Impact of instrument and experiment parameters on reproducibility and repeatability of peaks within ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter

Melissa C. Kido Soule^a, Krista Longnecker^a, Stephen J. Giovannoni^b, and Elizabeth B. Kujawinski^{a,*}

(a) Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole MA 02543, USA

(b) Department of Microbiology, Oregon State University, Corvallis OR 97331, USA

* - Corresponding author:

Tel: (508) 289-3493

Email address: ekujawinski@whoi.edu (Elizabeth B. Kujawinski)

Keywords: DOM, electrospray ionization, Fourier-transform, FT-ICR MS

ABSTRACT

Natural dissolved organic matter (OM) is a complex heterogeneous mixture of compounds that have defied traditional characterization using standard analytical methods. Electrospray ionization mass spectrometry, particularly ultrahigh resolution mass spectrometry, has provided a new platform for compositional assessment of this important pool of the Earth's reduced carbon. Here, we propose a framework for optimization of instrument and experiment parameters for high quality data acquisition using Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). We report the impact of these parameters on reproducibility of peak detection and repeatability of peak height in replicate injections of Suwannee River fulvic acid (FA), a common terrestrial OM standard. In addition, we examine the variability in peak detection and peak height among different types of experimental replicates of dissolved OM derived from laboratory cultures of *Candidatus Pelagibacter ubique*, a ubiquitous marine α -proteobacterium.

1 Introduction

Dissolved organic matter (DOM) refers to reduced-carbon compounds in aquatic environments. The size of this pool (~700 Pg) in the oceans alone rivals that of atmospheric CO₂ (~750 Pg) and thus its spatial and temporal dynamics in all aquatic regions are of paramount importance for mechanistic understanding of the global carbon cycle (Hedges, 2002). Residence times of DOM are intimately tied to its reactivity toward biotic and abiotic degradation processes, which is in turn determined by its composition. As a result, numerous investigations have examined DOM composition in order to predict DOM variability as a function of biogeochemical processes.

A recent advance in the study of DOM composition has been the introduction of electrospray ionization (ESI) coupled to Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). The technique provides the first tool with the power to examine individual molecules within DOM and to resolve the molecular-level impact of different biogeochemical processes on DOM composition. Electrospray ionization, which preferentially ionizes water-soluble, hydrophilic compounds, is a “soft” ionization technique with low incidence of fragmentation for DOM molecules (Rostad and Leenheer, 2004). The ultrahigh resolution and mass accuracy of FT-ICR MS provides molecular masses accurate to within 1 ppm, which often enables the determination of elemental formulae from the mass measurement alone (Kim et al., 2006). Numerous investigators have used ESI FT-ICR MS to characterize DOM collected from rainwater (Altieri et al., 2009), atmospheric aerosols (Wozniak et al., 2008), glaciers (Grannas et al., 2006; Bhatia et al., in press), freshwater (Kim et al., 2006; Sleighter et al., 2009), the coastal ocean (Sleighter and Hatcher, 2008; Kujawinski et al., 2009), the open ocean (Dittmar and Koch, 2006; Koch et al., 2008), sediment porewater (Schmidt et al., 2009), and culture studies (Kujawinski et al., 2004; Rossello-Mora et al., 2008).

Obtaining and interpreting ESI FT-ICR mass spectra of DOM is, however, not without analytical challenges. Extraction protocols to isolate and concentrate DOM from marine samples are complicated by the low concentration of OM relative to inorganic salts. Sample matrix effects and large differences in ionization efficiency among compounds mean that not all components of the DOM mixture are quantitatively represented in the ESI FT-ICR mass

spectrum. At ultrahigh resolution, a typical DOM mass spectrum contains thousands of peaks with tens of peaks per nominal mass. Each peak represents a distinct elemental formula with numerous structural isomers, posing a challenge for analysis and interpretation. To date, most ESI FT-ICR MS analyses of DOM have addressed analytical challenges associated either with sample preparation and ion production, or with data analysis and interpretation. Researchers have made important contributions in these areas by optimizing protocols for DOM extraction and concentration (Kim et al., 2003; Koprivnjak et al., 2006; Dittmar et al., 2008; Morales-Cid et al., 2009), chromatographic pre-separation (Fievre et al., 1997; Koch et al., 2008; Reemtsma et al., 2008; Stenson, 2008; Gaspar et al., 2009), and ESI direct infusion (Brown and Rice, 2000; Kujawinski et al., 2002; Stenson et al., 2003; Sleighter et al., 2008). Major advances have also been made for elemental formula assignments (Koch et al., 2005; Kujawinski and Behn, 2006; Koch et al., 2007; Kunenkov et al., 2009), graphical and compositional analysis (Kim et al., 2003; Stenson et al., 2003; Wu et al., 2004; Koch and Dittmar, 2006), and multivariate statistical analysis (Koch et al., 2005; Koch et al., 2008; Rossello-Mora et al., 2008; Gougeon et al., 2009; Kujawinski et al., 2009).

In contrast, few studies have focused on the data acquisition stage of FT-ICR MS analysis of DOM, i.e. the variability in the measurement system itself. Although some studies have improved mass resolution, sensitivity and mass accuracy by optimizing select instrument and experimental parameters (Kujawinski et al., 2002; Stenson et al., 2002; Sleighter et al., 2008), only a few have examined the reproducibility of ion detection and/or the repeatability of peak heights within injection or experimental sample replicates (Parsons et al., 2009; Payne et al., 2009). Understanding the variability in these two parameters is essential for making valid spectral interpretations as well as for sample comparison within a dataset and across experiments, particularly when multivariate statistical tools are subsequently applied.

Here, we test instrument and experimental parameters to assess variability in ESI FT-ICR mass spectra of DOM. Using Suwannee River Fulvic Acid (SRFA), we optimize parameters associated with ion population size (ion accumulation time) and detection time (number of transients and mass resolution) under full and selected ion monitoring (SIM) data acquisition modes. We then apply these optimal instrument conditions to assess the reproducibility of ion

detection and repeatability of peak height in SRFA mass spectra under full and SIM analysis modes. Last, we examine a set of marine DOM samples to compare the variability between different experimental strategies for sample replication.

