assigned to mass peaks in the range 150-1000 m/z allowing for a maximum mass error of ±0.5 and the following elemental ranges: $C_{1\text{-}60}$, $^{13}C_{0\text{-}1}$, $H_{1\text{-}122}$, $O_{0\text{-}40}$, $N_{0\text{-}2}$, $S_{0\text{-}1}$, $^{34}S_{0\text{-}1}$. Additional filters were applied to the calculated molecular formulas: $0.3 < H/C < 2.5$, $0 < O/C < 1$, $0 < N/C < 1.5$, $0 < DBE < 25$ (double bond equivalent, $DBE = 1 + 1/2\,(2C - H + N)$ [41]), $-10 < DBE\text{-}O < 10$ [42], and element probability rules proposed by Kind and Fiehn [43]. Isotopologue formulas (^{13}C, ^{34}S) were used

for quality control but removed from the final data set as they represent duplicate chemical information. Molecular formulas assigned to peaks in LC-MS blanks were removed from the respective segments. For the applied element ranges and S/N threshold, this procedure resulted in only 4 formulas removed per segment on average.

RESULTS AND DISCUSSION

Chromatographic performance and chromatogram segmentation

To assess the chromatographic performance, five model compounds of a wide polarity range (logD: -2.82 to 1.14) were used. Although glutamic acid eluted close to the dead volume (t_M = 4.3 min), all model compounds were separated effectively as indicated by their extracted ion chromatogram (EIC, Figure S3). The chromatographic peak width of the model compounds ranged between 0.5 to 1 min, with stable retention times over multiple injections (Table S1). The near base line separation of the three isomeric model compounds with molecular weight 370.27 Da confirms that the applied RP-LC method is capable to efficiently retain and separate isomers with varying logD values (-1.25 – 0.01 at pH3). Adding the model compounds (0.02 - 2.7 ppm C) into a high concentration of SRFA (210 ppm C) did not shift the retention times of model compounds, confirming that the SRFA contained only a low proportion of highly non-polar compounds which are strongly retained by the RP-LC column.

On the used RP column, SFRA was separated into two distinct peaks based on the UV_{254} and mass chromatograms (Figure S3). A comparably small peak eluted at 5 min shortly after the dead volume, and a large, unstructured peak eluted from approx. 9.8 to 21 min. Due to the dimension of the column and low flow rate, the first MeOH from the solvent gradient reached the detector at 9.8 min. During the resulting isocratic separation range (pure water adjusted to pH3) between 4.3 min and 9.8 min the most polar compounds elute as described previously (Figure S4) [22,37,38]. Starting with the gradient separation range, the TIC and UV_{254} absorbance increased up to 40% MeOH (16 min), followed by a smooth UV_{254} absorbance and mass peak intensity drop until 75% MeOH (20min), confirming that the water soluble fraction of SRFA contained a low proportion of highly non-polar compounds which cannot be eluted with MeOH.

Due to the high structural complexity in DOM, individual isomers – from potentially tens of thousands for one molecular mass – largely overlap in their retention time [44]. This leads to the broad distribution of most EICs in any DOM sample [27,28,45]. The elution profile of individual m/z ratios in SRFA resulted in an inability to properly pick chromatographic peaks using standard software [46]. Accordingly, the retention time range was divided into 16 evenly spaced segments between 5 and 21 min, each 1 min segment consisted of 52 individual MS scans, which is sufficient to accurately model chromatographic peaks [27]. Although the peak width of the model compound (for a concentration range of 40 – 4700 ng/mL) was estimated to be 0.5 – 1 min in our system, we can readily assume that the (currently invisible) chromatographic peak width

of individual SRFA isomers is much smaller due to their order of magnitude lower concentrations [32]. Accordingly each segment may represent multiple isomeric SRFA compounds for each detected mass peak, while any individual isomer in SRFA may extend over two consecutive segments at max [26,44]. Hence each segment is considered to represent an own set of (almost exclusive) isomers and we conservatively count a detected mass peak in one segment as an individual isomer.

Between 6 and 10 minutes, no elution of compounds was detected in SRFA under standard MS conditions. These 4 segments were omitted from further analysis resulting in 12 considered segments for SRFA (Figure S3). However, non-extracted PPW sample showed higher UV_{254} and mass spectral intensities in these segments (Figure S4), which will be discussed below.

