Environmental Earth Sciences (2017) 76:828
https://doi.org/10.1007/s12665-017-7161-0

ORIGINAL ARTICLE

Profiling of dissolved organic compounds in the oil sands region using complimentary liquid–liquid extraction and ultrahigh resolution Fourier transform mass spectrometry

Yi Yi[1,2,6] · Jun Han[3] · S. Jean Birks[4] · Christoph H. Borchers[3,5] · John J. Gibson[1,2]

Received: 22 January 2017 / Accepted: 27 November 2017

Abstract

Understanding and characterizing organics in aquatic environments is a great challenge for environmental monitoring, especially for the oil sands industry due to the complexity and potential toxicity of dissolved organics in water. To date, significant efforts have been made in investigating the toxicity of naphthenic acids, although other compounds may also contribute to the toxicity of oil sands process-affected water (OSPW). Here, we present a case study showing a systematic approach for profiling the organic composition of OSPW and environmental water samples by concentrating and separating dissolved organics through complementary liquid–liquid extractions followed by positive- or negative-ion mode ultrahigh resolution mass detection. Our comparative investigation shows clear differences in the composition of dissolved organics (homologues particularly) not only between OSPW samples and environmental water samples, but also differences among oil sands operators. Sulfur-containing compounds (especially the SO_n *classes*) appear to have great potential to be used for evaluating the impact of OSPW, while our understanding of oxygen-only containing compounds should not be limited to O_2 (i.e., classic naphthenic acids), but rather can be broadened to include many other compound *classes* (for instance O_n, $n = 1$–9). Systematic profiling of water samples should be more widely implemented for monitoring the origin and transport of organics in aquatic ecosystems of the oil sands development region, northeastern Alberta, Canada.

Keywords Environmental forensics · Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) · Oil sands · Alberta

Introduction

Canadian oil sands contain over 168 billion barrels of proven unconventional petroleum reserves that are currently being extracted by either surface mining or in situ extraction at the rate of 1.8 million barrels per day (ERCB 2013). Surface-mined bitumen is processed via the Clark hot (79–93 °C) water method to separate oil from other constituents such as clay,

Yi Yi and Jun Han have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12665-017-7161-0) contains supplementary material, which is available to authorized users.

✉ Yi Yi
yiyi@uvic.ca

[1] InnoTech Alberta, Vancouver Island Technology Park, 3-4476 Markham St., Victoria, BC V8Z 7X8, Canada

[2] Department of Geography, University of Victoria, Victoria, BC V8W 3R4, Canada

[3] University of Victoria - Genome BC Proteomics Centre, Vancouver Island Technology Park, Victoria, BC V8Z 7X8, Canada

[4] InnoTech Alberta, Calgary, AB T2L 2A6, Canada

[5] Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8P 5C2, Canada

[6] Present Address: Environmental Monitoring and Science Division, Alberta Environment and Parks, Edmonton, Alberta T5J 5C6, Canada

Published online: 16 December 2017

sand, dissolved metals, and organic compounds (Clark 1944; Masliyah et al. 2004). The resulting oil sands process-affected water (OSPW) is stored within on-site tailings ponds. It has been estimated that approximately one trillion liters of OSPW are currently being retained in these manmade containment structures (Ferguson et al. 2009; Herman et al. 1994; Kannel and Gan 2012; Quagraine et al. 2005). As the volume of OSPW has accumulated, there is a growing concern about unintentional release of the OSPW into the Athabasca River through groundwater seepage or catastrophic failure of a containment dyke, which would have profound impacts on the aquatic ecosystems of northeastern Alberta and beyond (Ferguson et al. 2009; Giesy et al. 2010; Gosselin et al. 2010; Headley and McMartin 2004; Rowland et al. 2011; Woynilowicz et al. 2005).

The potential environmental impact of OSPW depends on the chemical composition of the dissolved constituents, which are not yet fully understood. Detecting and quantifying individual organics of OSPW on the molecular level is challenging, mainly due to the complexity of OSPW which contains thousands of chemicals from a wide range of compound classes. Ultrahigh resolution Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS), coupled with soft ionization techniques such as electrospray ionization (ESI) and atmospheric pressure photoionization (APPI), allows the characterization of complex mixtures at the molecular level, with the benefit of extremely high resolving power (> 1,000,000) and accurate mass measurements (ppb) (Marshall and Rodgers 2004, 2008). This technology has been widely used in analyzing natural organic matter of various environmental origins and has led to the recent resurgence in profiling and characterizing OSPW from the oil sands region (Barrow et al. 2010; Grewer et al. 2010; Headley et al. 2011, 2013a, b). Similar techniques but with lower resolving power are also under development (Bataineh et al. 2006; Martin et al. 2008; Frank et al. 2014; Huang et al. 2016).

Previous applications of FTICR-MS have primarily been focused on characterization of naphthenic acids (NAs) (Headley et al. 2013a). Negative-ion (−)ESI has emerged as the method most widely used for ionization (Headley et al. 2009). Compounds from (−)ESI profiling are dominated by oxygen-containing compounds (to a lesser degree by sulfur-containing compounds), making it particularly useful for the study of NAs (Grewer et al. 2010; Ross et al. 2012; Headley et al. 2013a; Nyakas et al. 2013; Pereira et al. 2013; Frank et al. 2014). Barrow et al. (2010) investigated a single OSPW sample via both negative-ion (−)ESI and positive-ion (+)ESI profiling. Distinct from (−)ESI mode, compounds detected in (+)ESI mode are mainly sulfur and nitrogen-containing compounds, potentially broadening the capacity of FTICR-MS profiling. Investigation of petroleum fractions via FTICR-MS also suggests that results from (−)ESI and (+)ESI are complementary to each other and provide a more comprehensive characterization of compounds in combination(Ávila et al. 2012). Recently, other ionization techniques, such as APPI, were also explored for detecting a broader range of compounds including nonpolar hydrocarbons in OSPW and environmental samples (Barrow et al. 2015).

There are also multiple choices for extracting and preparing samples for analytical procedures. The effects of pH and solvents have been discussed (Headley et al. 2007, 2013b; Barrow et al. 2015; Huang et al. 2016). Extraction methods for polar organics vary too. Solid phase extraction (SPE) has been successfully applied to extract polar organics from water samples (Headley et al. 2011; Reinardy et al. 2013; Frank et al. 2014; West et al. 2014; Barrow et al. 2015), while for studies requiring larger amounts of organics for multiple investigative purposes, liquid–liquid extraction (LLE) has been more commonly used (Grewer et al. 2010, Ross et al. 2012; Nyakas et al. 2013; Han et al. 2016). Overall, differences between the selected analytical approached can lead to significant variations in quantity and compositions of organic compounds in a given sample (Grewer et al. 2010; Yi et al. 2014).

Research activities at InnoTech Alberta and University of Victoria have been aimed at improving organic profiling in OSPW and environmental samples with a tractable consistency in sample preparation and analytical conditions (e.g., liquid–liquid extraction followed by ESI ionization, Gibson et al. 2011; Nyakas et al. 2013; Yi et al. 2015; Han et al. 2016). The objective of this study is to explore molecular profiling in both (−)ESI and (+)ESI modes in a complementary manner. Our results are expected to provide more compositional information on dissolved organics in OSPW at the molecular level, and to enable further characterization and differentiation of OSPW from the natural background for environmental monitoring and water resource management.

Materials and methods

Chemical and reagents

Water, methanol, ethyl acetate (EA), NaOH, formic acid, and ammonium hydroxide (≥ 25%) were LC/MS grade and obtained from Sigma-Aldrich (St. Louis, MO). Dichloromethane (DCM) and phosphoric acid were HPLC grade and were obtained from Sigma-Aldrich as well. Standard "ESI tuning mix" solution was purchased from Agilent Technologies (Palo Alto, CA).