2 Experimental Methods

2.1 Samples

2.1.1 SRFA.

SRFA was acquired from the International Humic Substances Society. A 1 mg/mL stock solution was prepared with ultrapure water (Optima LC-MS, Fisher Scientific) and stored in the dark at 4 °C. For negative ion mode analysis, it was diluted to 0.1 mg/ml in 50:50 MeOH/water (v/v). For positive ionization mode, it was diluted to 0.25 mg/ml in 50:50 MeOH/water (v/v) and 0.1% formic acid. All working solutions were prepared immediately prior to FT-ICR MS analysis and were used within 45 min to limit ester formation (McIntyre and McRae, 2005).

2.1.2 Marine DOM.

For this work, we examined DOM in laboratory cultures of *Candidatus Pelagibacter ubique* (HTCC1062), a common marine heterotrophic bacterium in the SAR11 clade of the α -Proteobacteria (Rappé et al., 2002). The bacteria were grown in sterilized seawater collected from the Oregon coast in three 10 l polycarbonate carboys under a variable light cycle (Connon and Giovannoni, 2002). Cell growth was monitored until the culture reached maximum density at which time the cells were removed by 0.2- μ m filtration. DOM from the filtrate was extracted according to published methods (Kim et al., 2003). In brief, filtrate was acidified with concentrated HCl until the pH ranged between 2 and 3. The filtrate was passed through stacked C₁₈ and SDB extraction disks, conditioned according to manufacturer's instructions. Disks were washed with 10-20 ml pH 2 Milli-Q water and DOM was eluted using 8 ml of 70% MeOH (v/v). Extracts were concentrated by vacuum centrifugation, re-dissolved in a known volume of 70% MeOH and frozen until analysis. We assessed the variability in DOM composition from two experimental approaches to sample replication. Samples were obtained either as "extraction

replicates” (three extractions from a single carboy) or as “treatment replicates” (one extraction from each of three carboys). Hereafter, the DOM from this experiment will be referred to as SAR11 DOM.

The salt levels in the SAR11 DOM samples remained high following C₁₈/SDB extraction and we could not achieve stable electrospray. We desalted the samples further using microscale extraction with ZipTip® pipette tips containing C₁₈ resin (Millipore). Our ZipTip protocol consisted of 5 x 10 µl pre-conditioning rinses with MeOH followed by an equal volume of pH 3 water (adjusted with HCl). After conditioning, 150-200 µl of sample was passed over the resin in 10 µl aliquots. The sample was then rinsed with an equal volume of pH 3 water (adjusted with HCl). DOM was eluted from the C₁₈ resin with 5 µl of 70% MeOH, and then diluted (with 70% methanol) up to a final volume of 100 µl. The ZipTip-extracted SAR11 samples were used without further dilution for positive ion mode ESI FT-ICR MS analysis.

For negative ionization mode, chromatographic pre-separation was used for more extensive desalting of the SAR11 DOM samples. Liquid chromatography (LC) and fraction collection was performed on an Agilent HPLC system with a C₁₈ reversed phase column (Thermo Hypersil Gold, 5 µm, 150 x 2.1 mm). The LC solvents were Solvent A: water with 0.05% formic acid and Solvent B: methanol with 0.05% formic acid. In the gradient, solvent B increased linearly from 5% to 100% between 3 and 22 min, set to 100% B for 10 min to rinse the column, and was re-equilibrated at 5% B for 13 min. The run time was 45 minutes and the flow rate was 250 µl min⁻¹. A single fraction between 4 and 24 min was collected to cover the range when most of the DOM eluted from the column, as evidenced by a single broad feature in the chromatogram. Fractions from two 25 µl injections were combined and dried. Prior to ESI FT-ICR MS negative ion mode analyses, the LC-desalted DOM samples were re-suspended in 150 µl of 70% MeOH.

2.2 FT-ICR MS data acquisition.

All samples were analyzed using a hybrid 7 Tesla linear ion trap (LTQ) Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer (LTQ FT Ultra, Thermo Electron Corp.) The instrument was externally calibrated to a mass accuracy < 2 ppm using a standard solution

from Thermo Fisher Scientific; spectra were not further calibrated except where otherwise noted. All spectra within each comparison were collected using the same external calibration and shifts in m/z values >1 ppm were not observed. Samples were infused into the ESI source with a flow rate of $4 \mu\text{l min}^{-1}$. The ESI spray voltages were 4.2 kV (negative mode) and 4.4 kV (positive mode) for SRFA. For SAR11 DOM, the spray voltage was 4.15 kV under both positive and negative ionization. The capillary temperature was 265°C . Data were acquired using a conventional full scan mode (m/z 200-1000) and two selected ion monitoring (SIM) modes, wide and narrow SIM. For this instrument, the mass window width for wide SIM mode must be between m/z 30 and 600, and for narrow SIM, the mass window width is limited to 30 m/z . A resolving power of 400,000 ($m/\Delta m_{50\%}$ at m/z 400, where $\Delta m_{50\%}$ is the full peak width at half peak height) was used, corresponding to a transient collection time of ca. 4 s. Data were collected as individual transient files (the time domain ICR signal) and as a RAW file using Xcalibur 2.0. Transients were processed using custom-written MATLAB code provided by Southam et al. (2007). Briefly, for each sample the individual transient files were averaged if the total ion current (TIC) was $\geq 20\%$ of the maximum TIC for each set of transients collected. The data were Hanning apodized, zero-filled once, and fast Fourier transformed. Peaks with a signal/noise ratio ≥ 5 were detected, where the noise is defined as the root-mean-square (RMS) calculated for a signal-free region of the spectra (Southam et al., 2007). The resulting data for each sample are a list of m/z values and their corresponding peak heights. Data from individual windows in either SIM mode were stitched together using the approach detailed in Southam et al. (2007). Except where otherwise noted, each sample injection was repeated three times.