The RP-LC method is highly reproducible in terms of elution profile and TIC/UV_{254} intensity as determined from triplicate SRFA injections (Figure S5). In addition, the reproducibility of the detection of molecular formulas in individual segments (# shared peaks among triplicates / mean # peaks for triplicates) was better than 75% for segments with high TIC values (5-6 min and 10-20 min) and within the range reported for DI-ESI [33]. Segments with lower TIC values (i.e. 20-21 min) and correspondingly lower number of mass peaks also have a lower fraction of shared peaks (Figure S6).

Detection of polar compound with LC-FT-ICR-MS

Counter gradient aids in the ionization of polar compounds

The total number of unique MF (in the range 0-1000 Da and $N_{0-2}S_{0-1}$) identified in SRFA with the counter gradient (CG) and standard gradient (SG) mode were 7318 and 7000, with 5278 MF shared between the two modes. Taking the presence of MFs in all 12 segments into account, a minimum estimate of 26010 and 24600 different isomeric compounds were detected in SRFA for CG and SG, respectively.

The post-column addition of a reversed solvent mixture in the CG mode resulted in 1.5 fold higher peak intensities in the first segment (5-6) of the isocratic range, as compared to the SG mode. Correspondingly, the number of assigned formulas were 22% higher for the CG mode (Figure 1b). Similarly, in the first seven segments of the gradient range (10-17 min), a substantially higher TIC intensity (mean increase: 48%) of the CG measurement was observed (Figure 1a). The number of assigned formulas was on average 20% higher for the CG mode in these segments (Figure 1b), and the compounds shared in both modes also showed higher intensity in CG (Figure 2). The benefit of the CG mode was particularly pronounced for the heteroatom-containing formula classes, e.g. as a 36% (58%) increase in the number (intensity) of CHNO MFs. After the MeOH reached 50% in the eluent (approx. 17 min), the TIC values and # MF in the SG was slightly higher than those in the CG segments (Figure 1b, Figure S7).

