Supporting information

Molecular characterization of water-soluble humic-like substances in smoke particles emitted from combustion of biomass materials and coal using ultrahigh-resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry

Jianzhong Song[a], Meiju Li[a,c], Bin Jiang[a], Siye Wei[a,c], Xingjun Fan[a,b], Ping'an Peng[a]

[a] State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, P. R. China

[b] College of Resource and Environment, Anhui Science and Technology University, Anhui 233100, P. R. China

[c] Graduate School of Chinese Academy of Sciences, Beijing 100049, P. R. China

* Corresponding author e-mail: songjzh@gig.ac.cn

Samples, Data processing, 2 Tables, 6 Figures, 15 pages in total

Contents:

10. Figure S6. Double bond equivalent (DBE) vs C number for all the CHOS compounds of the three BB-smoke HULIS samples. The color bar and marker size denote the number of O atoms and the peak intensities of the compounds. The formulae in the box a are $C_{17}H_{27}O_3S_1$, $C_{18}H_{29}O_3S_1$, and $C_{19}H_{31}O_3S_1$, corresponding to alkane sulfonates that are known to exist widely in the atmospheric environment and the background pollution during FT-ICR MS measurement.

1. Samples

In this study, four types of smoke particle samples were collected from the combustion of three biomass materials (i.e., corn straw, rice straw, and pine branch) and a lignite coal. Rice straw and corn straw were collected at Liuzhou, Guangxi Zhuang Autonomous Region, China. The pine branch was obtained at Qingyuan, Guangdong province, China. The lignite coal was obtained at Pingdingshan, Henan province, China.

After collected, corn straw, rice straw, and pine branch were firstly washed with tap water three times to remove dust and then air-dried for 7 days. The coal sample was also air-dried for 7 days. Finally these four samples were stored in bags until used for combustion.

2. Data processing

The mass spectra were processed using custom software, which has been described elsewhere.[1-3] All the ions with relative abundance greater than 6 times the standard deviation of the baseline noise and in the m/z range from 150 to 800 were exported. Subsequently, all mathematically possible formulae for these ions were calculated using a mass tolerance of ± 1 ppm. The general formula $C_cH_hO_oN_nS_s$ was used to indicate the assigned compounds, in which the C, H, O, N, S indicate carbon, hydrogen, oxygen, nitrogen and sulfur respectively and the c, h, o, n, s represent the number of atoms of carbon, hydrogen, oxygen, nitrogen and sulfur, respectively.

The calculated formulae were further characterized by the number of rings plus

74 double bonds (i.e., the double bond equivalents (DBE)) calculated from the elemental
75 composition $C_cH_hO_oN_nS_s$: [4, 5]

76 $$DBE=(2c+2-h+n)/2$$

77 DBE reflects the degree of unsaturation (hydrogen deficiency) in a given compound.[6]
78 All the calculated formulae with DBE < 0 and >20 and that disobey the nitrogen rules
79 for the even electron ions were excluded from the lists.

80 Each molecular formula was also estimated using the modified aromaticity index
81 (AI_{mod}) based on the system proposed by Koch and Dittmar:[7]

82 $$AI_{mod} = (1+c-0.5o-s-0.5h)/(c-0.5o-s-n)$$

83 Here, formulae are conservatively classified as non aromatic (AI < 0.5), aromatic (AI >
84 0.5) and condensed aromatic (AI ≥ 0.67). Because of heteroatoms such as oxygen,
85 sulfur and nitrogen were taken into account as potential contributors, this ratio is a
86 suitable parameter for the identification of aromatic structures in natural organic
87 matter.

88 It is noted that each molecule should present with different intensity, the overall
89 properties of the four HULIS were done with relative abundance weighted. The
90 relative abundance weighted elemental ratios, organic mass to organic carbon ratios,
91 DBE, and AI_{mod} were calculated from the intensity (Int) of each assigned peak (i)
92 using the following equations: [5, 8]