2.3 Instrument optimization.

Instrument parameters that were optimized include: i) wide SIM isolation window size, ii) the automatic gain control (AGC) target level, which corresponds to the number of ions transferred from the linear ion trap to the FT-ICR cell, iii) the number of transients acquired, and iv) the mass resolution. These parameters were chosen for optimization because of their key role in the acquisition of FT-ICR MS data (Marshall et al., 1998). To identify optimal instrument settings, we evaluated the transient quality, the number of peaks found, the proportion (%) of

peaks shared between replicates, and the spectral noise levels. All statistical analysis for these optimizations was conducted in Matlab 7.5 (Mathworks).

To optimize the number of SRFA ions detected in wide SIM mode, we tested five mass window widths, m/z 50, 100, 200, 400, and 600, each centered at two nominal masses, m/z 400 and 600. The tests were conducted in negative ionization mode; the AGC level was set at 5×10^5 , and the mass resolution was 400,000 throughout. The same number of transients was collected and processed for each of the SIM windows tested. To determine the optimal number of transients for full scan mode, 500 transients were collected in triplicate between m/z 200 and 1000. For wide SIM, 200 transients were collected in triplicate using a 100 m/z window centered at two values, m/z 400 and 600 (negative ion mode) or m/z 360 and 600 (positive ion mode). In narrow SIM, 100 transients were collected in triplicate using a 30 m/z window centered at m/z 400 and 600 (negative ion mode) or at m/z 360 and 560 (positive ion mode).

2.4 *Data processing and statistical analysis.*

Peak lists from each comparison set were aligned using MATLAB code from Mantini et al. (2007). The alignment code creates a single master peak list of the m/z values in each sample set, where peaks within 1 ppm across different samples are considered to be the same. Tests with the alignment code reveal that it aligns m/z values correctly when the actual instrument error is up to twice the expected error (data not shown). The aligned data were used to determine (i) the total number of peaks in each sample, (ii) the number of peaks shared between three replicate injections for each sample, (iii) the proportion (%) of peaks shared between replicates, and (iv) the relative standard deviation (RSD; Parsons et al., 2009; Payne et al., 2009), also known as the coefficient of variation, in peak height across the three replicate injections. The proportion (%) of peaks shared was defined as the total number of peaks shared between the three replicates divided by the average number of peaks found in the three replicates. To calculate RSD for peaks found in all three injections, the standard deviation in peak height was divided by the mean peak height. A median value was obtained for the entire spectrum and was multiplied by 100. A low RSD value indicates less variability (greater reproducibility) in peak height and a high RSD means greater variability in peak height. Peak heights were $\log_{10}$ -transformed prior to

analysis with one-way ANOVAs followed by multiple comparisons with a Bonferroni correction.

3 Results and Discussion

First, we needed to determine the optimal window for wide SIM mode data acquisition. We were then able to optimize our instrument parameters for all data modes systematically: full scan mode, wide SIM mode and narrow SIM mode. Below, we present the results of our optimization efforts for SRFA. This sample has been used in other applications as a terrestrial DOM standard and is commonly analyzed as a comparative tool for other DOM samples (Stenson et al., 2003; Bhatia et al., in press). A summary of our recommendations for optimal instrument parameters is provided in Table 1. We then used our optimized instrumental parameters to assess DOM spectral variability associated with different solvent systems and different experimental strategies for sample replication, in particular extraction replicates and treatment replicates.

3.1 Wide SIM window optimization.

As the width of the window increased, the mean number of peaks identified at each nominal mass decreased (Fig. 1A). For example, a decrease in the window size from 200 to 50 m/z led to an approximate three-fold increase in SRFA peak numbers per nominal mass. The enhanced peak detection observed in smaller wide-SIM windows is reasonable given that an equal number of ions (AGC level: 5×10^5) are transferred to the FT-ICR cell at each SIM window setting. Detection of low abundance ions should be enhanced within increasingly narrow windows because the total ion population consists of fewer different masses with sufficient ion number. Considered alone, the impact of window size on the average number of peaks detected would favor the smallest mass window for wide SIM data collection. However, we observe “edge effects” similar to those reported by Southam et al. (2007), such that ions near the edge of the SIM window are lost or significantly reduced in peak height. We quantified this observation by calculating the proportion of peaks lost (“% of m/z lost”: Eq. 1, where m/z_{set} is the m/z value

of the window edge, m/z_{obs} is the m/z value of the first ion observed from the edge and w_{win} is the width of the window):

$$\% m/z \text{ lost} = (m/z_{\text{set}} - m/z_{\text{obs}}) / w_{\text{win}} \quad (1)$$

A higher proportion of peaks is lost at the edges for the smaller wide SIM mass windows, with more peaks lost at the upper end of the mass window than at the lower end (Fig. 1B). Indeed, as the wide SIM window size decreases, the proportion of peaks lost as a result of edge effects increases. For example, a 50 m/z window, while providing high peak coverage, results in $\sim 16\%$ of peaks lost at the upper edge of the window. As a percentage, this is approximately 4 times as great an ion loss as for a 200 m/z window (Fig. 1B). Given the opposing constraints of peak numbers and proportion of peaks lost in Fig. 1, we conclude that a 100 m/z wide SIM window is an optimal choice for SRFA and other DOM samples.

In addition to peak loss, we also observed reduced peak heights at the wide SIM window edges. Because individual SIM windows are ultimately stitched together to cover the full mass range, this tailing in ion peak height can be mitigated by acquiring data across overlapping adjacent mass regions (Southam et al., 2007). We determined the optimum overlap by dividing the peak height of peaks found in a wide SIM window by the peak heights of the same peaks in a narrower window. For example, the height ratio between peaks in a 200 to 100 m/z window (both centered at m/z 400 and 600) is shown in Fig. 2. There is no peak tailing in the center of the 100 m/z mass range and the height ratio is essentially unity. However, because of “edge effects”, the ratio becomes greater than one and increases toward the edge. Visual inspection shows that the increase in the height ratio extends about 10-15 m/z at each edge, indicating that a total overlap of 30 m/z is necessary. Similarly, we determined that wide SIM windows of m/z 50, 200, and 400 would require overlaps of 30 m/z , 40 m/z and 100 m/z , respectively (data not shown).