Sample collection

Ten water samples were analyzed in this study. These samples were obtained from the oil sands region of northeastern Alberta, Canada, via multiple sampling campaigns coordinated by Alberta Environment in 2009. The OSPW samples were collected from tailings ponds at three mine sites by employees from InnoTech Alberta, with the assistance of on-site operational staff. Lake and seepage samples were grab samples collected by InnoTech Alberta and environmental

Table 1 Sample information analyzed in this study

Sample ID	Water type	Conductivity (μS/cm)	TOC mg/L	Notes
PAW001	OSPW sample	2530	43.7	Operator A
PAW002	OSPW sample	1535	58.6	Operator A
PAW003	OSPW sample	3880	62.7	Operator B
PAW004	OSPW sample	3780	63.0	Operator B
PAW007	OSPW sample	2325	N/A	Operator C
PAW008	OSPW sample	1001	N/A	Operator C
S07A	Environmental sample	4600	21.5	Seepages
S01A	Environmental sample	2670	10.9	Seepages
S09-239	Environmental sample	232	12.4	Lake
S09-092	Environmental sample	25	35.6	Lake

Water samples were generally categorized into two types: OSPW samples collected from tailings ponds versus environmental samples collected from natural water bodies. Routine measurements for bulk inorganic (i.e., conductivity) and organic content (i.e., TOC) are included

consultants during geophysical surveys (Gibson et al. 2013). During the sampling campaigns, field blank samples were collected with distilled water following identical sampling procedures for the purpose of QA/QC control.

Routine measurements of electrical conductivity and total organic carbon in water samples are listed in Table 1. Samples were refrigerated immediately after collection and were shipped to InnoTech Alberta, Victoria Lab, BC for processing and redistribution. In this investigation, we consider OSPW samples as water heavily impacted by the industrial processes of oil sands development, while samples from lakes and seepages are referred as environmental samples, representing local natural aquatic environments presumably with minimal industrial impact.

Liquid–liquid extraction

Two liquid–liquid extractions (LLEs) were performed using EA-DCM mixtures to enrich extractable organic compounds prior to FTICR-MS analysis. Extractions were performed under contrasting pH conditions (acidic vs basic) to separate dissolved organics in water samples into two fractions based on acidity/basicity. Figure 1 illustrates the schematic workflow of the two extractions. For the acidic extraction (i.e., LLE-1), 3-mL aliquots of individual water samples in 5 mL borosilicate glass tubes were added with 15 μL formic acids and extracted with 1 mL of EA-DCM (10:2, v/v). For the basic extraction (i.e., LLE-2), 3-mL aliquots of individual water samples were thoroughly mixed with 60 μL of 25% ammonium hydroxide solution, followed by extractions with 1 mL of EA-DCM (10:2, v/v). After a 30-s vortex mixing

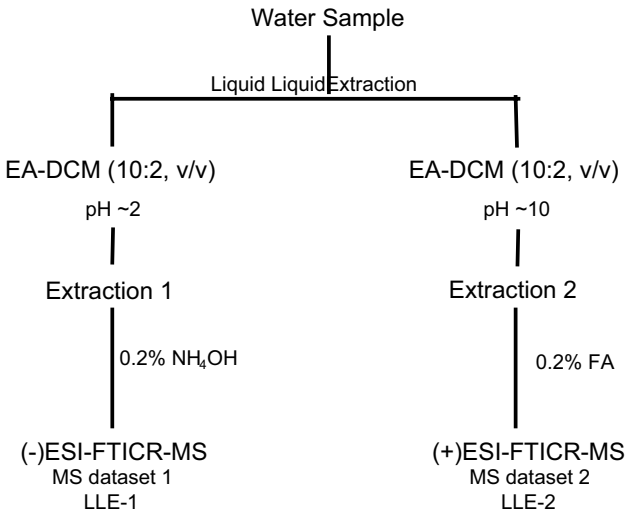

Fig. 1 Schematic workflow for liquid–liquid extractions to fractionate dissolved organics into two extractions, LLE-1 and LLE-2

and subsequent clarification by centrifugation at 4000 rpm for 10 min, the samples were placed at room temperature for 10 min. A 0.5-mL aliquot of the upper organic phases from each extraction was carefully transferred to a 3-mL borosilicate glass tube without disturbing the organic/aqueous interface and then dried under a gentle nitrogen flow at room temperature in a nitrogen evaporator. The residue of LLE-1 extraction was reconstituted in 0.5 mL of a mixed acetonitrile–water-25% ammonium hydroxide (60:39.2:0.8, v/v) solution containing 2 μL of the "ESI tuning mix" solution. The residue of LLE-2 extraction was reconstituted in 0.5 mL of a mixed acetonitrile–water-formic acid (60:39.8:0.2, v/v) solution containing 2 μL of the "ESI tuning mix" solution. After vortex mixing for 10 s and centrifugation at 10,000 rpm for 1 min in a microcentrifuge, the resulting clear solution was directly infused into the FTICR-MS and analyzed in the (−)ESI mode for the LLE-1 fraction and in the (+)ESI mode for the LLE-2 fraction, respectively.

FTICR-MS

All mass spectra were acquired on a Bruker 12-Tesla Apex-Qe hybrid quadrupole FTICR mass spectrometer (Billerica, MA). Each sample was infused into the MS instrument with a syringe pump at a flow rate of 3 μL/min. The instrument was operated in either (+) or (−)ESI mode within a scan range of *m/z* 150–1100. Each mass spectrum was recorded from an accumulation of 400 scans with the broadband acquisition and a data acquisition size of 1 MBps. Typical ESI–MS parameters were capillary electrospray voltage of 3800 V; spray shield voltage of 3500 V; source ion accumulation time of 0.1 s; and collision cell ion accumulation time of 0.2 s.

Data processing

Each set of the raw mass spectra from the OSPW and environmental samples were batch-processed using a custom VBA script, which has been described elsewhere (Han et al. 2008), within the instrument vendor's data processing software, *DataAnalysis*®. With this VBA script, the raw mass spectra were internally calibrated with the reference masses of the "ESI tuning mix" solution. Monoisotopic peaks corresponding to the isotopic distribution patterns on each spectrum were determined, and those with signal-to-noise ratios ≥ 10 were picked to generate a peak list which contained the individual *m/z* and the peak intensity values for each monoisotopic peak. The peak list from each mass spectrum was further batch-processed with another custom software program written within the LabView® development suite (National Instruments, Austin, TX) to align and combine monoisotopic masses across different samples within an allowable mass error of ± 1 ppm. The combined peak table from each mass spectral set was then exported and saved as a Microsoft Excel file which was used for subsequent data presentation and multivariate statistics. The peaks that showed higher peak intensities in the solvent blank than in any of the water samples were excluded and removed from the table.

Data analysis and presentation

The Kendrick mass defect (KMD) plots are used in this study to present and illustrate the mass distribution of the detected compounds in individual samples. The Kendrick mass and KMD values (Kendrick 1963; Marshall and Rodgers 2008) for each peak were calculated from the Système international d'unités (SI) mass [Eqs. (1) and (2), respectively]:

$$\text{Kendrick Mass} = \text{SI mass} \times 14/14.01565 \quad (1)$$

$$\text{Kendrick Mass defect} = (\text{nominal Kendrick mass} - \text{Kendrick mass}) \times 1000 \quad (2)$$

where nominal Kendrick mass is the Kendrick mass which is rounded to the nearest integer (Marshall and Rodgers 2008). In a KMD plot, a series of homologous compounds differing from each other by a repeating unit (e.g., a CH_2 group), yet having the same KMD value, can be easily identified as having common *y*-axis values. The identification of homologue series further assists with subsequent elemental composition assignments (Hughey et al. 2001). The relative contributions (RC), which are calculated as the intensity of an individual peak divided by the total intensity of all detected peaks, were used to characterize the organic compositions. It is important to note that the RC values cannot be treated as accurate measures of absolute concentration. This is because different compounds may have different ionization efficiencies in such complex mixture, which are generally referred as ion depression or matrix effect (Barrow et al. 2010). However, RC does generally reflect compositional changes in a complex mixture of organics (Hughey et al. 2007). Principle component analysis (PCA), using SIMCA-P + (V12.0, Umetrics AB Umeå, Sweden), was carried out based on the RC values in order to statistically compare compositional differences among water samples. Molecular formulas of the detected homologues were computed based upon the accurately measured monoisotopic masses with a custom algorithm written in the MATLAB R2011 suite (Yi et al. 2015). Heuristic filtering based on seven golden rules was applied in the algorithm in order to infer rationale molecular formulae with the assistance of identification of homologues (Kind and Fiehn 2007; Koch et al. 2007). The procedure for data presentation, compositional sorting, and graphical imaging was programmed and standardized using Matlab.