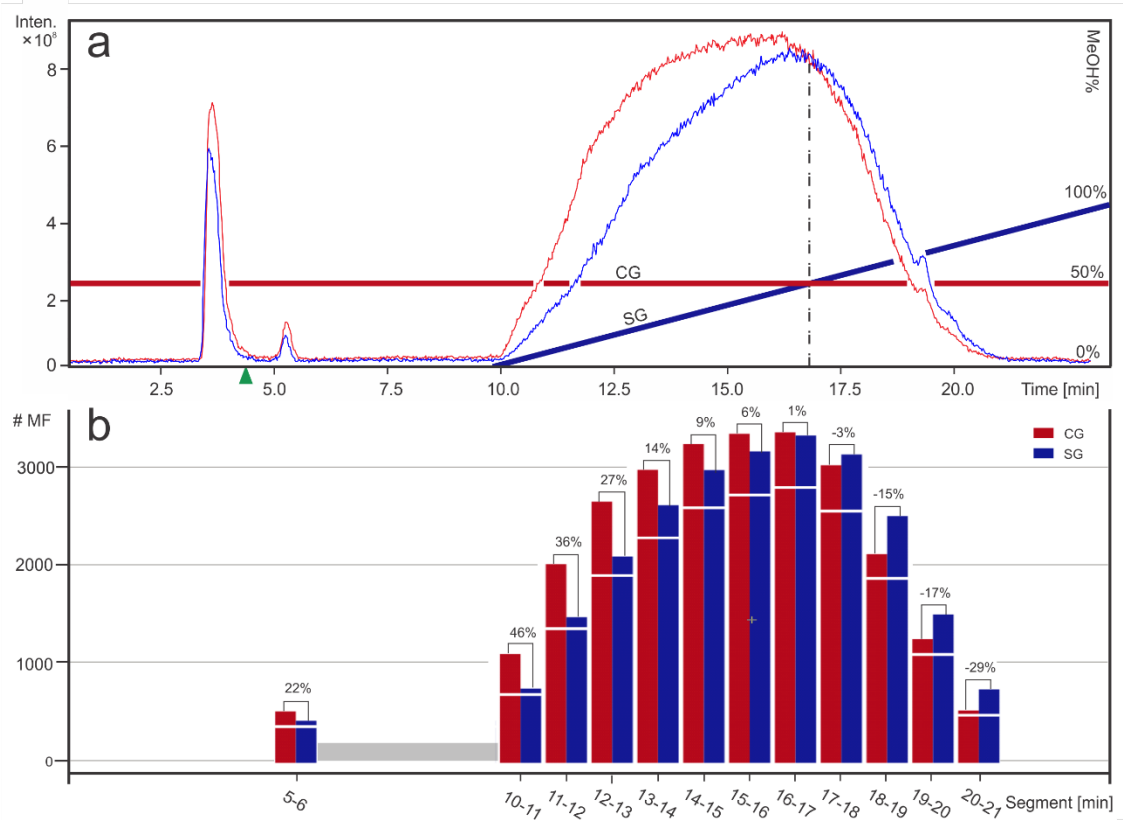


Figure 1. (a) Total ion chromatogram (TIC) of a Suwanee River Fulvic Acid standard (SRFA, injected amount: 21 µg) obtained with the LC-FT-ICR-MS method in counter gradient (CG, red) and standard gradient (SG, blue) mode (both: HSST3, 0.05% formic acid, 0.2 ml/min). The approximate solvent composition during electrospray ionization for both modes is shown. (b) Assigned number of molecular formulas in each segment and both modes. The relative difference in assigned MFs is indicated and the white lines correspond to the number of shared MF between the modes. Grey bar: No assignment was made for the segments 6-10 min due to overall low intensity of SRFA compounds.

Although highly polar compounds often have higher response factors as compared to less polar compounds in negative mode ESI , a too high water fraction in the eluent may also hamper their efficient ionization [33]. It can be assumed that the high aqueous solubility for the most polar compounds in DOM (which are already present as ions at ambient pH) results in a less efficient gas phase transfer during ESI process. In addition, evaporation of water droplets in the ESI source is reduced as compared to organic solvents. The stabilization of the water-to-MeOH ratio at 50:50 across the entire retention time range counteracts this apparent lower ionization efficiency of the highly polar DOM compounds in water, thus enhancing their detection in LC-MS methods.

Notably, this effect of higher ionization efficiency for the highly polar compounds is solely based on the solvent ratio during electrospray ionization since the post-column added solvent was buffer free in both modes. Adding a buffer to the eluent is important to ensure a long-time stable pH and reproducible separation of compounds in DOM. Hence, also for the SG mode, adding a buffer-free solvent prior ESI

introduction dilutes the buffer concentration, counteracting adverse effects of salts during electrospray ionization [32]. The non-linear reduction of ion suppression due to buffer dilution outcompetes the parallel reduction in analyte concentration and is recommended for LC-MS of NOM [47].

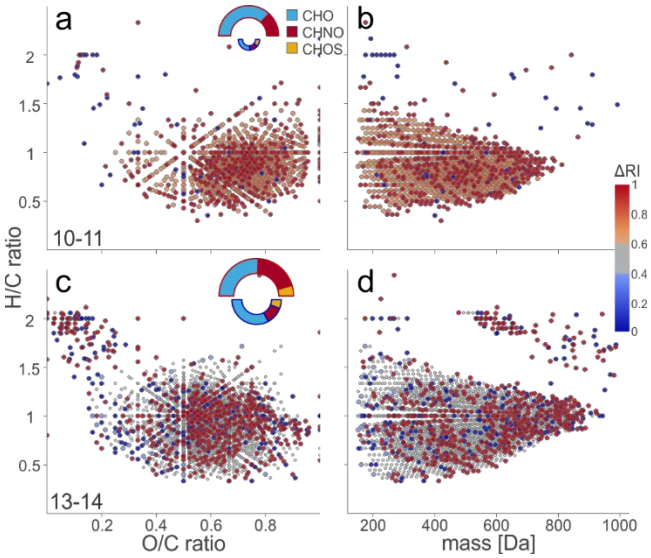


Figure 2. Assigned molecular formulas (MF) in SRFA for the segments (a, b) 10-11 min and (c, d) 13-14 min displayed by their (a, c) H/C vs O/C ratios and (b, d) H/C vs mass. The color indicates relative intensity ratios between CG (red) and SG (blue) mode. A ratio between 0.4 and 0.6 is considered as similar intensity. In (a, c), the uniquely detected MF in both modes, their number is indicated by the size of the circle (CG: upper half, SG: lower half) and the formula class distribution is color coded as CHO (blue), CHNO (red) and CHOS (yellow). Complementary figures for later eluting segments can be found in Figure S7.

The solvent ratio also influences the reproducibility of DOM peak detection. When the MeOH ratio in the solvent was lower than 15%, the MF detection reproducibility in CG segments was on average 5% higher than in SG segments (Figure S6). Low water-to-MeOH ratios were shown to reduce peak detection reproducibility for DI measurements of SRFA, and 50% of MeOH achieved highest reproducibility [33].