3.2 AGC target level optimization.

The number of ions transferred from the ion-trap to the FT-ICR detection cell (determined by the AGC setting) was optimized to give the best sensitivity and spectral quality. SRFA was analyzed in negative ion mode using a range of AGC target values for the three scan types: full, wide SIM and narrow SIM. For full scan mode, the AGC levels tested were 1×10^5 , 2×10^5 , 5×10^5 ,

1×10^6 , 2×10^6 and 3×10^6 . For the wide SIM mode, the AGC target values were 1×10^5 , 2×10^5 , 5×10^5 , 1×10^6 and 2×10^6 . Each wide SIM analysis was done using a 100 m/z window centered at m/z 400 and 600. For narrow SIM analysis we used a 30 m/z window (Southam et al., 2007), centered at m/z 400 and 600 to test AGC values 1×10^4 , 2×10^4 , 5×10^4 , 1×10^5 , 2×10^5 and 5×10^5 . To determine the optimal AGC setting for each scan type, we evaluated the effects of ion number on transient quality, number of peaks, percent of peaks shared between replicate scans (i.e. peak reproducibility) and the RMS noise level.

Fig. 3 shows transients of 200 co-added scans for SRFA collected under full scan negative ion mode at selected AGC target values. We did not test AGC values $> 3 \times 10^6$ because ion accumulation times exceeded the maximum time set by the manufacturer and we could not accurately determine the number of ions introduced into the cell for each transient. A good quality transient signal is one that lasts for the duration of the collection period, i.e. the signal intensity continually decays with time, forming a cone-like profile. At the 1×10^5 AGC target value, the full scan transient (Fig. 3A) is of poor quality because the transient signal decays only weakly at the start of the time domain and is fully flat by the second half of the signal (Fig. 3D). However, at AGC values of 1×10^6 and 3×10^6 (Fig. 3B, C), the transient lasts throughout the entire collection period, as evidenced by the cone-like tapering of the signal in the second half of the transient (Fig. 3E, F). This effect of ion number on transient quality can be explained by considering the stability of ion populations in the FT-ICR cell. Ions injected into the ICR cell are excited to greater cyclotron orbits by the application of an rf electric field. Ions of differing mass-to-charge ratios form separate ion clouds, which oscillate at specific cyclotron frequencies. The decay of all ion cloud orbits in the cell over time is recorded as a transient. Ion population size plays a key role in ion cloud stability and transient duration, such that the time domain signal decays more rapidly for clouds with fewer ions (Bresson et al., 1998). Thus, at the low AGC target values tested, the transient of smaller ion clouds decays well before the end of the detection interval.

In addition to transient quality, we looked at the effect of ion population size on the number of peaks detected, the percent of peak overlap between replicates, and RMS noise. Fig. 4 shows these results for SRFA in full scan negative ion mode. By all measures shown in Fig. 4, the AGC

value of 1×10^5 provided insufficient ions for good quality spectra. Too few ions in the FT-ICR cell reduced sensitivity and led to poor peak detection, a high RMS noise level and the lowest percentage of peaks shared between replicate spectra. In comparison, at the 1×10^6 and 3×10^6 AGC targets, larger ion population sizes produced higher peak numbers, lower noise, and better reproducibility. Of the two AGC levels, the number of peaks is highest and increases most steeply with number of co-added scans at the 1×10^6 level.

Given the effects of ion number on transient quality, peak number, peak reproducibility, and RMS noise (Figs. 3 and 4), we conclude that an AGC value of 1×10^6 is optimal for SRFA in full scan mode. We performed the same series of evaluations for SRFA in wide SIM and narrow SIM modes (Supplementary Figs. 1 and 2) and find that for wide SIM and narrow SIM, the optimal AGC target levels are 5×10^5 and 5×10^4 , respectively. Because the AGC target values correspond simply to the number of ions transferred to the FT-ICR cell, the optimal AGC levels tested in negative ion mode can also be extended to positive ion mode.

3.3 Scan number optimization.

We optimized the minimum number of scans needed for SRFA by evaluating the effects of signal averaging on spectral parameters as above. Analyses were conducted in both positive and negative ion mode using the optimal AGC settings already described. The results for full scan mode show that as the number of averaged transients increases, both the number of peaks detected (Fig. 5A) and the number of peaks shared (Fig. 5B) increase while the RMS noise level decreases (Fig. 5C). Although more signal averaging, even beyond 500 scans, is potentially better in terms of peak detection, lower noise, and RSD levels (Fig. 5C, D), it comes at the expense of longer analysis time and lower throughput. In addition, spectral reproducibility, as shown by the percent of peaks shared between replicates, does not increase beyond 87% reached at ~150 scans. Thus, we find that for full scan in both positive and negative ion modes, averaging 150-200 transients per spectrum is sufficient. We repeated this analysis for wide SIM and narrow SIM modes (Supplementary Figs. 3 and 4). For narrow SIM, we recommend the co-addition of 50 scans per 30 m/z window. For wide SIM with a 100 m/z window, we find that 100 scans are optimal at the lower mass range where peak height is highest (<500 m/z), but that 150

scans are better at higher m/z values. As with the full scan data, these conclusions are based both on the likelihood of repeatedly detecting the same peak (percent of peaks shared) and on the variability in peak height for each set of detected peaks (RSD).

3.4 Resolution optimization.

To optimize mass resolution for analysis of SRFA and other DOM samples, we collected data (in triplicate) in full scan, wide SIM and narrow SIM modes at a resolving power of $m/\Delta m_{50\%} = 400,000$, using the instrument settings already described. Because lower resolution in FT-ICR MS is achieved simply by collecting a transient for less time, the full transient was halved or quartered prior to Fourier-transformation to simulate spectra collected at resolving power of $m/\Delta m_{50\%} = 200,000$ or $100,000$, respectively. Table 2 shows that in full scan mode for both polarities, higher resolving power corresponds to more reproducible peak detection between replicates. The proportion of peaks shared increases 7-8% between resolutions of 100,000 and 400,000. However, higher resolving power results in fewer total peaks detected in the full scan spectrum (Table 2). This appears to be counter-intuitive since higher mass resolving power will separate more peaks per nominal mass, so should result in the observation of more peaks. However, higher resolving power also requires longer transient time. If ion signals do not last for the entire scan time, noise will be averaged into the Fourier transform and reduce spectral sensitivity for those ions. This would be most pronounced for low abundance ions because their less stable ion clouds are more susceptible to early break-up (Bresson et al., 1998). Thus the full scan mode data shows that at higher resolution, fewer total peaks may be detected, but their detection is more reproducible.