Results and discussion

Complementary liquid–liquid extraction

Water samples such as OSPW and environmental samples from the oil sands region are compositionally complex, containing dissolved organic compounds and inorganic salts (Table 1).The aim of this work is to apply direct infusion ESI-FTICR-MS for organic profiling of OSPW and to investigate both (−)ESI and (+)ESI results in a systematic manner to enable comparison with environmental samples from the region. Ten water samples, including six OSPW samples and four environmental samples were selected for this study. We designed a sample preparation procedure using two LLEs to fractionate dissolved organic compounds into two extractions under different pH conditions (i.e., acidic vs basic as shown in Fig. 1), which was followed by corresponding (+) and (−) ESI-FTICR-MS determination via direct infusion. The choice of (+)ESI or (−)ESI mode for downstream FTICR-MS detection was dependent on the nature of the extracted organic compounds. For the LLE-1 fraction that was performed at low pH conditions, it was anticipated that the extracted organics would best be characterized by analysis using (−)ESI–MS after alkylating with ammonium hydroxide. On the contrary, for the LLE-2 fraction that was performed at high pH conditions, it was anticipated that the extracted organic compounds would best be characterized by analysis using (+)ESI–MS after acidifying with formic acid. For both LLE-1 and LLE-2 fractions, the same organic solvent of EA/DCM (10:2, v/v) was used to exclude the potential solvent effects (Headley et al. 2007, 2013b; Huang et al. 2016). It is also important to note that we intentionally chose a solvent that is lighter than water, making it possible to transfer the organic extracts in an agitation-free manner.

In this way, organic compounds dissolved in water samples were extracted into two fractions on the basis of their acidity/basicity followed by subsequent FTICR-MS analysis in either (+) or (−) ion detection mode, while the inorganic salts were left in the aqueous phase, so as not to interfere with downstream FTICR-MS detections.

Thousands of compounds (i.e., peaks) were detected in the water samples (Fig. 2). Between 1486 and 2414 compounds were identified in LLE-1, while between 1095 and 3129 compounds were detected in LLE-2 (Table 2). The number of detected compounds was consistent with the range reported for (−)ESI detections (Nyakas et al. 2013; Yi et al. 2015), but significantly less than the number of compounds detected by APPI (between 6000 and 15,000 as reported by Barrow et al. 2015). Our approach does not profile the compositions of organics in the most comprehensive manner as compared to results presented by Barrow et al. (2015). But this study certainly improved the profiling results by ESI ionization, and the advantage of continuing to explore the ESI detection in parallel with the development of APPI method will be discussed in the following section.

It is very interesting to note that LLE-1 and LLE-2 appear to target different sets of compounds. Figure 2b demonstrates minor overlaps in detected peaks when raw spectra for LLE-1 and LLE-2 are compared in detail, peak by peak. Table 2 further lists the number of common peaks that were detected in both LLE-1 and LLE-2. From sample to sample, 64–100 compounds are found to be common in both fractions, which accounts for less than 6% of compounds in LLE-1 and less than 9% in LLE-2. The low percentage of overlapping compounds suggests that LLE effectively separates different organic compounds into two fractions.

Visualizations in KMD plots provide another line of evidence of the complementary nature of LLE-1 and LLE-2. Taking PAW001 as an example, apparent differences are observed between LLE-1 (Fig. 3a) and LLE-2 (Fig. 3b). Major peaks (RC > 2000 ppm) in LLE-1 are characterized by KMD values ~ 200 with an oval pattern in mass

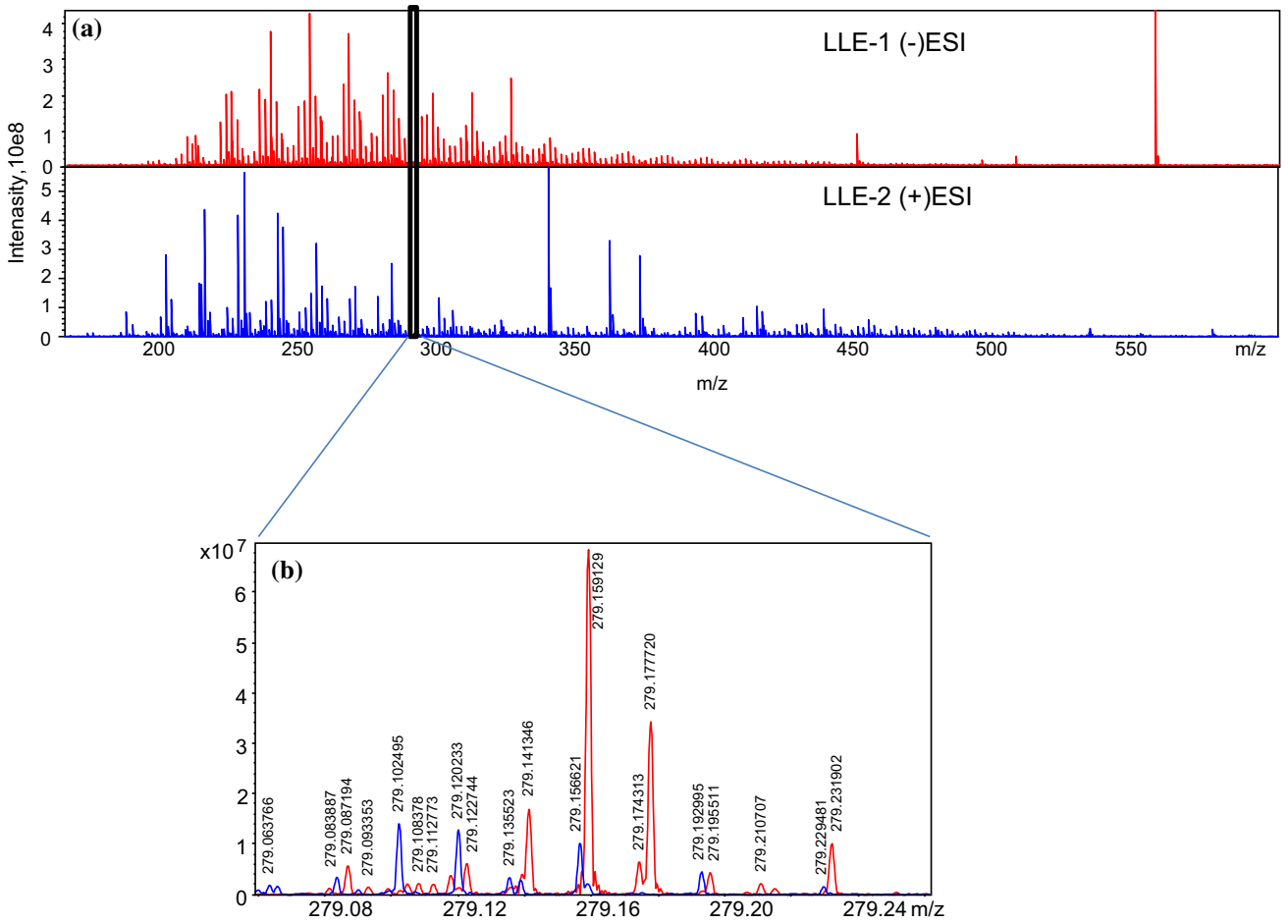

Fig. 2 Raw mass spectra acquired for LLE extracts, **a** broadband mass spectra for LLE-1 and LLE-2, respectively; **b** a comparative zoom-in (0.2 Da window) for both LLE-1 and LLE-2, demonstrating no overlaps of peaks acquired under different ESI modes [ESI(−) vs ESI(+)] and indicating complementary natures between LLE-1 and LLE-2

Table 2 Number of compounds (all peaks and homologues) detected in two LLE fractions

Fractions	PAW001	PAW002	PAW003	PAW004	PAW007	PAW008	S01A	S07A	SB09-092	SB09-239
LLE-1										
All peaks	2193	2143	2340	2414	2123	2317	1486	1620	1955	1753
Homologues	776	723	805	821	732	828	568	622	639	568
LLE-2										
All peaks	1691	1812	1777	1906	3129	2772	1095	1144	1203	1111
Homologues	588	601	647	672	1234	1095	273	290	306	292
Common peaks	66	68	98	108	74	64	86	81	100	73

