Suspect Screening of Disinfection Byproduct Precursors in Water Treatment Processes using Ultrahigh Resolution Mass Spectrometry

Juliana R. Laszakovits[1,2], Allison A. MacKay[1]
[1]Department of Civil, Environmental, and Geodetic Engineering, The Ohio State University, Columbus, OH, 43210, USA
[2]Institute of Biogeochemistry and Pollutant Dynamics,Department of Environmental Systems Science, ETH-Zuerich, Zuerich, 8002, CH

Table of Contents

Text S1: Bulk dissolved organic matter (DOM) property characterization.

Dissolved organic carbon (DOC) and TDN. A Shimadzu TOC-CSVN total organic carbon (TOC) analyzer with a total dissolved nitrogen (TDN) attachment and an ASI autosampler was used. Samples were acidified below pH 2 with 2% 2M HCl and sparged for 2 minutes prior to analysis. An injection volume of 50 µL was used.

UV-Vis. Ultraviolet-visible light (UV-Vis) spectra were obtained between 200 – 600 nm using a Cary 4000 UV-Vis spectrophotometer. The instrument was blanked with air v. air with no cuvette inserted. A DI blank was measured and manually subtracted from each sample spectrum to ensure a 1-cm quartz cuvette was clean and to account for absorbance by water.

E2/E3 Ratio and SUVA-254. The absorbance data was used to determine E2/E3 ratios (Equation 3.1) and specific ultraviolet absorption at 254 nm (SUVA-254, L mg_C^{-1} m^{-1}) values (Equation 3.2) with the dissolved organic carbon (DOC, mg_C L^{-1}) data.

$$E2/E3\,\text{Ratio} = \frac{Abs_{250nm}}{Abs_{365nm}} \qquad \text{Equation 3.1}$$

$$\text{SUVA-254} = \frac{Abs_{254nm}}{DOC} \times 100 \qquad \text{Equation 3.2}$$

where Abs_{250nm} is the absorbance at 250 nm and Abs_{365nm} is the absorbance at 365 nm.

Text S2. DOM characterization by ultrahigh resolution mass spectrometry.

Data Collection. For electrospray ionization (ESI), solid samples were dissolved in 1:1 water acetonitrile and sprayed at a rate of 2.5 μL min^{-1}. The Fourier Transform-Ion cyclotron resonance mass spectrometer (FT-ICR-MS) is housed in the Ohio State University Mass Spectrometry and Proteomics Campus Chemical Instrument Center and is used daily by a variety of researchers. Because of this, the FT-ICR-MS is calibrated daily for ESI using a polypropylene glyocol mix. Prior to data collection, our spectra were calibrated externally using known background contaminants (i.e., fluorotelomers). For laser desorption/ionization (LDI), samples were dissolved in methanol and plated on a 384-fold stainless steel matrix assisted laser desorption/ionization (MALDI) plate. Prior to data collection the FT-ICR-MS was calibrated externally using a PepMix standard. The methanol was evaporated prior to analysis and a Yag/Nd (351 nm) laser was used to ionize the solid sample with varying laser power. The ESI spectra were internally calibrated using known fatty acids and background contaminants. Chemical formulae for each mass-to-charge ratio obtained were assigned using Bruker's Data Analysis software (version 5.1).

Data Analysis. Ion formulae were predicted using Bruker's Compass Data Analysis software (v 5.0). Peaks with an absolute intensity greater than 100, relative base peak intensity greater than 0.05% and a signal-to-noise (S/N) ratio greater than 5 were considered. All ions were assumed to have a charge of 1 (singly charged ions). The elemental composition considered by the software was set such that:

$C \leq 50$	Equation S1
$H \leq 102$	Equation S2
$N \leq 25$	Equation S3
$O \leq 60$	Equation S4
$P \leq 5$	Equation S5
$S \leq 10$	Equation S6

For positive ion modes, M+Na and M+H ions were considered; for negative ion modes only M-H ions were considered. Molecular ions were also considered for both LDI and ESI (odd electron configurations). Up to 1000 formulae were considered for every peak. The data was exported in an excel spreadsheet and subsequently formulae were eliminated using a Matlab code.