Q-isolation improves sensitivity for low abundance compounds in LC-MS

To further increase the sensitivity for the early eluting (segments 6-10 min) highly polar compounds, we applied a Q-isolation (CASI) mass spectral stitching approach with increased IAT (here: 0.4 s). To demonstrate the improved sensitivity for highly polar compounds, non-extracted peat pore water (PPW) samples were used with CG mode.

Highly polar compounds eluted across all early segments of PPW. This is in contrast to SRFA, where isocratic elution segments from 6 to 10 min did not show detectable elution of compounds (Figure S4). Over 2 times more MF could be assigned with CASI LC-MS across all mass windows in the segments from 6 to 10 min than as compared to full scan mode (Figure 3a). The improvement was largest for heteroatom

containing MF (CHNO, CHOS, CHNOS formula classes), which generally have lower relative intensities in DOM samples as compared to the CHO formula class (Figure S8). For example, during segment 7 – 8 min, peaks at m/z 269.0415 ($C_{10}H_{10}N_2O_7$) and 290.0882 ($C_{11}H_{17}N_1O_8$) were not detected in full scan mode (Figure S9). Increased detection of especially N-containing compounds due to sample fractionation has been observed for marine and aerosol samples likely due to reduced charge competition for the weak ionizers in ESI-(-) [26,48].

Notably, all segments showed a broad distribution of molecular mases, ranging up to 1000 Da even for the most polar segments (Figure S10a, a-f, i-n). However the large molecular mass of the early eluting DOM compound is balanced by their very high O/C ratios – a class of compounds which were previously not detected in terrestrial DOM with DI-ESI methods (Figures S8, S10b). In addition, larger molecular mass PPW compounds of the early eluting segments have on average a lower H/C ratio as compared to low-molecular weight compounds, indicating that a condensed carbon backbone contributes to the high polarity and aqueous solubility (Figure S11).

The benefits of CASI mode for LC-MS were markedly different across retention times representing changes in the mean polarity of the eluting DOM compounds. Segments of later eluting, less polar DOM profited less from CASI mode (Figure 3a). This indicates that the bias against highly polar compounds resulting from variable solvent composition during SG mode LC-MS is even larger than already suggested from the comparison with CG mode and DI, as discussed below.

The overall elution profile was also markedly different between non-extracted PPW and SRFA likely due to the different abundances and ionization efficiencies of highly polar compounds as compared to compounds with lower polarity (Figure S4). This may in part reflect the differences in sample preparation, since SRFA represents only the (water soluble) part of retained DOM on a polar XAD-8 resin at pH 2. Also PPL, the most widely used solid phase extraction sorbent for DOM, only recovers up to 65% of DOC from fresh water samples and is in addition biased against highly polar and basic compounds [18,49].

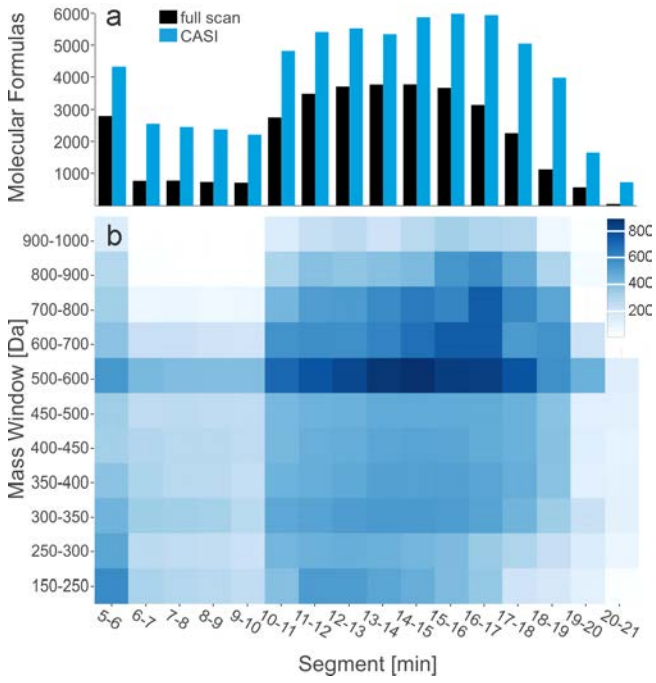


Figure 3. (a) Assigned molecular formulas (MF) for the peat pore water (PPW, injected amount: 27 µg) in segments 5-21 min obtained with the LC-CG-FT-ICR-MS method (HSST3, 0.05% formic acid, 0.2 ml/min) in CASI (blue) and full scan (black) mode. (b) Assigned MFs in different segments and mass windows (note different mass window sizes).