The number of common peaks, detected in both fractions, is also presented

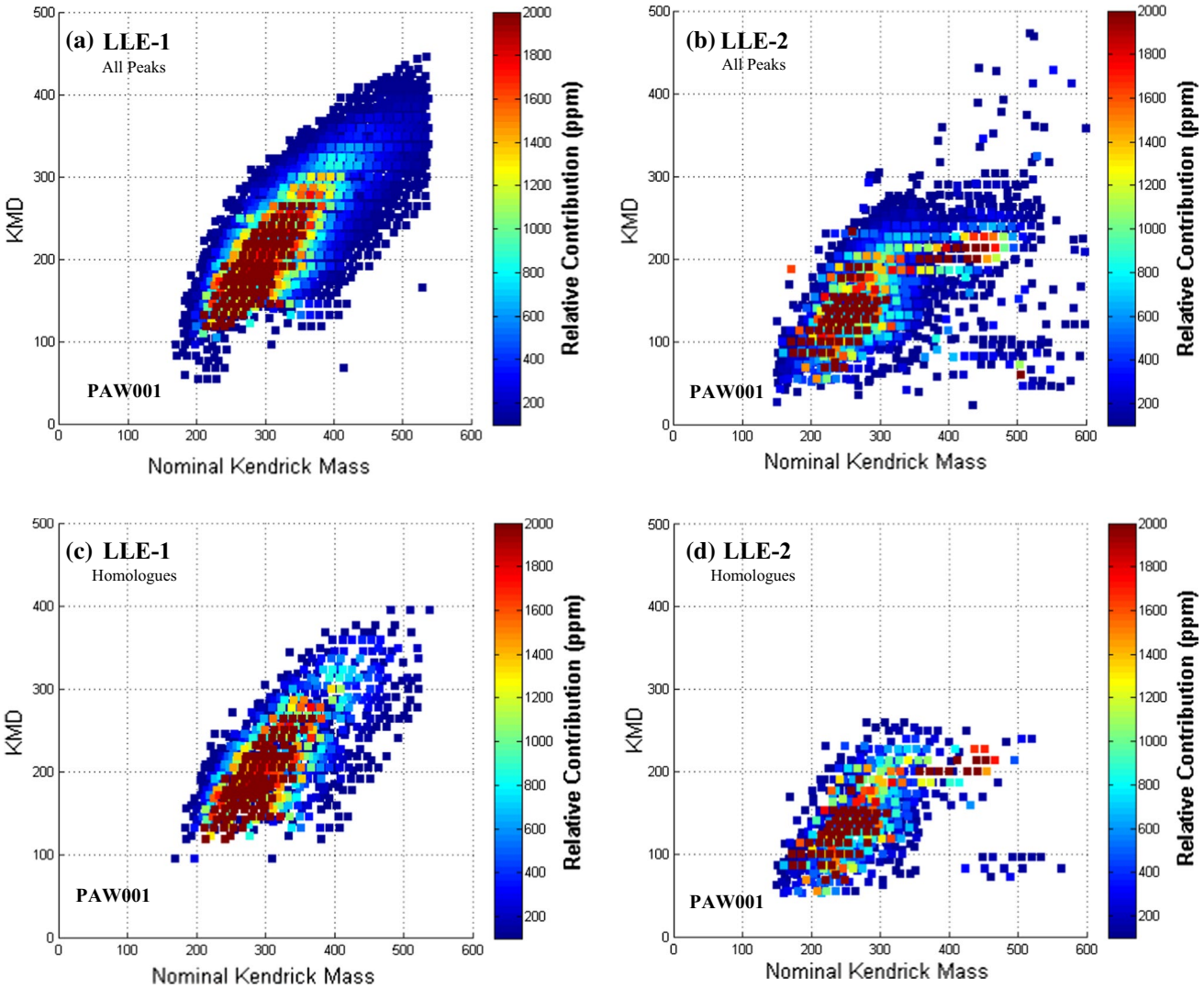

Fig. 3 Kendrick plots visualize composition of all detected peaks (**a**, **b**) and homologous series (**c**, **d**) in PAW001. A significant proportion of detected peaks is homologues, which show similar mass distribution patterns as demonstrated in KMD plots. **a** All peaks detected in LLE-1; **b** all peaks detected in LLE-2; **c** homologous series identified in LLE-1; **d** homologous series identified in LLE-2

distribution. In contrast, major peaks in LLE-2 are characterized by lower KMD values (between 100 and 200), demonstrating a sporadic pattern in mass distribution. Similar effects are observed for the homologues series (Fig. 3c, d), which will be discussed in the following sections. It is important to note that there are only 66 common compounds detected in both fractions, whereas 2193 compounds were detected in LLE-1 and 1691 compounds were detected in LLE-2 (Table 2). Collectively, 3813 compounds were detected in the combined LLE-1 and LLE-2 fractions for PAW001 as demonstrated in Fig. 3.

Nyakas et al. (2013) showed ultrahigh-performance liquid chromatography (UHPLC) fractionation prior to FTICR-MS was highly beneficial for detecting and characterizing more organic compounds in an OSPW sample. As illustrated, ~ 2200 to ~ 2800 peaks were detected in the OSPW sample if a single (or double) UHPLC was applied (Nyakas et al. 2013). However, the UHPLC fractionation was time-consuming. In this study, 2123–2414 compounds were detected in the LLE-1 fraction for OSPW samples (Table 2), which is comparable to the offline UHPLC fractionation. In addition, over 1500 compounds were detected in LLE-2 by (+)ESI mode. As shown here, profiling two LLE fractions could be a time-efficient alternative to UPHLC fractionations for comprehensively characterizing dissolved organic compounds in OSPW.

Abundant homologues in both LLE-1 and LLE-2 results are also a key feature in efforts to perform elemental assignments with accurate mass measurements (Hughey et al. 2001; Koch et al. 2007; Marshall and Rodgers 2008). Table 2 lists numbers of detected compounds, as well as numbers of homologues in both LLE-1 and LLE-2 fractions. Approximately 30% of peaks are found to be homologous series in this investigation, which is consistent with the previous study using LLE sample preparation (Yi et al. 2015), but a significant increase in proportion compared to those without LLE pre-treatment (~ 15% in Gibson et al. 2011). More importantly, mass distribution patterns for the homologous series resemble the distribution for all peaks. Taking PAW001 as an example, Fig. 3 illustrates the mass distribution of all detected compounds (2193 compounds for LLE-1 Fig. 3a; 1691 compounds for LLE-2 Fig. 3b) compared with the distribution of homologues alone (776 homologues for LLE-1 Fig. 3c and 588 homologues for LLE-2 Fig. 3d). We find the mass distributions of homologues are very similar to mass distributions for all detected compounds, suggesting profiles of homologues capture the main characteristics of dissolved organic compounds in OSPW. Furthermore, using homologue distributions only (Supplementary Figure 1) appears to capture notable differences between OSPW samples and environmental samples in both LLE-1 and LLE-2 fractions. Given that homologous series provide significant advantages via improved confidence in elemental assignments, we further focus discussions on a comparison between OSPW and environmental samples using homologues only (both LLE-1 and LLE-2).

Overall, the lack of overlap between LLE-1 and LLE-2 fractions, and high percentages of homologous series detected in the profiles are two advantages of using complementary LLE fractions for characterizing water samples such as OSPW that contain complex mixtures of organics.

Comparison of OSPW with environmental samples

Principle component analysis (PCA) is used as an exploratory tool to examine and illustrate differences among samples based upon hundreds of homologues. Score plots of the first two principal components (PC2 versus PC1) for LLE-1 (Fig. 4a) and LLE-2 (Fig. 4b) are shown for OSPW and environmental samples. As shown in the score plots, OSPW samples are clearly separated from environmental samples. In the LLE-1 (Fig. 4a), compositional differences in dissolved constituents explain 81% of the variance in the dataset (PC1, 63%; PC2, 18%). The distinction between environmental samples and OSPW is largely based on PC1, while the variance among the OSPW samples is mainly explained by PC2. OSPW samples from operators A and B (Table 1) are more similar to each other than to samples from operator C. Based on the score plot for LLE-2 (Fig. 4b), significant variance observed within the dataset (77%) could be explained by the first two PCs; PC1 and PC2 accounting for 57 and 20%, respectively, of the variation. Again, four environmental samples cluster closely, while the OSPW samples show relatively large variability in PC1 and PC2 space.