93 $$MW_w = \Sigma(MW_i * Int_i)/\Sigma Int_i$$

94 $$O/C_w = \Sigma(O/C_i * Int_i)/\Sigma Int_i$$

95 $$H/C_w = \Sigma(H/C_i * Int_i)/\Sigma Int_i$$

$S/C_w = \Sigma(S/C_i * Int_i)/ \Sigma Int_i$

$O/N_w = \Sigma(O/N_i * Int_i)/ \Sigma Int_i$

$O/S_w = \Sigma(O/S_i * Int_i)/ \Sigma Int_i$

$OM/OC_w = \Sigma(OM/OC_i * Int_i)/ \Sigma Int_i$

$DBE_w = \Sigma(DBE_i * Int_i)/ \Sigma Int_i$

$AI_{mod,w} = \Sigma(AI_{mod,i} * Int_i)/ \Sigma Int_i$

for elemental compositions of $C_cH_hN_nO_oS_s$, where Int_i is the intensity for each individual molecular formula, *i*

Reference

(1) Shi, Q.; Pan, N.; Long, H. Y.; Cui, D. C.; Guo, X. F.; Long, Y. H.; Chung, K. H.; Zhao, S. Q.; Xu, C. M.; Hsu, C. S., Characterization of Middle-Temperature Gasification Coal Tar. Part 3: Molecular Composition of Acidic Compounds. Energ Fuel 2013, 27, (1), 108-117.

(2) Jiang, B.; Liang, Y. M.; Xu, C. M.; Zhang, J. Y.; Hu, M.; Shi, Q., Polycyclic Aromatic Hydrocarbons (PAHs) in Ambient Aerosols from Beijing: Characterization of Low Volatile PAHs by Positive-Ion Atmospheric Pressure Photoionization (APPI) Coupled with Fourier Transform Ion Cyclotron Resonance. Environ Sci Technol 2014, 48, (9), 4716-4723.

(3) Li, Y. Y.; Fang, Z.; He, C.; Zhang, Y. H.; Xu, C. M.; Chung, K. H.; Shi, Q., Molecular Characterization and Transformation of Dissolved Organic Matter in Refinery Wastewater from Water Treatment Processes: Characterization by Fourier

Transform Ion Cyclotron Resonance Mass Spectrometry. *Energ Fuel* 2015, 29, (11), 6956-6963.

(4) Lin, P.; Rincon, A. G.; Kalberer, M.; Yu, J. Z., Elemental Composition of HULIS in the Pearl River Delta Region, China: Results Inferred from Positive and Negative Electrospray High Resolution Mass Spectrometric Data. *Environ Sci Technol* 2012, *46*, (14), 7454-7462.

(5) Zhao, Y.; Hallar, A. G.; Mazzoleni, L. R., Atmospheric organic matter in clouds: exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry. *Atmos Chem Phys* 2013, *13*, (24), 12343-12362.

(6) Hockaday, W. C.; Grannas, A. M.; Kim, S.; Hatcher, P. G., Direct molecular evidence for the degradation and mobility of black carbon in soils from ultrahigh-resolution mass spectral analysis of dissolved organic matter from a fire-impacted forest soil. *Org Geochem* 2006, *37*, (4), 501-510.

(7) Koch, B. P.; Dittmar, T., From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun Mass Sp* 2006, *20*, (5), 926-932.

(8) Sleighter, R. L.; Hatcher, P. G., 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. *Mar Chem* 2008, *110*, (3-4), 140-152.

Table S1. Percentage (%) of different compound category in total unique chemical formulae.

Formulae subgroups	Corn	Rice	Pine	Coal
CHO	57.8	14.6	56.1	6.6
CHON	30.2	81.2	30.7	15.3
CHOS	9.0	1.5	8.4	43.3
CHONS	3.0	2.7	4.8	34.8

Table S2. Number of formulae in each compound category and the average values of elemental ratios, molecular weight, double-bond equivalents (DBEs) and aromaticity index (AI_{mod}).