Considering the complete elution profile of PPW (5-21 min), CASI-LC-MS detected 64379 isomeric DOM compounds (15216 unique MF) with up to 2 nitrogen and 1 sulfur atom in the PPW, yet covering over 99% of isomers detected in full scan mode (33906 isomers of 8998 unique MF, Figure S8). In full scan mode, the sensitivity for the low abundance highly polar compound was limited as the IAT was optimized for the major elution peak (around 15 min). When applying CASI mode with 50-100 Da windows, more ions can be loaded into the ICR cell per mass window, increasing the sensitivity for low abundance ions [50,51]. The sensitivity of ultra-high resolution mass spec trometry is typically limited by number and distribution of ions in the analyzer cell and not the number of ions produced in the ESI process. Similarly, the scan speed is determined by the transient length which depends on the desired mass resolution. Hence CASI LC-MS with high IAT can be performed without loss of scan rate and chromatographic resolution. Although mass spectral segmentation is time consuming it reveals up to 2.5 times more formulas than full scan LC-MS due to the possibility of increasing the number of selected ion species in the analyzer cell. This potentially allows MS/MS experiments with even on the lowest abundance compounds in DOM.

292 **Comparison of on-line LC with direct infusion FT-ICR-MS**

293 To highlight the advantages of LC-FT-ICR-MS for the detection of polar compounds, we compared the
294 online LC-*CG*-FT-ICR-MS method with a DI measurement (4428 MF detected). The on-line LC-*CG*-FT-
295 ICR-MS method detected 3884 MF not found in the DI method, which mainly represent highly polar
296 compounds with O/C ratios > 0.6 and H/C ratios < 1 and higher than average mass (Figure 4a,b). The
297 number of detected isomeric highly polar compounds was more than 21 times larger for CG mode as
298 compared to DI (5715 vs 266 MF). Of those unique highly polar compounds 69% (3924) eluted in the early
299 segments (5 – 14 min), before the MeOH ratio reached 25%. Consequently, the mean O/C ratio of formulas

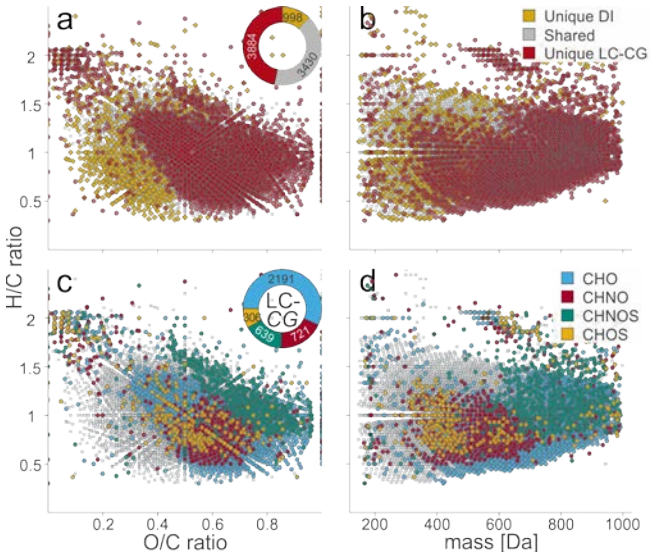


Figure 4. (a, b) Assigned shared (grey) and unique molecular formulas (MF) in SRFA measured with the LC-CG-FT-ICR-MS (sum of all segments, unique: red) and DI-FT-ICR-MS method (unique: orange). (c, d) Unique assigned MF via LC method color coded by formula class (CHO: blue, CHNO: red, CHNOS: cyan, and CHOS: yellow), with their distribution indicated by the circle in (c). A complementary figure can be found in Figure S13. Individual MF are displayed according to their (a, c) H/C vs O/C ratios and (b, d) H/C vs mass.

300 assigned in *CG* is higher than DI (0.53 vs 0.45, Table S4). In addition, also peaks with high m/z and close
301 to average H/C ratios were preferentially detected by the LC method as compared to the DI method (Figure
302 4b). Suppression of the high m/z compounds in DI measurement due to the presence of low m/z compounds
303 has been reported earlier [14,52]. In terrestrial systems compounds with similar molecular properties
304 (unsaturated, oxygen rich) represent early stage degradation products from plant derived biomass like
305 polyphenols [5]. Their molecular compositions renders them as efficient ligands for multivalent cations with
306 strong affinity to Fe mineral phases, potential transport vectors for toxic elements due to their overall higher
307 mobility in aqueous media, and precursors for disinfection byproducts [6-8]. Better access to this compound

class via LC-*CG*-FT-ICR-MS will hence improve our understanding of soil organic matter formation as and well as DOM cycling in the environment in general.