It is interesting to note that excellent grouping of OSPW samples from different operators is observed from the LLE-2 fraction. As shown in Fig. 4b, the samples from operators A and C can be distinguished based mainly on PC1, while operator B can be readily distinguished based on PC2. Although based on a limited set of samples ($n = 10$), the PCA analysis suggests the potential of LLE-2 for tracing OSPW from various operators.

Barrow et al. (2010) analyzed a single OSPW sample in both negative-ion mode and positive-ion mode with two ionization methods (ESI and APPI). Because more peaks were detected with APPI, Barrow et al. (2015) further investigated multiple samples (including OSPW and environmental samples) with (−)APPI and (+)APPI. The PCA analyses also demonstrate significant differences between environmental samples (such as river water) and OSPW. However, the clustering of OSPW between operators based upon APPI is not as distinct as those presented in ESI (Figure 5 & 9 in Barrow et al. 2015 vs Fig. 4 in this study). Although more peaks are profiled by APPI in both positive- and negative-ion modes, use of ESI remains a practical alternative, especially when

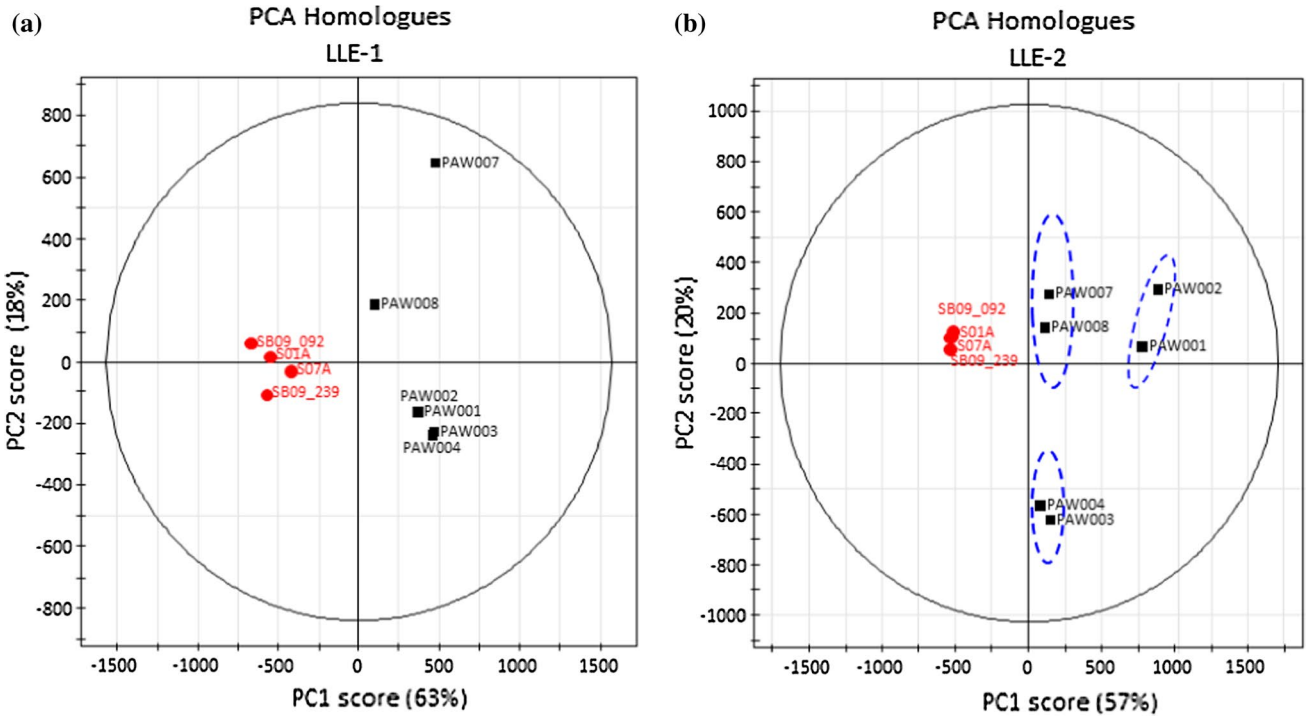

Fig. 4 PCA score plots of homologous series showing the compositional differences between OSPW and environmental samples. **a** PCA score plot of the results from LLE-1; **b** PCA score plot of results from LLE-2

ESI results (both LLE-1 and LLE-2) indicate the potential for fingerprinting individual operators.

Distinctions in elemental compositions

Given the statistical differences in organic compositions revealed by PCA, we further explore the elemental compositions of the detected organic compounds (homologues in particular) among the ten samples. Based upon the accurate masses determined by ultrahigh resolution FTICR-MS and an allowable mass tolerance of ± 1 ppm, it is possible to infer the rational molecular formulae for many of detected compounds, particularly for hundreds of the homologues identified by the KMD plots. To do this, the unique elemental compositions for one or a few low molecular weight members of a homologous series are computed and the assignments are then used to extrapolate the chemical formulae for the remaining compounds within the homologous series. During the molecular formula computation, several chemical rules were applied in the algorithm to filter out formulae not likely to be observed in nature. For example, the H/C and O/C ratios were limited to between 0.2–3.1 and 0–1.2, respectively, while the N/C and S/C ratios were restricted to the ranges of 0–1.3 and 0–0.8, respectively (Kind and Fiehn 2007). The number of rings plus double bond equivalents (DBE) was also examined. In this way, the unique molecular formulae for the majority of homologues (~ 80%) were generated.

The homologue compounds are categorized into five major heteroatom *classes*, including O_n, NO_n, N_2O_n, SO_n, and S_2O_n (where $n = 1$–12 in all cases), which is consistent with the classification of Headley et al. (2011). For example, SO_n refers to compounds that contain only one sulfur and multiple oxygen atoms varying between 1 and 12. Other compound *classes* are defined similarly. Taking PAW001 and SB09-239 as representatives of the OSPW and environmental samples, respectively, Fig. 5 shows a typical RC distribution for the LLE-1 fraction, and Fig. 6 shows typical results for the LLE-2 fraction. In general, dissolved organic compounds in all samples are found to be dominated by oxygen-containing classes (O_n) in LLE-1, with the appearance of one nitrogen (NO_n) and one sulfur (SO_n) classes, which is consistent with the findings of previous investigations (Barrow et al. 2010, 2015). The RCs of two nitrogen (N_2O_n) and two sulfur (S_2O_n) classes are minor if not absent.

The O_2 class is commonly regarded as "classic" naphthenic acids (Holowenko et al. 2002). Lee (1940) introduced the term "oxy-naphthenic acids" to describe compounds which are formed after mild oxidation of the classic naphthenic acids (Lee 1940). Other studies reported mono- and dioxide naphthenic acids (i.e., O_3 and O_4 classes) in extracts from OSPW (Han et al. 2009) and Grewer et al. (2010) urged investigators to broaden the measurements of potentially