A Matlab script developed in-house was used to remove unlikely or impossible formulae and subsequently to combine chemical formulae obtained by different ionization techniques.[1-5] We excluded formulae on the basis of an allowable error for the FT-ICR-MS used (≤0.5 ppm). We only considered formulae between m/z = 200 and 700 because of how the instrument was tuned and because we could reliably assign formulae in this mass range. We removed background impurity formulae based on a known background spectrum. Only formulae that met the following criteria were included based on previous work and adjusted (*) to eliminate unlikely chemical formulae[2-5]:

$2 \leq H \leq 2C+2$	Equation S7
$O \leq C+2$	Equation S8
$O/C < 1.2$	Equation S9
$0.33 \leq H/C \leq 2.25$	Equation S10
$N/C < 0.5$	Equation S11
$S/C < 0.2$	Equation S12
$P/C < 0.05$	Equation S13
$(S+P)/C < 0.1$	Equation S14

$$(N+P)/C<0.4 \quad \text{Equation S15}$$

$$DBE \geq 0 \text{ and an integer} \quad \text{Equation S16}$$

$$-10 \leq DBE + O \leq 10 \quad \text{Equation S17}$$

$$AI<1 \text{ and a real number*} \quad \text{Equation S18}$$

Where DBE is the double bond equivalents and AI is the aromaticity index .[6, 7]

$$DBE = 1 + \frac{1}{2}(2C - H - Cl + N + P) \quad \text{Equation S19}$$

$$AI = \frac{1 + C - O - S - \frac{1}{2}(N+P+H)}{C - O - N - S - P} \quad \text{Equation S20}$$

We considered elemental ratios (i.e., H/C, O/C, N/C, etc.) rather than number of the element in a formula so that the same constraints could be used for all ionization techniques. Specifically, peptides that contain amines readily ionize in positive mode and, as such, it is possible to ionize compounds that contain greater than 3 nitrogen atoms, which is a typical constraint for negative mode ESI.[3] From the elemental ratios, van Krevelen diagrams were constructed and used for comparison.[8, 9]

Text S3. Disinfection byproduct (DBP) quantification.

Haloacetic Acids. Haloacetic acids were analyzed using an adapted Environmental Protection Agency (EPA) method 557 using a Exion liquid chromatograph and Sciex 3500 mass spectrometer. A ^{13}C-bromoacetic acid internal standard (resulting concentration of 40 ppb) and 0.5 mL of concentrated sulfuric acid were added to 10 mL of the chlorinated samples, prior to analysis. The haloacetic acids were separate using a Luna Polar C18 column (1.6 µm particle size), a flow rate of 0.3 mL min^{-1}, and a gradient that begins with 95:5 water (0.1% formic acid) : acetonitrile (0.1% formic acid) that changes to 60:40 water (0.1% formic acid) : acetonitrile (0.1% formic acid) over 4 minutes and then to 100% acetonitrile (0.1% formic acid) over 1 minute and is held for 1 minute. The analytes were ionized using negative mode electrospray ionization. The source gas parameters are summarized in the table below.

Haloacetic acid source gas parameters.

Source Gas Parameter	Value
Curtain Gas (CUR)	30
Collision Gas (CAD)	5
IonSpray Voltage (IS)	-4500
Temperature	600
Ion Source Gas 1 (GS1)	30
Ion Source Gas 2 (GS2)	50

The analytes were monitored using a product ion scan. The ion transitions and entrance ion optics are summarized in the table below.

Haloacetic acid retention times and product ions.

Analyte	Retention Time (min)	Ion Type	Q1	Q3	DP	CE	EP	CXP
Chloroacetic Acid	1.5	Quantifying	93	35	-20	-14	-10	-11
		Qualifying	95	37	-20	-18	-10	-11
Dichloroacetic Acid	1.4	Quantifying	127	83	-17	-12	-10	-11
		Qualifying	129	85	-17	-11	-10	-11
Trihloroacetic Acid	2.4	Quantifying	163	119	-11	-7	-10	-11
		Qualifying	161	117	-13	-9	-10	-11
Bromoacetic Acid	1.83	Quantifying	137	79	-15	-14	-10	-11
		Qualifying	139	81	-15	-17	-10	-11
Dibromoacetic Acid	1.8	Quantifying	217	173	-22	-13	-10	-11
		Qualifying	219	175	-21	-12	-10	-11
Tribromoacetic Acid	3.32	Quantifying	253	81	-25	-30	-10	-11
		Qualifying	251	79	-25	-27	-10	-11
Bromochloroacetic Acid	1.55	Quantifying	173	71	-30	-28	-10	-11
		Qualifying	171	79	-42	-24	-10	-11
Bromodichloroacetic Acid	2.65	Quantifying	161	79	-20	-12	-10	-11
		Qualifying	163	81	-20	-12	-10	-11
Dibromochloroacetic Acid	2.95	Quantifying	207	79	-15	-18	-10	-11
		Qualifying	209	81	-15	-15	-10	-11
Internal Standard	1.83	Quantifying	138	79	-15	-15	-10	-11