In total, 998 unique MF were detected in DI which were not detected with the counter gradient LC method. The MF have on average a lower S/N ratio (in the DI measurement) than the MF uniquely detected in LC-*CG*-MS (Figure S12). In the comparison of the CASI, LC-*CG* and DI data of the PPW sample, more than 60% of the unique MFs detected by the DI but not detected by the LC-*CG* were detected in the CASI mode. This indicates that the dilution on the LC column and corresponding spread of multiple isomers over several segments renders some m/z ratios with low abundance in DI undetectable by LC-MS methods [53]. However, the gain in information both on absolute number and with respect to chemical information (e.g. retention time) when applying LC-MS for DOM highly outcompetes this effect.

LC-*CG*-FT-ICR-MS is especially advantageous for the detection of heteroatom-containing compounds. In SRFA, 2154 and 1153 heteroatom-containing compounds were detected by LC-MS and DI, respectively (Table S4). Out of those, 1360 unique MF containing nitrogen (in formula classes CHNO and CHNOS) and 306 sulfur containing unique MF (CHOS) were detected by LC-MS (Figure 4c,d), being 2-3 times higher than the number of unique MF detected in DI (440 CHNO+CHNOS and 135 CHOS, Figure S13). While the maximum number of detected CHO MFs was between 16 and 18 min, the CHNO MF preferentially eluted at lower retention times. For instance, in segment 13-14 min and segment 17-18 min the CHO/CHNO ratio increased from 3.5 to 17.4 albeit a similar number of MFs were detected in total (2942 vs 2971). Together, this indicates that CHNO compounds tend to be more polar than CHO compounds although they have similar average H/C and O/C values in DI measurements.

Challenges in the handling and processing of data for LC-FT-ICR-MS application to NOM go beyond standard DI measurements. They include feature selection or meaningful retention time segmentation for unresolvable isomers, alignment of m/z and retention times across multiple samples as well as storage of large LC-FT-ICR data sets. These challenges remain to be resolved by developing widely applicable measurement protocols as well as standard guidelines for data processing by the community as has been put forward for imaging mass spectrometry [54].

CONCLUSIONS

A new on-line LC-FT-ICR-MS methods for the characterization of DOM was presented. Utilizing an isocratic elution step and a post-column counter gradient, the new method reveals highly polar compounds and heteroatom compounds in DOM, which are typically suppressed using DI or standard gradient LC-MS methods or have low recoveries in commonly applied SPE sample pretreatments. Further improvement in

the detection of low abundance, highly polar compounds was achieved using Q-isolation of selected mass ranges.

Overall, our online LC-*CG*-FT-ICR-MS methods allows to study the DOM in much greater detail as compared to conventional DI methods, especially the most polar fraction of DOM which, to date, was inaccessible for non-target ultra-high resolution MS. Since this fraction contains the highly mobile, yet low abundance fraction of compounds in DOM, using LC-FT-ICR-MS with counter gradient allows novel insights into this dynamic pool of DOM.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Chemicals and detailed methods description, LC-FT-ICR mass accuracy, EICs of the model compounds, TIC and UV_{254} of SRFA and PPW, VK diagrams of different segments and formula classes, comparison of assignments for LC-*CG* and DI, and signal-to-noise (S/N) ratio distribution of the unique molecular formulas in LC-*CG* and DI.

Author Contributions

The manuscript was written with contributions of all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors thank Klaus-Holger Knorr (University of Münster, Ecohydrology and Biogeochemistry Research Group) for providing the PPW sample, the Chinese Scholarship Council, the ProVIS Center for Chemical Microscopy within the Helmholtz Center for Environmental Research Leipzig, which is supported by European regional development funds (EFRE—Europe Funds Saxony) and the Helmholtz Association.