Samples	Elemental compositions	Number of formulas	MW_w	H/C_w	O/C_w	N/C_w	S/C_w	O/N_w	O/S_w	OM/OC_w	DBE_w	DBE/C_w	$AI_{mod,w}$
Corn	CHO	1514	296	1.33	0.41	-	-	-	-	1.66	6.23	0.41	0.31
	CHON	1640	335	1.13	0.43	0.085	-	5.83	-	1.76	8.59	0.55	0.44
	CHOS	227	328	1.48	0.41	-	0.072	-	5.57	1.86	4.83	0.33	0.13
	CHONS	146	311	1.21	0.57	0.087	0.087	6.67	6.67	2.19	6.17	0.53	0.27
Rice	CHO	1733	311	1.13	0.40	-	-	-	-	1.62	7.58	0.47	0.38
	CHON	2405	343	1.10	0.45	0.093	-	5.66	-	1.79	9.00	0.57	0.46
	CHOS	339	300	1.28	0.46	-	0.081	-	5.65	1.94	5.07	0.40	0.18
	CHONS	248	330	1.21	0.55	0.081	0.081	6.91	6.91	2.15	6.17	0.48	0.22
Pine	CHO	2296	337	1.30	0.21	-	-	-	-	1.39	8.02	0.40	0.35
	CHON	1527	365	1.07	0.42	0.082	-	5.99	-	1.74	9.68	0.57	0.45
	CHOS	522	342	1.41	0.41	-	0.068	-	6.03	1.84	5.62	0.36	0.18
	CHONS	276	328	1.24	0.61	0.086	0.086	7.19	7.19	2.24	6.21	0.51	0.23
Coal	CHO	918	270	1.13	0.31	-	-	-	-	1.51	7.48	0.51	0.45
	CHON	610	295	1.01	0.27	0.070	-	4.15	-	1.52	9.57	0.60	0.56
	CHOS	669	277	1.26	0.42	-	0.085	-	5.02	1.89	5.80	0.46	0.31
	CHONS	430	304	1.09	0.44	0.082	0.082	5.31	5.31	1.98	7.52	0.58	0.45

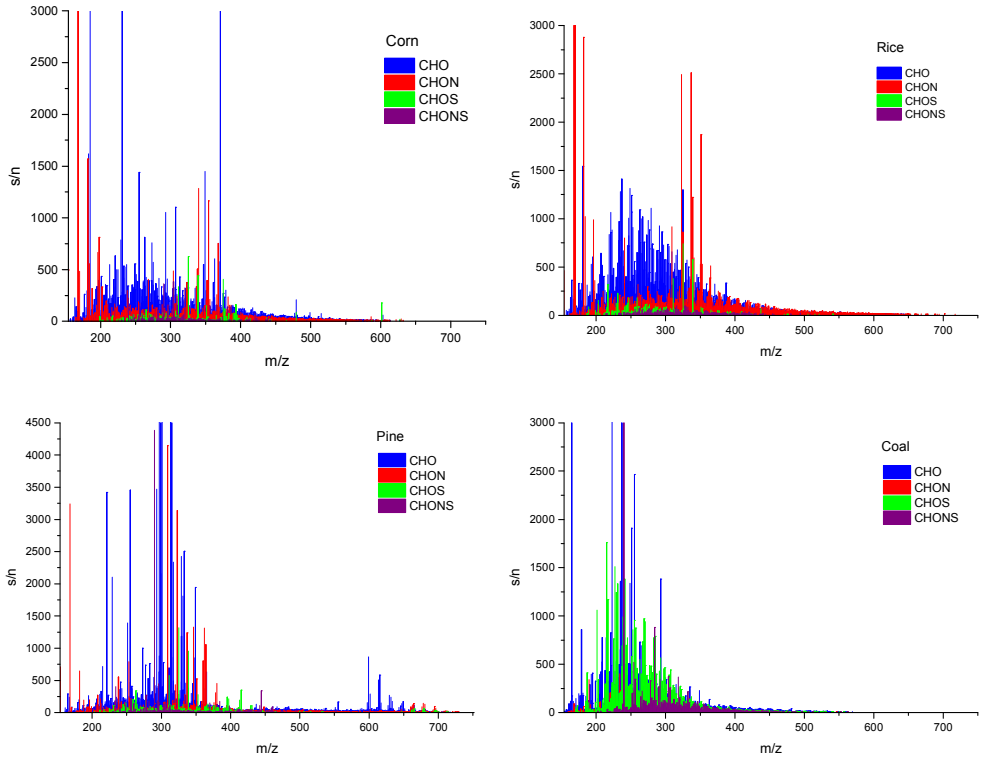

Figure S1. Reconstructed Fourier-transform ion cyclotron resonance (FT-ICR) mass spectra for the four smoke HULIS samples.