Although more peaks are observed under full scan mode with lower mass resolving powers, we use a resolution of $m/\Delta m_{50\%} = 400,000$ on the 7 T LTQ FT Ultra for DOM analysis. High mass resolution is critical for DOM measurements since it provides the peak separation necessary for fully characterizing complex mixtures, particularly those with heteroatom contributions such as S and N (Kujawinski et al., 2002; Stenson et al., 2002; Kim et al., 2006). To mitigate the impact of lower peak detection under full scan mode at this resolution, we use signal averaging to increase the signal/noise ratio and to detect more peaks.

Under SIM mode (narrow and wide), increased resolution has a different impact on the number of peaks detected in the SRFA mass spectrum. The positive ion mode spectra show that the number of peaks increases with increasing mass resolution for both narrow and wide SIM modes. This result is consistent with the previously observed enhancement in spectral dynamic range provided by SIM methods when applied to metabolite analysis (Southam et al., 2007). Because only a narrow mass-to-charge window is analyzed in SIM mode, fewer individual ion species contribute to the AGC target value. This leads to better sensitivity of low abundance ions and more ions detected per nominal mass, relative to full scan mode. Furthermore, the ion cloud population for a given ion species is likely to be larger in SIM mode, and therefore less susceptible to premature decay during transient collection. Consequently, in positive ion mode we find that the highest resolution examined (400,000) is able to separate the most peaks per nominal mass for both SIM mode types, a desirable achievement for DOM analysis. In contrast, the negative ionization SIM mode data does not completely follow the expected trend in the number of peaks detected with increasing resolution. We do not have a good explanation for this result. For example, the mean number of peaks increases between 100,000 and 200,000 mass resolution, but then decreases at 400,000 resolution.

The effect of resolving power on peak height reproducibility was also evaluated from calculations of RSD for all three scan modes under positive and negative mode ionization. The data generally show that peak height reproducibility is slightly better at lower mass resolution, although we did not observe large variances in peak height reproducibility.

3.5 *Reproducibility in peak height and detection.*

To assess the peak height reproducibility, spectral peaks from triplicate SRFA injections were aligned and then assigned to one of three groups: i) peaks found in all three replicates, ii) peaks found in two replicates and iii) peaks only found in one replicate. Optimized instrument settings were used throughout. Mean peak heights were then calculated for each peak. We find that peaks found in all three replicates have significantly higher peak heights compared to peaks found in one or two replicates (one-way ANOVA, $p \ll 0.0001$ followed by multiple

comparisons with a Bonferroni correction). Thus, the dominant peaks are the ones most reliably present in repeat injections. We find this to be the case for full and SIM scan types in both positive and negative ion modes. The variability in peak heights for ions found in one, two, or three injection replicates is further illustrated in Fig. 6 for the three scan types in negative ion mode (also Supplementary Fig. 5; positive ion mode data for all scan types in Supplementary Fig. 6). For full scan mode, the majority of peaks fall along the 1:1 line, except, notably, when peak heights are low and peaks are found in only one or two replicates. The greater deviation of peaks from the 1:1 line in SIM modes illustrates that peak heights are less reproducible than in full scan, which is also expressed as higher RSD values for wide SIM and narrow SIM compared to full scan data (Table 2).

In addition to peak height variability, we also determined the reproducibility of peak detection (Table 2). In both positive and negative ion modes, the percent of peaks shared is greatest for full scan, compared to wide and narrow SIM scans. High peak overlap combined with the low peak height variability show that spectral reproducibility is best under full scan conditions. However, we find that the number of peaks detected across the same mass range is higher for both SIM scan types than for full scan. For example, we observe small peaks at ~0.5 Da mass defect between m/z 277 and 1000 in the wide SIM negative ion mode SRFA spectrum. We find these masses only in negative ion mode, at 400,000 resolution for wide SIM and at lower mass resolution for narrow SIM. These additional peaks are not present in full scan mode under any conditions. We attribute these masses to multiply (negatively) charged SRFA compounds, consistent with previous observations (Brown and Rice, 2000; Gaspar et al., 2009). These peaks are only observed when ion diversity is low and noise levels are sufficiently reduced. From this observation, we conclude that different advantages are afforded by selecting either full or SIM scan types; narrow or wide SIM conditions give the best peak coverage (at lower sample throughput), while full scan provides the best repeatability with the highest throughput.

3.6 *Effect of solvent composition on peak detection.*

Solution composition has been shown to alter the apparent molecular weight distribution of DOM mass spectra (Brown and Rice, 2000; Gaspar et al., 2009). Here, we examine the effects of solvent composition on spectral reproducibility in terms of the number of peaks and the percent of peaks shared between injection replicates of negative ion mode SRFA spectra. SRFA was diluted to 0.1 mg/mL in 25:75, 50:50, and 70:30 methanol/water (v/v). Solvent compositions containing 50-70% methanol afford the best reproducibility: 87% overlap in 50% methanol and 83% overlap in 70% methanol. (Table 3) A greater number of peaks under full scan mode was found using 25% methanol, but spectra under such solvent conditions were less reproducible: 68% of peaks shared between replicates. The decreased reproducibility in 25% methanol might be due to the challenges that a higher proportion of water impose on solvent evaporation from ESI droplets as they travel from atmospheric pressure to vacuum conditions at the spectrometer inlet. Therefore, selecting a solvent composition with a higher proportion of methanol is advantageous in terms of reproducibility for ESI FT-ICR MS analysis of SRFA, and likely for other DOM samples. We did not observe significant changes in elemental composition with different solvent systems (Supplementary Fig. 7).