REFERENCES

(1) Lipczynska-Kochany, E. *Science of the Total Environment* **2018**, *640*, 1548-1565.
(2) Hertkorn, N.; Frommberger, M.; Witt, M.; Koch, B. P.; Schmitt-Kopplin, P.; Perdue, E. M. *Analytical Chemistry* **2008**, *80*, 8908-8919.
(3) Fasching, C.; Behounek, B.; Singer, G. A.; Battin, T. J. *Scientific Reports* **2014**, *4*, 4981.
(4) Kalbitz, K.; Schwesig, D.; Schmerwitz, J.; Kaiser, K.; Haumaier, L.; Glaser, B.; Ellerbrock, R.; Leinweber, P. *Soil Biology and Biochemistry* **2003**, *35*, 1129-1142.
(5) Waggoner, D. C.; Hatcher, P. G. *Organic Geochemistry* **2017**, *113*, 315-325.
(6) Raeke, J.; Lechtenfeld, O. J.; Seiwert, B.; Meier, T.; Riemenschneider, C.; Reemtsma, T. *Environmental Science & Technology* **2017**, *51*, 5523-5530.
(7) Lau, S. S.; Wei, X.; Bokenkamp, K.; Wagner, E. D.; Plewa, M. J.; Mitch, W. A. *Environmental Science & Technology* **2020**, *54*, 5729-5736.
(8) Farre, M. J.; Day, S.; Neale, P. A.; Stalter, D.; Tang, J. Y. M.; Escher, B. I. *Water Research* **2013**, *47*, 5409-5421.
(9) Gonsior, M.; Powers, L. C.; Williams, E.; Place, A.; Chen, F.; Ruf, A.; Hertkorn, N.; Schmitt-Kopplin, P. *Water Research* **2019**, *155*, 300-309.
(10) Mangal, V.; Stock, N. L.; Gueguen, C. *Analytical and Bioanalytical Chemistry* **2016**, *408*, 1891-1900.
(11) Reemtsma, T. *Journal of Chromatography A* **2009**, 3687–3701.
(12) Kellerman, A. M.; Dittmar, T.; Kothawala, D. N.; Tranvik, L. J. *Nature Communications* **2014**, *5*.
(13) Kujawinski, E. B.; Freitas, M. A.; Zang, X.; Hatcher, P. G.; Green-Church, K. B.; Jones, R. B. *Organic Geochemistry* **2002**, *33*, 171-180.
(14) Rowland, S. M.; Robbins, W. K.; Corilo, Y. E.; Marshall, A. G.; Rodgers, R. P. *Energy Fuels* **2014**, *28*, 5043-5048.
(15) Belov, M. E.; Anderson, G. A.; Smith, R. D. *International Journal of Mass Spectrometry* **2002**, *218*, 265-279.
(16) Smith, D. F.; Podgorski, D. C.; Rodgers, R. P.; Blakney, G. T.; Hendrickson, C. L. *Analytical Chemistry* **2018**, *90*, 2041-2047.
(17) Merder, J.; Freund, J. A.; Feudel, U.; Niggemann, J.; Singer, G.; Dittmar, T. *Analytical Chemistry* **2019**.
(18) Sandron, S.; Rojas, A.; Wilson, R.; Davies, N. W.; Haddad, P. R.; Shellie, R. A.; Nesterenko, P. N.; Kelleher, B. P.; Paull, B. *Environmental Science-Processes & Impacts* **2015**, *17*, 1531-1567.
(19) Stenson, A. C. *Environmental Science & Technology* **2008**, *42*, 2060-2065.
(20) Huber, S. A.; Balz, A.; Abert, M.; Pronk, W. *Water Research* **2011**, *45*, 879-885.
(21) Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P. *Analytical and Bioanalytical Chemistry* **2007**, *389*, 1311-1327.
(22) Koch, B. P.; Ludwichowski, K. U.; Kattner, G.; Dittmar, T.; Witt, M. *Marine Chemistry* **2008**, *111*, 233-241.
(23) Hawkes, J. A.; Sjoberg, P. J. R.; Bergquist, J.; Tranvik, L. J. *Faraday Discuss.* **2019**, *218*, 52-71.
(24) Nebbioso, A.; Piccolo, A.; Lamshöft, M.; Spiteller, M. *RSC Advances* **2014**, *4*, 23658-23665.
(25) Koch, B. P.; Ludwichowski, K.-U.; Kattner, G.; Dittmar, T.; Witt, M. *Mar. Chem.* **2008**, *111*, 233-241.
(26) Spranger, T.; Pinxteren, D. v.; Reemtsma, T.; Lechtenfeld, O. J.; Herrmann, H. *Environmental Science & Technology* **2019**, *53*, 11353-11363.
(27) Patriarca, C.; Bergquist, J.; Sjoberg, P. J. R.; Tranvik, L.; Hawkes, J. A. *Environmental Science & Technology* **2018**, *52*, 2091-2099.
(28) Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. **2017**, *4*.
(29) Hawkes, J. A.; Sjöberg, P. J. R.; Bergquist, J.; Tranvik, L. J. *Faraday Discussions* **2019**, *218*, 52-71.
(30) Kim, D.; Kim, S.; Son, S.; Jung, M. J.; Kim, S. *Analytical Chemistry* **2019**, *91*, 7690-7697.