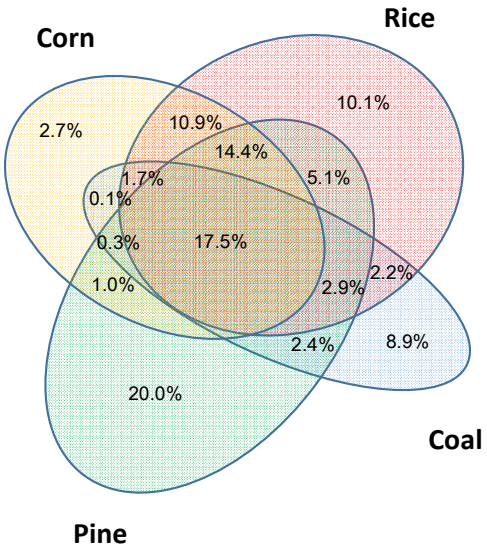

Figure S2. Venn diagrams showing the relative distributions of all molecular formulae (7291 total) present in the four smoke HULIS samples. Percentages in areas of overlap are percentages of molecular formulae that appear in both, three or all four of those samples. Percentages in areas with no overlap are unique to that individual sample.

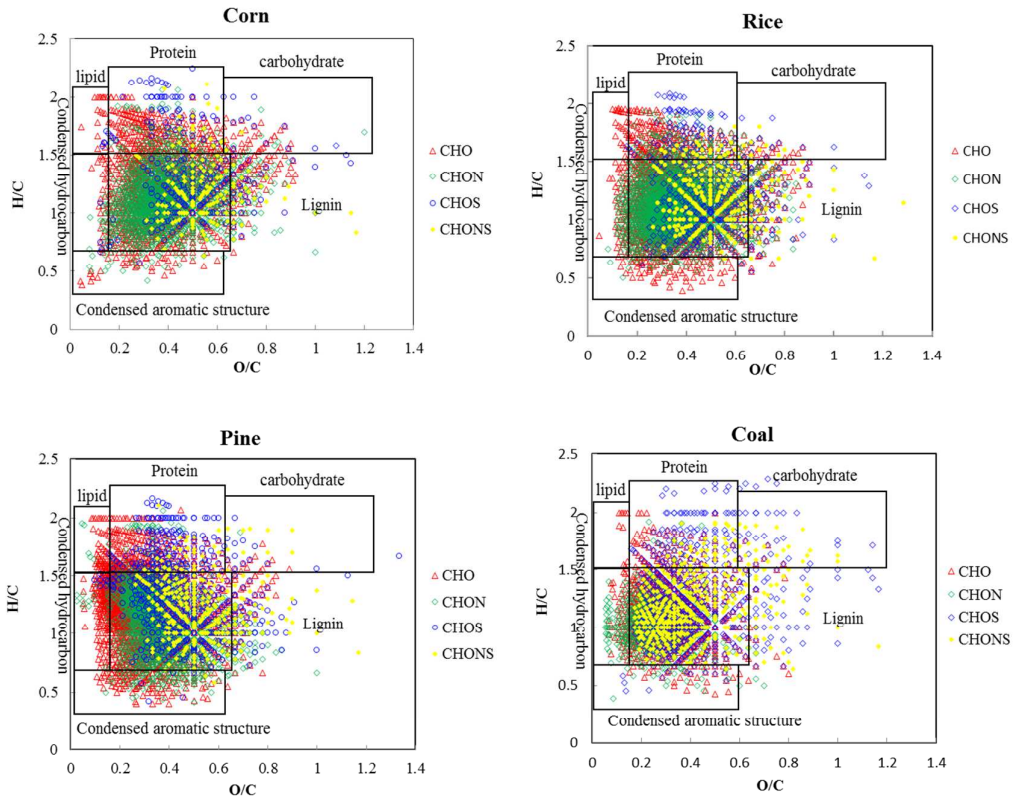

Figure S3. Van Krevelen diagrams of the four smoke HULIS samples. The composition domains in coal-smoke HULIS are the molecular formulae similar to lipids, proteins, lignins, and carbohydrates.