3.7 *Variability associated with extraction or experimental treatment.*

We also tested the impact of different experimental approaches to sample replication on the reproducibility of mass spectra using DOM from a bacterial culture. SAR11 DOM samples were obtained from extraction replicates (three extractions from a single carboy) and from treatment replicates (one extraction from each of three carboys). Data for desalted samples were collected under full scan mode using the optimized instrument settings. Full scan mode without replicate injections, rather than either wide or narrow SIM modes, was chosen on the basis of the need to balance sample throughput and spectral coverage. Table 4 summarizes the data for the two sample types, extraction replicates and treatment replicates, under positive and negative ionization modes. For both ionization modes, the reproducibility of peak detection, as shown by the proportion (%) of peaks shared, is greater for the extractions from different carboys (74-76%) than for three extractions from the same carboy (68-69%). That is, the treatment replicates show

less variability than the extraction replicates. Comparison of the full scan mode data at 400K resolution in Tables 2 and 4 also shows that the reproducibility of peak detection and peak height (RSD) of the treatment replicates is similar to instrument (injection) reproducibility for positive ion mode, but slightly worse in negative ion mode.

4 Conclusions

DOM characterization using ESI FT-ICR MS is expensive and labor-intensive and has therefore been limited to few laboratories. However, as the technique gains prominence in the Earth sciences and more laboratories acquire the capability, formal assessment of the technique's caveats and limitations will be necessary, particularly those that affect a peak's reproducibility and peak height variability. This has become increasingly important with the recent application of advanced statistical tools to ESI FT-ICR MS data interpretation (Koch et al., 2008; Rossello-Mora et al., 2008; Gougeon et al., 2009; Kujawinski et al., 2009). Appropriate use of these tools requires an estimate of underlying experimental variability. These estimates are rarely provided because repeat injections or experimental sample replicates are often non-existent. However, these estimates are critical for the appropriate design of experiments, from sample acquisition to data analysis and interpretation. Here, we have outlined a protocol to formulate a "best practice" method for MS analysis of environmental OM. The specific values of the instrument parameters optimized here provide a framework for evaluating these parameters on related systems. For example, we have shown that spectral variability associated with treatment replicates and with injection replicates are comparable in magnitude in terms of peak detection. As a result, in our facility, we now collect experimental replicates when possible and assess peak reproducibility through repeat injections of selected samples. With this type of assessment on different instruments as well as a formal instrument intercomparison, the field of DOM characterization using ESI FT-ICR MS will mature and will increasingly provide important insights into the role of DOM in global biogeochemical cycles.

Acknowledgements

We gratefully acknowledge the funding sources for this work: the National Science Foundation (Major Research Instrumentation program: OCE-0619608 (E.B.K.), the Chemical Oceanography program: CAREER-OCE-0529101 (E.B.K.) and OCE-0751897 (E.B.K. and S.J.G.)), the Marine Microbiology Initiative at the Gordon and Betty Moore Foundation (S.J.G.) and supporting funds from the WHOI Director of Research (E.B.K.). The manuscript was improved by constructive comments from two anonymous reviewers.

References

- Altieri, K.E., Turpin, B.J. and Seitzinger, S.P., 2009. Oligomers, organosulfates, and nitrooxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry. *Atmospheric Chemistry and Physics* 9, 2533-2542.
- Bhatia, M.P., Das, S.B., Charette, M.A., Longnecker, K. and Kujawinski, E.B., in press. Molecular characterization of dissolved organic matter associated with the Greenland ice sheet. *Geochimica Et Cosmochimica Acta*.
- Bresson, J.A., Anderson, G.A., Bruce, J.E. and Smith, R.D., 1998. Improved isotopic abundance measurements for high resolution Fourier transform ion cyclotron resonance mass spectra via time-domain data extraction. *Journal of American Society of Mass Spectrometry* 9, 799-804.
- Brown, T.L. and Rice, J.A., 2000. Effect of experimental parameters on the ESI FT-ICR mass spectrum of fulvic acid. *Analytical Chemistry* 72, 384-390.
- Connon, S.A. and Giovannoni, S.J., 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied and Environmental Microbiology* 68, 3878-3885.
- Dittmar, T., Koch, B., Hertkorn, N. and Kattner, G., 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography-Methods* 6, 230-235.
- Dittmar, T. and Koch, B.P., 2006. Thermogenic organic matter dissolved in the abyssal ocean. *Marine Chemistry* 102, 208-217.
- Fievre, A., Solouki, T., Marshall, A.G. and Cooper, W.T., 1997. High-resolution Fourier transform ion cyclotron resonance mass spectrometry of humic and fulvic acids by laser desorption / ionization and electrospray ionization. *Energy and Fuels* 11, 554-560.
- Gaspar, A., Kunenkov, E.V., Lock, R., Desor, M., Perminova, I. and Schmitt-Kopplin, P., 2009. Combined utilization of ion mobility and ultra-high-resolution mass spectrometry to identify multiply charged constituents in natural organic matter. *Rapid Communications in Mass Spectrometry* 23, 683-688.
- Gougeon, R.D., Lucio, M., Frommberger, M., Peyron, D., Chassagne, D., Alexandre, H., Feuillat, F., Voilley, A., Cayot, P., Gebefugi, I., Hertkorn, N. and Schmitt-Kopplin, P., 2009. The chemodiversity of wines can reveal a metabiogeography expression of cooperage oak wood. *Proceedings of the National Academy of Sciences USA* 106, 9174-9179.

Grannas, A.M., Hockaday, W.C., Hatcher, P.G., Thompson, L.G. and Mosley-Thompson, E., 2006. New revelations on the nature of organic matter in ice cores. *Journal of Geophysical Research* 111, D04304.

Hedges, J.I., 2002. Why dissolved organics matter. *Biogeochemistry of Marine Dissolved Organic Matter*. In: Hansell, D.A. and Carlson, C.A. (Eds.). San Diego, CA, Elsevier Science. pp. 1-34.

Kim, S., Kaplan, L.A. and Hatcher, P.G., 2006. Biodegradable dissolved organic matter in a temperate and a tropical stream determined from ultra-high resolution mass spectrometry. *Limnology and Oceanography* 51, 1054-1063.

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

Kim, S., Rodgers, R.P. and Marshall, A.G., 2006. Truly "exact" mass: Elemental composition can be determined uniquely from molecular mass measurement at similar to 0.1 mDa accuracy for molecules up to similar to 500 Da. *International Journal of Mass Spectrometry* 251, 260-265.