(31) Chandran, J.; Zheng, Z.; Thomas, V. I.; Rajalakshmi, C.; Attygalle, A. B. *Analyst* **2020**, *145*, 5333-5344.
(32) Brown, T. L.; Rice, J. A. *Analytical Chemistry* **2000**, *72*, 384-390.
(33) Kido Soule, M. C.; Longnecker, K.; Giovannoni, S. J.; Kujawinski, E. B. *Organic Geochemistry* **2010**, *41*, 725-733.
(34) Gaspar, A.; Schrader, W. *Rapid Communications in Mass Spectrometry* **2012**, *26*, 1047-1052.
(35) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. *Analytical Chemistry* **2007**, *79*, 4595-4602.
(36) Krajewski, L. C.; Rodgers, R. P.; Marshall, A. G. *Analytical Chemistry* **2017**, *89*, 11318-11324.
(37) Lechtenfeld, O. J.; Koch, B. P.; Geibert, W.; Ludwichowski, K.-U.; Kattner, G. *Analytical Chemistry* **2011**, *83*, 8968-8974.
(38) Lechtenfeld, O. J.; Koch, B. P.; Gašparović, B.; Frka, S.; Witt, M.; Kattner, G. *Marine Chemistry* **2013**, *150*, 25-38.
(39) Patriarca, C.; Balderrama, A.; Može, M.; Sjöberg, P. J. R.; Bergquist, J.; Tranvik, L. J.; Hawkes, J. A. *Analytical Chemistry* **2020**.
(40) Gavard, R.; Palacio Lozano, D. C.; Guzman, A.; Rossell, D.; Spencer, S. E. F.; Barrow, M. P. *Anal. Chem.* **2019**, *91*, 15130-15137.
(41) Koch, B. P.; Kattner, G.; Witt, M.; Passow, U. *Biogeosciences* **2014**, *11*, 4173-4190.
(42) Herzsprung, P.; Hertkorn, N.; von Tümpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. *Analytical and Bioanalytical Chemistry* **2014**, *406*, 7977-7987.
(43) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2007**, *8*, 105.
(44) Hawkes, J. A.; Patriarca, C.; Sjöberg, P. J. R.; Tranvik, L. J.; Bergquist, J. *Limnology and Oceanography Letters* **2018**, *3*, 21-30.
(45) These, A.; Reemtsma, T. *Analytical Chemistry* **2003**, *75*, 6275-6281.
(46) Gavard, R.; Jones, H. E.; Palacio Lozano, D. C.; Thomas, M. J.; Rossell, D.; Spencer, S. E. F.; Barrow, M. P. *Analytical Chemistry* **2020**, *92*, 3775-3786.
(47) Lohse, M.; Blaser, S.; Vetterlein, D.; Schlűter, S.; Oburger, E.; Reemtsma, T.; Lechtenfeld, O. J. *Analytical Chemistry* **2020**.
(48) Reemtsma, T.; These, A.; Linscheid, M.; Leenheer, J.; Spitzy, A. *Environmental Science & Technology* **2008**, *42*, 1430-1437.
(49) Raeke, J.; Lechtenfeld, O. J.; Wagner, M.; Herzsprung, P.; Reemtsma, T. *Environmental Science: Processes & Impacts* **2016**, *18*, 918-927.
(50) Southam, A. D.; Weber, R. J. M.; Engel, J.; Jones, M. R.; Viant, M. R. *Nature Protocols* **2017**, *12*, 310-328.
(51) Riedel, T.; Dittmar, T. *Analytical Chemistry* **2014**, *86*, 8376-8382.
(52) Reemtsma, T.; These, A. *Anal. Chem.* **2003**, *75*, 1500-1507.
(53) Spranger, T.; Pinxteren, D. V.; Reemtsma, T.; Lechtenfeld, O. J.; Herrmann, H. *Environ. Sci. Technol.* **2019**, *53*, 11353-11363.
(54) Schramm, T.; Hester, Z.; Klinkert, I.; Both, J.-P.; Heeren, R. M. A.; Brunelle, A.; Laprévote, O.; Desbenoit, N.; Robbe, M.-F.; Stoeckli, M.; Spengler, B.; Römpp, A. *Journal of Proteomics* **2012**, *75*, 5106-5110.