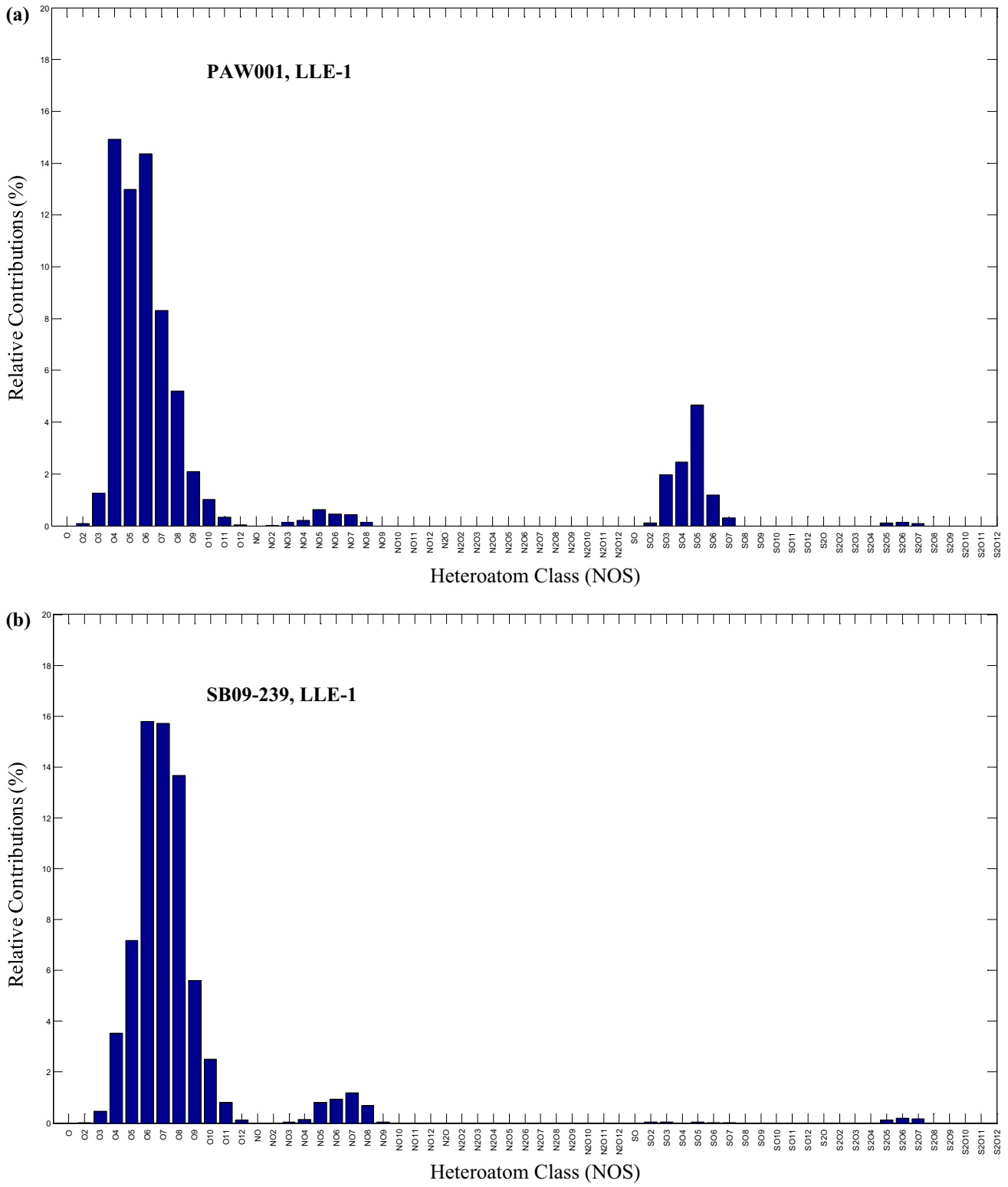

Fig. 5 Distribution of compound classes (homologous series) for representative profiles obtained from LLE-1. **a** PAW001 which represents OSPW; **b** SB09-239 which represents environmental samples

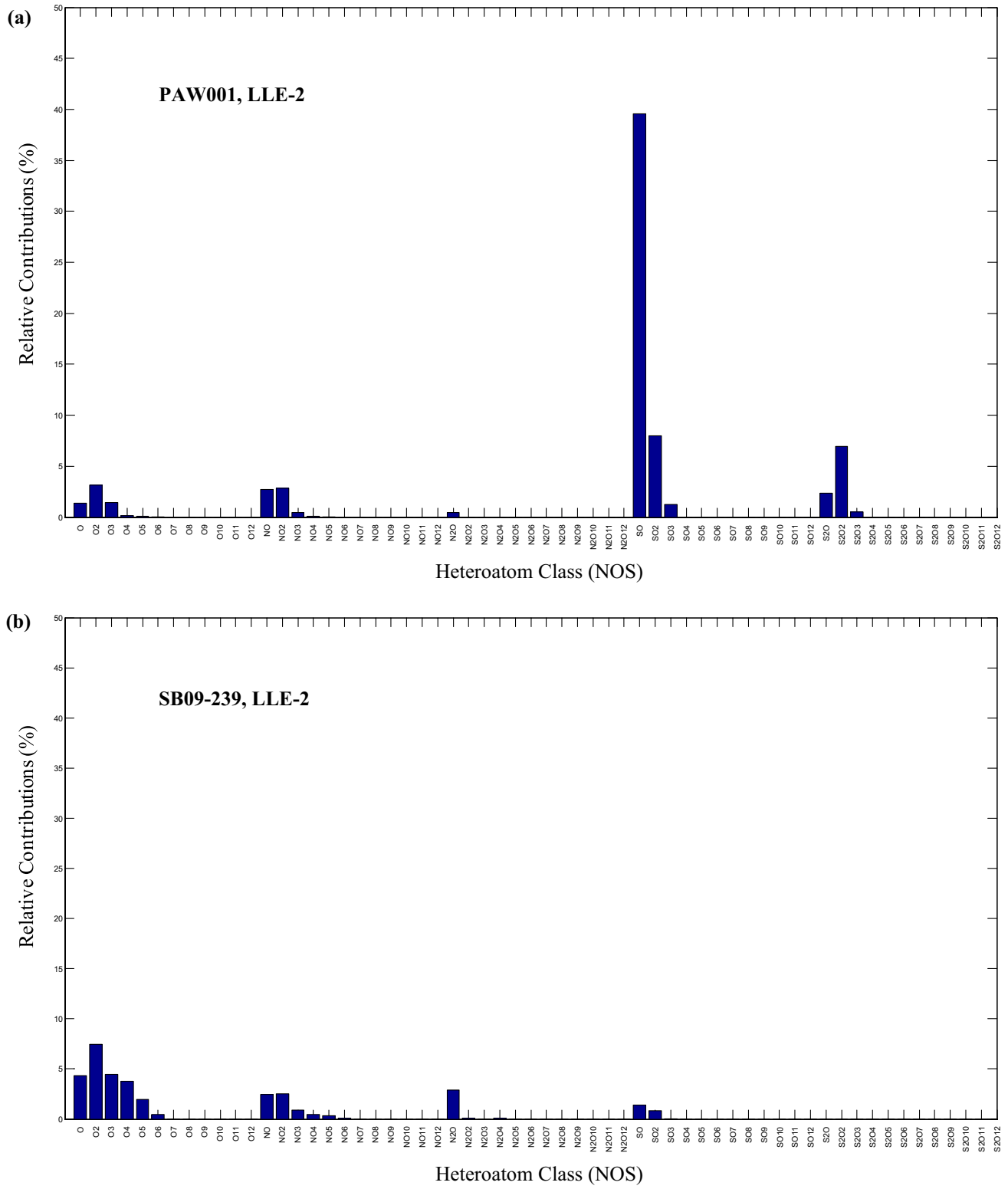

Fig. 6 Distribution of compound classes (homologue series) for representative profiles obtained from LLE-2. **a** PAW001 represents OSPW; **b** SB09-239 represents environmental samples

toxic compounds in OSPW beyond the O_2 classes. Other oxidized acids were included in their investigation, as they demonstrated for the O_3 to O_5 classes (Grewer et al. 2010). Our results also show a broad spectrum of oxygen-containing compounds detected in OSPW. Among O_n (n = 1–12) classes, the O_4–O_6 classes are dominant in PAW001 (Fig. 5a), while the major classes in environmental samples (e.g., SB09-239) are O_6–O_8 (Fig. 5b). This compositional difference between OSPW and environmental samples is consistent with the visual differences presented in the KMD plots (Supplementary Figure 1a&c), where O_6–O_8 classes are characterized by higher KMD values compared to the O_4 class (Hughey et al. 2001). Similar observations on distribution differences in O_n classes were noted by others (Headley et al. 2011; Barrow et al. 2015).

On the other hand, the relatively minor contribution of O_2 compounds to the overall O_n classes in this study (Fig. 5) is different from others (Headley et al. 2011; Barrow et al. 2015). According to previous studies, even in OSPW samples, relative contributions of O_2 to the overall O_n classes can vary significantly from sample to sample. Headley et al. (2011) reported considerably higher O_2 contribution (relative to O_4 contributions) in OSPW profiling. But Pereira et al. (2013) and Barrow et al. (2015) reported noticeable lower O_2 contributions as compared to the O_4 class. From the perspective of environmental samples, Yi et al. (2015) reported relative lower O_2 contribution in lakes comparing to river and snow samples. Certainly, the slight differences in instrumentation (FTICR-MS vs Orbitrap) and analytical procedures (such as pH, solvents and SPE vs LLE) could contribute to these differences (Grewer et al. 2010). Meanwhile, this also suggests that the profiling results, in most cases, at the current stage of development shall be interpreted qualitatively (or semiquantitatively at the best). Caution should be used in quantitative applications of these results (Frank et al. 2014; Yi et al. 2014). In the LLE-1 fraction, it is also important to note the distribution of SO_n and NO_n classes. Environmental samples (such as SB09-239) lack the SO_n classes (Fig. 5b), while OSPW samples (such as PAW001) clearly demonstrate the presence of SO_n compounds, particularly SO_3 to SO_6 (Fig. 5a). The strong presence of SO_n (n = 3–6) classes is confirmed by other (−)ESI (Headley et al. 2011) and (−)APPI investigations (Barrow et al. 2015). This sharp contrast (presence or absence) highlights the potential of SO_n classes in distinguishing OSPW from environmental samples, although the stability of these sulfur-containing classes—as well as which compounds should be targeted—needs further investigation.