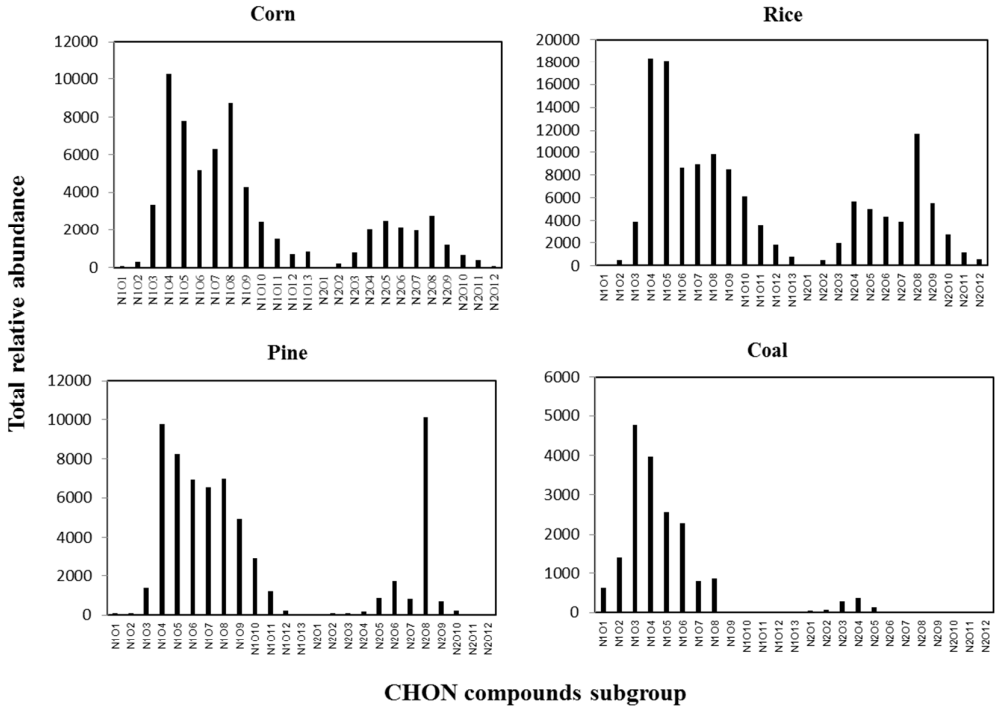

Figure S4. Classification of CHON compounds into different subgroups according to the number of N and O atoms in their molecules. The high of column is the sum of peak intensities in each subgroup.

