## **Supporting Information:**

# Two dimensional correlation analysis of Fourier transform ion cyclotron resonance mass spectra of dissolved organic matter: A new graphical analysis of trends

Hussain A.N. Abdulla<sup>\* §</sup>, Rachel L. Sleighter<sup>§</sup> and Patrick G. Hatcher<sup>§</sup>

§ Department of Chemistry & Biochemistry, Old Dominion University, Norfolk, VA 23529, USA

\* Corresponding author contact: (757) 683-4106; fax: (757) 683-4628;

email: habdulla@odu.edu

Number of Pages: 15 Number of Tables: 2 Number of Figures: 10

#### Supporting Text to Experimental Section

The spray voltages and parameters of the FTICR-MS were optimized for each sample to maintain stable, consistent ion currents from one sample to the next. This resulted in a shield current of approximately 140 nA and a capillary current of 15-20 nA for all samples. Exactly 300 transients, collected with a 4-Megaword time domain, were co-added. The summed free induction decay signal was zero-filled once and sine-bell apodized prior to fast Fourier transform and magnitude calculation using the Bruker Daltonics Data software. Average resolving powers of the broadband mass spectra are greater than 400,000.

All assigned molecular formulas have an error of  $\leq 1.0$  ppm (see Figure S-1). To remove formulas that are unlikely to occur in nature, we applied a modified version of rules set by Kind and Fiehn<sup>20</sup>, using H/C < 2.50, O/C  $\leq 1.20$ , O/P  $\geq 3.00$ , N/C < 0.50. All assigned formulas were tested for the physical existence of chemical structures using LEWIS and SENIOR chemical rules, and to the best of our knowledge, it is the first time that these chemical rules have been applied for assigning molecular formulas to the mass spectra of dissolved organic matter (DOM) mixtures. We also validate the molecular formula assignment using <sup>13</sup>C isotope peaks (when they are detected above the S/N threshold) and the chemical building block approaches described by Koch et al.<sup>19</sup>. Approximately 90-93% of all m/z values detected in each spectrum were assigned a unique molecular formula.

Table S-1.	The	defined	regions o	f each	compound	class on	the van	Krevelen	diagram.
			•		1				<u> </u>

Class	Definition
Lipid	O/C 0.0-0.2
	H/C 1.7-2.2
Protein	O/C 0.2-0.6
	H/C 1.5-2.2
	N/C > 0.05
Carbohydrate	O/C 0.6-1.2
	H/C 1.5-2.2
Lignin/CRAM	$O/C > 0.1- \le 0.6$
	H/C 0.6- < 1.7
	*AI < 0.67
Tannins	O/C > 0.6-1.2
	H/C 0.5- < 1.5
	*AI < 0.67
Unsaturated hydrocarbon	O/C 0.0- $\leq$ 0.1
	H/C 0.7-1.5
Condensed Aromatic	O/C 0.0-0.1
	H/C 0.3-0.7
	$^{*}AI \ge 0.67$

\*AI = aromaticity index, calculated according to Koch and Dittmar, *Rapid Commun. Mass Sp.* **2006**, *20(5)*, 926-932.

m/z	С	Н	H+1	N	0	S	Р
227.201654	14	27	28	0	2	0	0
241.217304	15	29	30	0	2	0	0
253.217304	16	29	30	0	2	0	0
255.232954	16	31	32	0	2	0	0
267.232954	17	31	32	0	2	0	0
269.248604	17	33	34	0	2	0	0
279.232954	18	31	32	0	2	0	0
281.248604	18	33	34	0	2	0	0
283.264254	18	35	36	0	2	0	0
285.080218	13	17	18	0	5	1	0
297.152989	16	25	26	0	3	1	0
299.095868	14	19	20	0	5	1	0
311.168639	17	27	28	0	3	1	0
313.111518	15	21	22	0	5	1	0
317.212218	20	29	30	0	3	0	0
319.227868	20	31	32	0	3	0	0
325.184289	18	29	30	0	3	1	0
333.207133	20	29	30	0	4	0	0
335.222783	20	31	32	0	4	0	0
337.238433	20	33	34	0	4	0	0
339.199939	19	31	32	0	3	1	0
351.217698	20	31	32	0	5	0	0
367.358154	24	47	48	0	2	0	0
375.182181	22	31	32	0	1	2	0

Table S-2. List of the 24 outlier peaks that were removed before the normalization and 2D correlation steps.

#### **Figure captions**

Figure S-1. Formula errors (in ppm) vs. m/z for all peaks assigned an molecular formula from all 20 UDOM samples.

Figure S-2. Graphical representation of the sulfur containing compounds along the O/C and S/C axes.

Figure S-3. The synchronous plots generated by applying hetero 2D correlation on 20 sub-spectra of S/C vs. O/C of the sulfur containing compounds.

Figure S-4. The synchronous map generated from all FTICR mass spectra (n=20) of UDOM isolated along the salinity transect. a) The entire mass spectral range (220-600 m/z), and the mass spectral ranges of b) 300-350 m/z, c) 400-450 m/z, and d) 500-550 m/z. The correlation was based on the normalized magnitude of each individual peak. Red represents positive correlations and green represents negative correlations; higher color intensity indicates a stronger positive or negative correlation.

Figure S-5. The correlation intensity of the individual peaks along the diagonal line of the 2D synchronous map (Figure S-4).

Figure S-6. The average normalized magnitude of the individual peaks for the FTICR mass spectra of all 20 samples.

Figure S-7. The correlation intensities of all the individual peaks with m/z = 341.066676 ( $C_{18}H_{13}O_7^{-}$ ). The red lines indicate the standard deviations that separate the weak vs. strong correlations of both positive and negative correlations.

Figure S-8. The van Krevelen diagram of the molecular formulas that show strong positive correlations with  $m/z = 341.066676 (C_{18}H_{13}O_7)$ .

Figure S-9. The van Krevelen diagram of the molecular formulas that show strong negative correlations with  $m/z = 341.066676 (C_{18}H_{13}O_7)$ .

Figure S-10. The synchronous map of the 2D hetero-spectral correlation analysis of the normalized magnitudes of individual FTICR-MS peaks and <sup>13</sup>C-NMR spectra (n=20). Positive correlations are represented by red and negative correlations are represented by green. The left side displays the average <sup>13</sup>C-NMR spectrum for all samples and is only shown as a reference.

Figure S-1





Figure S-3



Figure S-4

a)



b)



m/z

S-10



m/z

d)

c)



m/z

S-11

Figure S-5



Figure S-7



Figure S-8



S-13

### Figure S-9