NO_n compounds are relatively more abundant in environmental samples than in OSPW. Organic nitrogen compounds have been reported in natural aquatic environments, coming from precipitation and fog (Altieri et al. 2009; Mazzoleni et al. 2010). The high O/N ratios (≥ 3), including the dominance of NO_5 to NO_8 within the NO_n classes (Fig. 5b), are consistent with nitro (–NO_2) or nitrooxy (–ONO_2) groups in the natural environment. The high number of O atoms (more than in the –ONO_2 functional group) implies that these compounds may also possess other oxygenated functional groups.

From the LLE-2 fraction, detected compounds are not dominated by O_n classes, but rather show abundant sulfur-containing and nitrogen-containing compounds. SO_n (n = 1–3), S_2O_n (n = 1–2), NO_n (n = 1–2) and O_n (n = 1–3) are the major compound classes (Fig. 6). Unlike (+)APPI methods (Barrow et al. 2015), NSO_n classes are not detected by the (+)ESI method. One sulfur-containing compound (i.e., SO_n) is particularly abundant in the LLE-2 fraction, accounting for more than 30% of the RC in the sample from OSPW. The ubiquitousness and importance and of SO_n compounds can be further demonstrated when OSPW from individual operators is compared (Supplementary Figure 2). It is likely that variations in SO_n and S_2O_n compositions would largely explain the statistical distinctions between OSPW and environmental samples as demonstrated in Fig. 3b. This is consistent with the assertion suggested by Headley et al. (2011) that the ratios of the SO_n classes may be useful for environmental forensics. More rigorous investigation of the persistence and transformation of the signals should be carried out.

Here, we also reiterate the complementary nature of LLE-1 and LLE-2 from an elemental composition perspective. O_n and SO_n are classes found in both LLEs fractions. In our results, highly oxygenated compounds (n > 3) are usually found in the LLE-1 fraction (Fig. 5), while compounds with fewer numbers of oxygen atoms (n < 3) are usually found in the LLE-2 fraction (Fig. 6). This finding has not been reported in other studies. Although a much larger number of compounds and a broader range of classes were reported by Barrow et al. (2015), the similar distribution of O_n and SO_n classes in both positive- and negative-ion modes raises a concern that same compounds may have been detected in the different modes used. In this study, multiple lines of evidence, including the comparison of raw spectra (Fig. 2), visualization with Kendrick plots (Fig. 3) and examinations of compositional assignments (Figs. 5, 6), strongly suggest that different compounds were detected and profiled in LLE-1 and LLE-2 separately.

Concluding remarks

A complementary LLEs approach was developed to concentrate and extract the organic compounds in water samples into two fractions (LLE-1 and LLE-2), and each fraction was characterized by ESI-FTICR-MS. As shown in this study, over three thousands organic compounds (represented by

their unique masses) were detected in individual OSPW samples and the organic compounds detected in the two LLE fractions were characterized by little overlap and targeted different compound classes. As such, profiles obtained from two LLE fractions are complementary to each another, while potentially being comprehensive when integrated. High percentages of homologous compound series in both LLE-1 and LLE-2 have assisted in the determination of the elemental compositions of these dissolved organics based on the measured accurate masses. Thus, the complementary LLEs followed by ESI-FTICR-MS analysis provides a systematic method for profiling of the organic compositions in water samples, which is particularly suitable for environmental monitoring studies in the Athabasca oil sands region. PCA analysis of the homologue compositions in the two LLE fractions suggests statistically significant differences between the OSPW and environmental water samples, with clear separation observed among different operators in LLE-2 fractions. The results of compositional sorting based upon compound classes further reveal the importance of SO_n classes as well as distribution differences in O_n classes when comparing OSPW samples with environmental samples. SO_n ($n = 4$–6) classes detected in LLE-1 appear to have potential use for indicating the source of OSPW, while the proportional ratios of SO_n ($n = 1$–3) in LLE-2 might be a means to evaluate the impact of the OSPW. It is important that further compilation and evaluation of profiling results across the oil sand region shall be carried out using carefully documented analytical methods and procedures, while standardized approaches remain under development.

Acknowledgements Research funding for sample collection and FTICR-MS measurements was provided by InnoTech Alberta (formerly Alberta Innovates Technology Futures), Alberta Environment and OSRIN (Oil Sands Research Information Network). The University of Victoria—Genome BC Proteomics Centre—is supported by Genome Canada, Genome British Columbia and Genome Alberta through grants for "Science & Technology Innovation Centre (S&TIC)" in proteomics and partially through "The Metabolomics Innovation Centre (TMIC)" in metabolomics. We especially thank Drs. Roger Foxall and Preston McEachern for championing the application of new methods for OSPW tracing, and Dr. Carol E. Parker for helpful comments on an earlier version of this manuscript.

References

Altieri KE, Turpin BJ, Seitzinger SP (2009) Composition of dissolved organic nitrogen in continental precipitation investigated by ultra-high resolution FT-ICR mass spectrometry. Environ Sci Technol 43:6950–6955

Ávila BMF, Vaz BG, Pereira R, Gomes AO, Pereira RCL, Corilo YE, Simas RC, Nascimento HDL, Eberlin MN, Azevedo DA (2012) Comprehensive chemical composition of gas oil cuts using two-dimensional gas chromatography with time-of-flight mass spectrometry and electrospray ionization coupled to Fourier transform ion cyclotron resonance mass spectrometry. Energy Fuels 26:5069–5079

Barrow MP, Witt M, Headley JV, Peru KM (2010) Athabasca oil sands process water: characterization by atmospheric pressure photoionization and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem 82:3727–3735

Barrow MP, Peru KM, Fahlman B, Hewitt LM, Frank RA, Headley JV (2015) Beyond naphthenic acids: environmental screening of water from natural sources and the Athabasca oil sands industry using atmospheric pressure photoionization fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom 26:1508–1521

Bataineh M, Scott AC, Fedorak PM, Martin JW (2006) Capillary HPLC/QTOF-MS for characterizing complex naphthenic acid mixtures and their microbial transformation. Anal Chem 78:8354–8361

Clark KA (1944) Hot-water separation of Alberta bituminous sand. Trans Can Inst Min Metall 47:18

ERCB (2013) Alberta energy reserves 2012 and supply-demand outlook 2013–2022. ERCB, Calgary

Ferguson GP, Rudolph DL, Barker JF (2009) Hydrodynamics of a large oil sand tailings impoundment and related environmental implications. Can Geotech J 46:1446–1460

Frank RA, Roy JW, Bickerton G, Rowland SJ, Headley JV, Scarlett AG, West CE, Peru KM, Parrott JL, Conly FM, Hewitt LM (2014) Profiling oil sands mixtures from industrial developments and natural groundwaters for source identification. Environ Sci Technol 48:2660–2670

Gibson JJ, Birks SJ, Moncur M, Yi Y, Tattrie K, Jasechko S, Richardson K, Eby P (2011) Isotopic and geochemical tracers for fingerprinting process-affected waters in the oil sands industry: a pilot study. Oil Sands Research and Information Network, University of Alberta, Edmonton

Gibson JJ, Fennell J, Birks J, Yi Y, Moncur MC, Hansen B, Jasechko S (2013) Evidence of discharging saline formation water to the Athabasca River in the oil sands mining region, northern Alberta. Can J Earth Sci 50:1244–1257

Giesy JP, Anderson JC, Wiseman SB (2010) Alberta oil sands development. P Natl Acad Sci USA 107:951–952