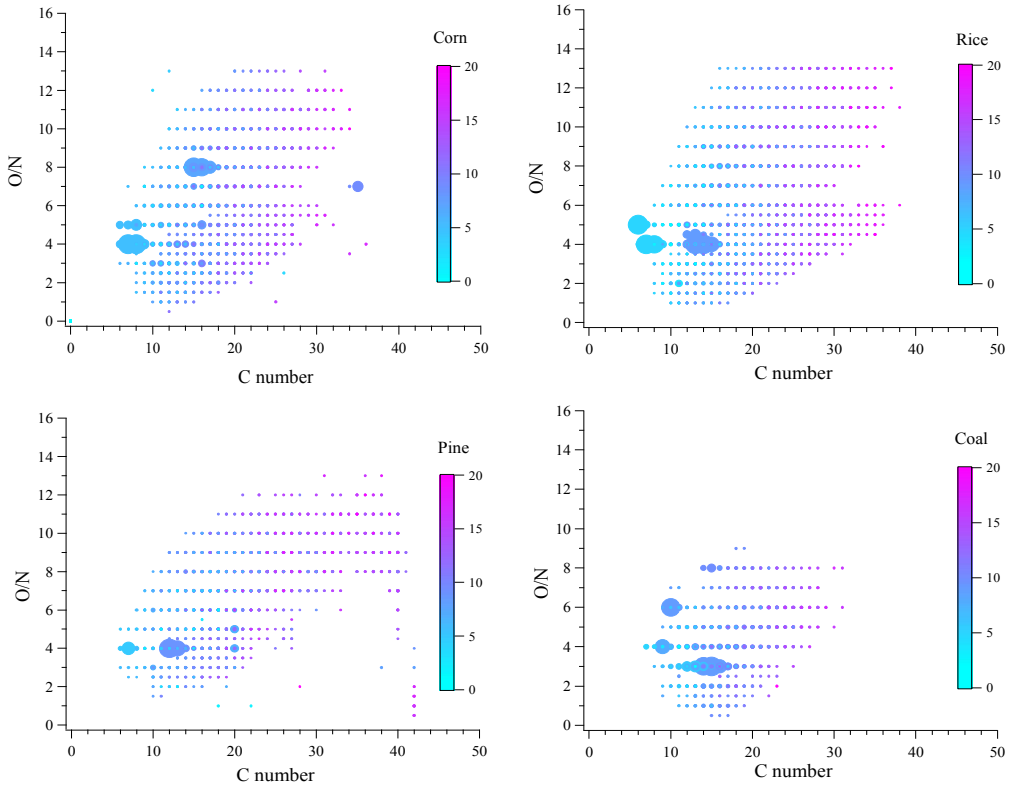

Figure S5. O/N vs C number for all the CHON compounds of the four smoke HULIS samples. The color bar and marker size denote the DBE values and the peak intensities of the compounds.

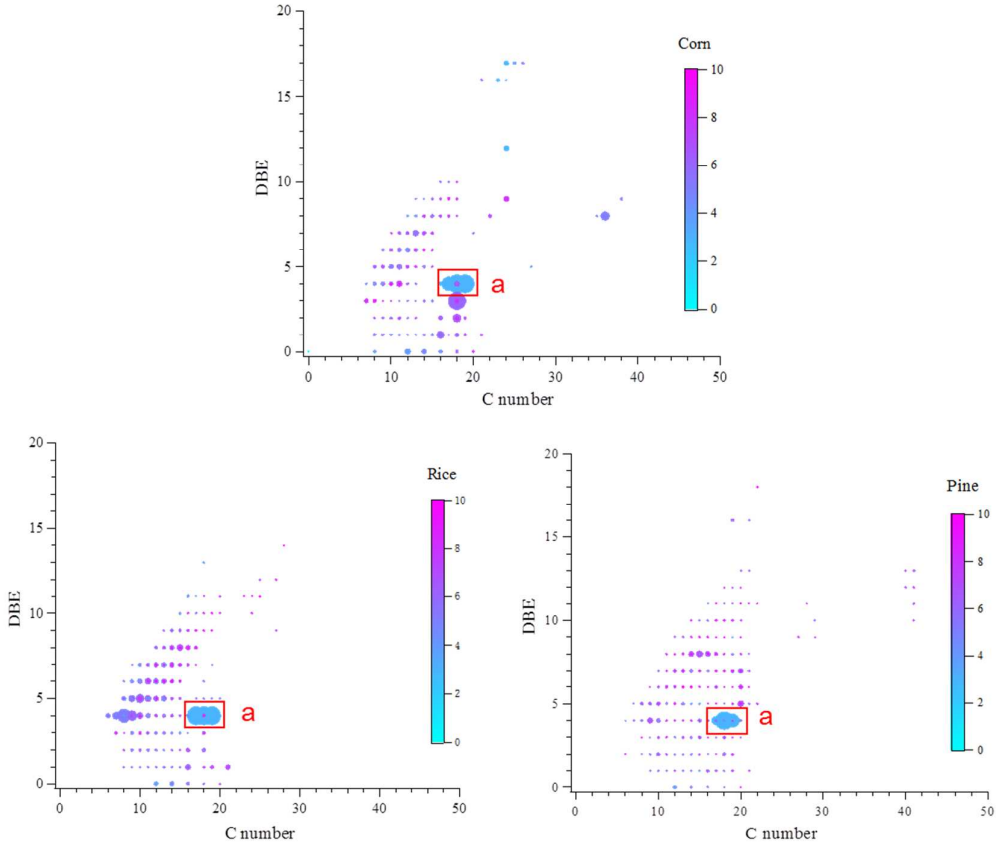

Figure S6. Double bond equivalent (DBE) vs C number for all the CHOS compounds of the three BB-smoke HULIS samples. The color bar and marker size denote the number of O atoms and the peak intensities of the compounds. The formulae in the box a are $C_{17}H_{27}O_3S_1$, $C_{18}H_{29}O_3S_1$, and $C_{19}H_{31}O_3S_1$, corresponding to alkane sulfonates that are known to exist widely in the atmospheric environment and the background pollution during FT-ICR MS measurement.