Kim, S., Simpson, A.J., Kujawinski, E.B., Freitas, M.A. and Hatcher, P.G., 2003. High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C18 solid phase disk. *Organic Geochemistry* 34, 1325-1335.

Koch, B.P. and Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry* 20, 926-932.

Koch, B.P., Dittmar, T., Witt, M. and Kattner, G., 2007. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Analytical Chemistry* 79, 1758-1763.

Koch, B.P., Ludwigowski, K.U., Kattner, G., Dittmar, T. and Witt, M., 2008. Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Marine Chemistry* 111, 233-241.

Koch, B.P., Witt, M., Engbrodt, R., Dittmar, T. and Kattner, G., 2005. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta* 69, 3299-3308.

Koprivnjak, J.F., Perdue, E.M. and Pfromm, P.H., 2006. Coupling reverse osmosis with electrodialysis to isolate natural organic matter from fresh waters. *Water Research* 40, 3385-3392.

Kujawinski, E.B. and Behn, M.D., 2006. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Analytical Chemistry* 78, 4363-4373.

Kujawinski, E.B., Del Vecchio, R., Blough, N.V., Klein, G.C. and Marshall, A.G., 2004. Probing molecular-level transformations of dissolved organic matter: Insights from electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry. *Marine Chemistry* 92, 23-37.

Kujawinski, E.B., Hatcher, P.G. and Freitas, M.A., 2002. High-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) of humic and fulvic acids: Improvements and comparisons. *Analytical Chemistry* 74, 413-419.

Kujawinski, E.B., Longnecker, K., Blough, N.V., Del Vecchio, R., Finlay, L., Kitner, J.B. and Giovannoni, S.J., 2009. Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta* 73, 4384-4399.

Kunenko, E.V., Kononikhin, A.S., Perminova, I.V., Hertkorn, N., Gaspar, A., Schmitt-Kopplin, P., Popov, I.A., Garmash, A.V. and Nikolaev, E.N., 2009. Total mass difference statistics algorithm: A new approach to identification of high-mass building blocks in electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter. *Analytical Chemistry* 81, 10106-10115.

Mantini, D., Petrucci, F., Pieragostino, D., Del Boccio, P., Di Nicola, M., Di Ilio, C., Federici, G., Sacchetta, P., Comani, S. and Urbani, A., 2007. LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinformatics* 8.

Marshall, A.G., Hendrickson, C.L. and Jackson, G.S., 1998. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* 17, 1-35.

McIntyre, C. and McRae, C., 2005. Proposed guidelines for sample preparation and ESI-MS analysis of humic substances to avoid self-esterification. *Organic Geochemistry* 36, 543-553.

Morales-Cid, G., Gebefugi, I., Kanawati, B., Harir, M., Hertkorn, N., Rossello-Mora, R. and Schmitt-Kopplin, P., 2009. Automated microextraction sample preparation coupled on-line to FT-ICR MS: application to desalting and concentration of river and marine dissolved organic matter. *Analytical and Bioanalytical Chemistry* 395, 797-807.

Parsons, H.M., Ekman, D.R., Collette, T.W. and Viant, M.R., 2009. Spectral relative standard deviation: a practical benchmark in metabolomics. *The Analyst* 134, 478-485.

Payne, T.G., Southam, A.D., Arvanitis, T.N. and Viant, M.R., 2009. A signal filtering method for improved quantification and noise discrimination in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-based metabolomics data. *Journal of the American Society for Mass Spectrometry* 20, 1087-1095.

Rappé, M.S., Connon, S.A., Vergin, K.L. and Giovannoni, S.J., 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630-633.

Reemtsma, T., These, A., Springer, A. and Linscheid, M., 2008. Differences in the molecular composition of fulvic acid size fractions detected by size-exclusion chromatography-on line Fourier transform ion cyclotron resonance (FTICR-) mass spectrometry. *Water Research* 42, 63-72.

Rossello-Mora, R., Lucio, M., Pena, A., Brito-Echeverria, J., Lopez-Lopez, A., Valens-Vadell, M., Frommberger, M., Anton, J. and Schmitt-Kopplin, P., 2008. Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *ISME Journal* 2, 242-253.

Rostad, C.E. and Leenheer, J.A., 2004. Factors that affect molecular weight distribution of Suwannee river fulvic acid as determined by electrospray ionization mass spectrometry. *Analytica Chimica Acta* 523, 269-278.

Schmidt, F., Elvert, M., Koch, B.P., Witt, M. and Hinrichs, K.-U., 2009. Molecular characterization of dissolved organic matter in pore water of continental shelf sediments. *Geochimica Et Cosmochimica Acta* 73, 3337-3358.

Sleighter, R.L. and Hatcher, P.G., 2008. Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Marine Chemistry* 110, 140-152.

Sleighter, R.L., McKee, G.A. and Hatcher, P.G., 2009. Direct Fourier transform mass spectral analysis of natural waters with low dissolved organic matter. *Organic Geochemistry* 40, 119-125.

Sleighter, R.L., McKee, G.A., Liu, Z. and Hatcher, P.G., 2008. Naturally present fatty acids as internal calibrants for Fourier transform mass spectra of dissolved organic matter. *Limnology and Oceanography-Methods* 6, 246-253.

Southam, A.D., Payne, T.G., Cooper, H.J., Arvanitis, T.N. and Viant, M.R., 2007. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Analytical Chemistry* 79, 4595-4602.

Stenson, A.C., 2008. Reversed-phase chromatography fractionation tailored to mass spectral characterization of humic substances. *Environmental Science & Technology* 42, 2060-2065.

Stenson, A.C., Landing, W.M., Marshall, A.G. and Cooper, W.T., 2002. Ionization and fragmentation of humic substances in electrospray ionization Fourier transform-ion cyclotron resonance mass spectrometry. *Analytical Chemistry* 74, 4397-4409.

Stenson, A.C., Marshall, A.G. and 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. *Analytical Chemistry* 75, 1275-1284.