Gosselin P, Hrudey SE, Naeth MA, Plourde A, Therrien R, van der Karaak G, Xu Z (2010) Environmental and health impacts of Canada's oil sands industry. Royal Society of Canada, Ottawa

Grewer DM, Young RF, Whittal RM, Fedorak PM (2010) Naphthenic acids and other acid-extractables in water samples from Alberta: what is being measured? Sci Total Environ 408:5997–6010

Han J, Danell RM, Patel JR, Gumerov DR, Scarlett CO, Speir JP, Parker CE, Rusyn I, Zeisel S, Borchers CH (2008) Towards high-throughput metabolomics using ultrahigh-field Fourier transform ion cyclotron resonance mass spectrometry. Metabolomics 4:128–140

Han XM, MacKinnon MD, Martin JW (2009) Estimating the in situ biodegradation of naphthenic acids in oil sands process waters by HPLC/HRMS. Chemosphere 76:63–70

Han J, Yi Y, Lin K, Birks J, Gibson J, Borchers CH (2016) Molecular profiling of naphthenic acids in technical mixtures and oil sands process water using a polar reversed-phase liquid chromatography-mass spectrometry. Electrophoresis 37(23–24):3089–3100

Headley JV, McMartin DW (2004) A review of the occurrence and fate of naphthenic acids in aquatic environments. J Environ Sci Health A 39:1989–2010

Headley JV, Peru KM, Barrow MP, Derrick PJ (2007) Characterization of naphthenic acids from Athabasca oil sands using electrospray ionization: the significant influence of solvents. Anal Chem 79:6222–6229

Headley JV, Peru KM, Barrow MP (2009) Mass spectrometric characterization of naphthenic acids in environmental samples: a review. Mass Spectrom Rev 28:121–134

Headley JV, Barrow MP, Peru KM, Fahlman B, Frank RA, Bickerton G, McMaster ME, Parrott J, Hewitt LM (2011) Preliminary fingerprinting of Athabasca oil sands polar organics in environmental samples using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom 25:1899–1909

Headley JV, Peru KM, Mohamed MH, Frank RA, Martin JW, Hazewinkel RRO, Humphries D, Gurprasad NP, Hewitt LM, Muir DCG, Lindeman D, Strub R, Young RF, Grewer DM, Whittal RM, Fedorak PM, Birkholz DA, Hindle R, Reisdorph R, Wang X, Kasperski KL, Hamilton C, Woudneh M, Wang G, Loescher B, Farwell A, Dixon DG, Ross M, Pereira ADS, King E, Barrow MP, Fahlman B, Bailey J, McMartin DW, Borchers CH, Ryan CH, Toor NS, Gillis HM, Zuin L, Bickerton G, McMaster M, Sverko E, Shang D, Wilson LD, Wrona FJ (2013a) Chemical fingerprinting of naphthenic acids and oil sands process waters—a review of analytical methods for environmental samples. J Environ Sci Health A 48:1145–1163

Headley JV, Peru KM, Fahlman B, Colodey A, McMartin DW (2013b) Selective solvent extraction and characterization of the acid extractable fraction of Athabasca oils sands process water by Orbitrap mass spectrometry. Int J Mass Spectrom 345–347:104–108

Herman DC, Fedorak PM, Mackinnon MD, Costerton JW (1994) Biodegradation of naphthenic acids by microbial-populations indigenous to oil sands tailings. Can J Microbiol 40:467–477

Holowenko FM, MacKinnon MD, Fedorak PM (2002) Characterization of naphthenic acids in oil sands wastewaters by gas chromatography-mass spectrometry. Water Res 36:2843–2855

Huang RF, McPhedran KN, Sun N, Chelme-Ayala P, Gamla El-Din M (2016) Investigation of the impact of organic solvent type and solution pH on the extraction efficiency of naphthenic acids from oil sands process-affected water. Chemosphere 146:472–477

Hughey CA, Hendrickson CL, Rodgers RP, Marshall AG, Qian KN (2001) Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal Chem 73:4676–4681

Hughey CA, Galasso SA, Zumberge JE (2007) Detailed compositional comparison of acidic NSO compounds in biodegraded reservoir and surface crude oils by negative ion electrospray Fourier transform ion cyclotron resonance mass spectrometry. Fuel 86:758–768

Kannel PR, Gan TY (2012) Naphthenic acids degradation and toxicity mitigation in tailings wastewater systems and aquatic environments: a review. J Environ Sci Health A 47:1–21

Kendrick E (1963) A, mass scale based on Ch2 = 14.0000 for high resolution mass spectrometry of organic compounds. Anal Chem 35:2146–2154

Kind T, Fiehn O (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinform 8:105

Koch BP, Dittmar T, Witt M, Kattner G (2007) Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. Anal Chem 79:1758–1763

Lee DD (1940) Thermosetting resin reaction product of furtural with an oxy-naphthenic acid. In: Office USP (ed) Standard Oil Co, p 4

Marshall AG, Rodgers RP (2004) Petroleomics: the next grand challenge for chemical analysis. Acc Chem Res 37:53–59

Marshall AG, Rodgers RP (2008) Petroleomics: chemistry of the underworld. P Natl Acad Sci USA 105:18090–18095

Martin JW, Han XM, Peru KM, Headley JV (2008) Comparison of high- and low-resolution electrospray ionization mass spectrometry for the analysis of naphthenic acid mixtures in oil sands process water. Rapid Commun Mass Spectrom 22:1919–1924

Masliyah J, Zhou ZJ, Xu ZH, Czarnecki J, Hamza H (2004) Understanding water-based bitumen extraction from Athabasca oil sands. Can J Chem Eng 82:628–654

Mazzoleni LR, Ehrmann BM, Shen XH, Marshall AG, Collett JL (2010) Water-soluble atmospheric organic matter in fog: exact masses and chemical formula identification by ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry. Environ Sci Technol 44:3690–3697

Nyakas A, Han J, Peru KM, Headley JV, Borchers CH (2013) Comprehensive analysis of oil sands processed water by direct-infusion fourier-transform ion cyclotron resonance mass spectrometry with and without offline UHPLC sample prefractionation. Environ Sci Technol 47:4471–4479

Pereira AS, Bhattacharjee S, Martin JW (2013) Characterization of oil sands process-affected waters by liquid-Chromatography Orbitrap mass spectrometry. Environ Sci Technol 47:5504–5513

Quagraine EK, Peterson HG, Headley JV (2005) In situ bioremediation of naphthenic acids contaminated tailing pond waters in the Athabasca oil sands region-demonstrated field studies and plausible options: a review. J Environ Sci Health A 40:685–722

Reinardy HC, Scarlett AG, Henry TB, West CE, Hewitt LM, Frank RA, Rowland SJ (2013) Aromatic naphthenic acids in oil sands process-affected water, resolved by GCxGC-MS, only weakly induce the gene for vitellogenin production in zebrafish (*Danio rerio*) larvae. Environ Sci Technol 47:6614–6620

Ross MS, Pereira AS, Fonnell J, Davies M, Johnson J, Sliva L, Martin JW (2012) Quantitative and qualitative analysis of naphthenic acids in natural waters surrounding the Canadian oil sands industry. Environ Sci Technol 46:12796–12805

Rowland SJ, Scarlett AG, Jones D, West CE, Frank RA (2011) Diamonds in the rough: identification of individual naphthenic acids in oil sands process water. Environ Sci Technol 45:3154–3159

West CE, Scarlett AG, Tonkin A, O'Carroll-Fitzpatrick D, Pureveen J, Tegelaar E, Gieleciak R, Hager D, Petersen K, Tollefsen KE, Rowland SJ (2014) Diaromatic sulphur-containing 'naphthenic' acids in process waters. Water Res 51:206–215

Woynilowicz D, Severson-Baker C, Raynolds M (2005) Oil sands fever. The environmental implications of Canada's oil sands rush. The Pembina Institute, Calgary

Yi Y, Gibson JJ, Birks S, Han J, Borchers CH (2014) Comments on "Profiling oil sands mixtures from industrial developments and natural groundwaters for source identification". Environ Sci Technol 48:11013–11014. https://doi.org/10.1021/es503498p

Yi Y, Birks SJ, Cho S, Gibson JJ (2015) Characterization of organic composition in snow and surface waters in the Athabasca Oil Sands Region, using ultrahigh resolution Fourier transform mass spectrometry. Sci Total Environ 518–518:148–158