Wozniak, A.S., Bauer, J.E., Sleighter, R.L., Dickhut, R.M. and Hatcher, P.G., 2008. Technical Note: Molecular characterization of aerosol-derived water soluble organic carbon using ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Atmospheric Chemistry and Physics* 8, 5099-5111.

Wu, Z., Rodgers, R.P. and Marshall, A.G., 2004. Two- and three-dimensional van Krevelen diagrams: A graphical analysis complementary to the Kendrick mass plot for sorting elemental compositions of complex organic mixtures based on ultrahigh-resolution broadband Fourier transform ion cyclotron resonance mass measurements. *Analytical Chemistry* 76, 2511-2516.

Table 1.

Recommendations for AGC values and minimum number of scans needed.

Scan type	Ion number (AGC)	Minimum # of scans
Full scan	1×10^6	200
Wide SIM (spectra width 100 m/z)	5×10^5	100 ($m/z < 500$) 150 ($m/z > 500$)
Narrow SIM (spectra width 30 m/z)	5×10^4	50

Table 2.

Variability in number of SRFA peaks (column 3), proportion of peaks shared (column 5), and relative standard deviation (RSD – column 6) for data collected at different mass resolution in positive ion mode or negative ion mode with full scan, narrow SIM, or wide SIM scan types.

	Scan type	Resolution	Mean # of peaks	Peaks	Peaks shared	RSD
			(SD)	shared (#)	(%)	
Positive Ion Mode	Full	100k	4090 (190)	2760	68	14
		200k	4000 (310)	3020	76	14
		400k	2970 (240)	2260	76	14
	Narrow SIM	100k	11750 (90)	8580	73	23
		200k	15850 (450)	9830	62	23
		400k	17340 (280)	11890	69	26
	Wide SIM	100k	11900 (180)	7650	64	14
		200k	15980 (370)	11700	73	16
		400k	18560 (580)	13700	74	17
Negative Ion Mode	Full	100k	3180 (60)	2550	80	6
		200k	3090 (70)	2660	86	7
		400k	2120 (40)	1860	87	8
	Narrow SIM	100k	10760 (470)	5770	54	26
		200k	11770 (570)	7840	67	25
		400k	7150 (180)	4860	68	25
	Wide SIM	100k	15050 (240)	9260	62	13
		200k	21970 (160)	15320	70	14
		400k	17650 (190)	12400	70	14

Table 3.

Effect of solvent composition (methanol/water) on number of SRFA peaks detected and proportion of shared peaks among three replicate injections.

Solvent	Run	Total peaks (#)	No. of peaks shared (%)
50% MeOH	1	2132	1856 (87%)
	2	2156	
	3	2080	
25% MeOH	1	5352	3157 (68%)
	2	3280	
	3	5253	
70% MeOH	1	2562	2187 (83%)
	2	2661	
	3	2692	

Table 4.

Assessment of impact of different strategies for sample replication on marine DOM composition (all data collected as the co-addition of 200 full scan transients at 400,000 resolution).

Ion mode	Treatment	Mean # of	No. of peaks	Peaks	RSD
		peaks (SD)	shared	shared (%)	
Positive	Treatment replicates	3380 (583)	2490	74	16
	Extraction replicates	3123 (363)	2144	69	29
Negative	Treatment replicates	2749 (434)	2083	76	36
	Extraction replicates	3017 (458)	2054	68	39

Figure legends

Fig. 1. Peaks observed in different wide SIM windows. Spectra were collected centered on either m/z 400 (circles) or 600 (diamonds). (A) Mean number of peaks per nominal mass is calculated over the entire window width for each spectrum (error bars represent 1 standard deviation). (B) Proportion (%) of peaks lost (Eq. 1) at the upper and lower edge of spectra centered on m/z 400 or 600.

Fig. 2. Edge effects in wide SIM mode. SRFA was run in negative ion mode with (A) data collection centered at m/z 400 or (B) data collection centered at m/z 600. The y-axis represents the ratio of peak heights in two wide SIM windows: 200 m/z (numerator) and 100 m/z (denominator). Values approximately equal to 1 (grey horizontal line) indicate no difference in peak height between the two windows. In contrast, increases in the ratio highlight the decrease in peak height in the smaller SIM window relative to the larger one. The gray vertical lines indicate recommended boundaries for wide SIM spectra with a width of 100 m/z .

Fig. 3. Untransformed SRFA transients at different AGC values. Data were collected in negative ion mode as full scan data. (A, D) have an AGC value of 1×10^5 , (B, E) have an AGC value of 1×10^6 and (C, F) have an AGC value of 3×10^6 . The top row of plots (A-C) displays the entire time duration of the transient, while the bottom row (D-E) focuses on the second half of the transient.

Fig. 4. Impact of AGC values on spectral parameters. Each plot shows the impact of different AGC values on the chosen parameter as a function of scans processed. Three AGC values were examined: 1×10^5 (squares), 1×10^6 (circles) and 3×10^6 (diamonds). Three parameters are shown: the total number of peaks found (A), the calculated noise level (B) and the proportion (%) of peaks shared (C). All data were collected in full scan negative ion mode.

Fig. 5. Impact of scan number on spectral parameters. Each plot shows the impact of scan number on the chosen parameter as a function of scans processed. Four parameters are shown: peak coverage (A), proportion (%) of peaks shared (B), noise level (C) and RSD (D). In (A), two methods of assessing peak coverage were used: total number of peaks found per replicate (circles) and total number of peaks shared between triplicate injections (diamonds). Error bars

(1 standard deviation) were calculated for (A). At most points, the error bars are smaller than the symbol. All data were collected in full scan negative ion mode.

Fig. 6. Reproducibility of peak detection and height in negative ion mode. For each plot, the peak heights in two runs are plotted against one another. The figure depicts two of the three SRFA replicate injections with the additional plots given in Supplementary Fig. 5. Symbols depict whether peaks were observed in all three replicates ($\bullet$), in two replicates ($\circ$) or in one replicate only ($\times$). A 1:1 line is provided for reference in all plots. All data were collected in negative ion mode. Full scan mode is depicted in panel (A), narrow SIM is depicted in panel (B) and wide SIM is depicted in panel (C).

Number of transients