Volatile DBPs. An adapted EPA method 551 was used to analyze: chloroform, 1,1,1-trichloroethane, carbon tetrachloride, trichloroacetonitrile, trichloroethylene, bromodichloromethane, dichloroacetonitrile, 1,1-dichloro-2-propanone, chloropicrin, dibromochloromethane, tetrachloroethylene, 1,2-dibromoethane, 1,1,1-trichloroacetone, 1,2-dibromopropane (extraction surrogate), bromoform, dibromoacetonitrile, bromofluorobenzene (internal standard), 1,2-dibromo-3-chloropropane using an Agilent 6890 GC with an autosampler and an Agilent 5973 mass spectrometer detector. The analytes were extracted from 50 mL of sample using 3 mL of methyl tertbutyl ether. To improve analyte stability and improve extraction efficiency, the headspace of the extraction vials was minimized and 30 g sodium sulfate and 5 mL pH 4.8 1M phosphate were added. The analytes were separated using a 30m, 0.25 mm inner diameter, 0.25 µm thick film DB-5 column, and a helium mobile phase at 1 mL min^{-1}. An injection volume of 1 µL was used with a splitless injection into an injector set at 120 °C. A temperature gradient was used. The initial temperature was 35 °C and was held for 4 minutes then ramped at 30 °C min^{-1} to 180 °C and finally ramped at 50 °C to 225 °C and held for 5 minutes at 225 °C. Selected ion monitoring was used for each analyte and are summarized with retention times in the table below. In this study, only of the volatile C-DBPs that the method could detect only chloroform, 1,1-dichloro-2-propanone, and 1,1-trichloroacetone were detected (belonging to the trihalomethane and haloketone classes). Of the volatile N-DBPs, only dibromoacetonitrile, dichloroacetonitrile, trichloroacetonitrile and chloropicrin were detected (belonging to the halonitrile and halonitromethane classes).

Volatile DBP retention times and ions monitored

Analyte	**Retention Time (min)**	**Ions Monitored (*m/z*)**
Chloroform	2.55	83
1,1,1-trichloroethane	2.79	97
Carbon tetrachloride	3.02	117
Trichloroacetonitrile	3.28	108
Trichloroethylene	3.76	130
Bromodichloromethane	4.09	83
Dichloroacetonitrile	4.47	74
1,1-dichloro-2-propanone	4.54	43
Chloropicrin	5.29	117
Dibromochloromethane	5.61	127
Tetrachloroethylene	5.64	164
1,2-dibromoethane	5.74	107
1,1,1-trichloroacetone	6.09	43
1,2-dibromopropane	6.31	123
Bromoform	6.65	173
Dibromoacetonitrile	6.93	118
Bromofluorobenzene	6.97	75
1,2-dibromo-3-chloropropane	8.15	157

Nitrosamines. Nitrosamines were analyzed according to EPA method 521 using an ExionLC and Sciex 3500 mass spectrometer. A N-Nitrosodi-n-propylamine-d14 internal standard (resulting concentration of 80 ppt) was added to 500 mL of the chlorinated samples and extracted using Enviro-Clean Method 521 SPE cartridges. The cartridges were primed with 8 mL methylene chloride, 8 mL methanol, and 20 mL

water and then the 500 mL of sample were run through the column at < 20 mL min^{-1}. High vacuum was pulled to dry the column for 10 minutes. The nitrosamines were eluted using 10 mL methylene chloride. The samples were dried with sodium sulfate and blown down to a volume of 1 mL. The samples were then analyzed by LC-MS. The nitrosamines were separated using a Kinetix F5 column (5 µm particle size) using a gradient elution with a flow rate of 0.8 mL min^{-1}. The initial mobile phase conditions were 95 water (0.1% formic acid) : 5 acetonitrile (0.1% formic acid) and were held for 1 minute, the acetonitrile percentage was increased to 99% over 5 minutes and held for 1 minute before returning to the initial conditions. The analytes were ionized using positive mode atmospheric pressure chemical ionization and the source gas parameters are summarized in the table below.

Nitrosamine source gas parameters

Source Gas Parameter	Value
Curtain Gas (CUR)	30
Collision Gas (CAD)	4
Nebulizer Current (NC)	2
Temperature	350
Ion Source Gas 1 (GS1)	35

The analytes were quantified using product ion scans and the entrance optic parameters are summarized in the table below.

Nitrosamine retention times and product ions.

Analyte	Retention Time (min)	Ion Type	Q1	Q3	DP	CE	EP	CXP
N-nitrosodimethylamine	0.95	Quantifying	75	43	53	23	10	6
		Qualifying	75	58	53	17	10	6
N-nitrosomethylethylamine	1.65	Quantifying	89	61	60	17	10	6
		Qualifying	89	43	60	20	10	6
N-nitrosodiethylamine	3.25	Quantifying	103	75	60	15	10	6
		Qualifying	103	29	60	23	10	6
N-nitrosodipropylamine	4.30	Quantifying	131	89	60	15	10	6
		Qualifying	131	43	60	22	10	6
N-nitrosodibutylamine	5.00	Quantifying	159	103	70	15	10	6
		Qualifying	159	57	70	20	10	6
1-Nitrosopiperidine	3.40	Quantifying	115	69	73	21	10	6
		Qualifying	115	41	73	34	10	6
1-Nitrosopyrrolidine	1.55	Quantifying	101	59	70	24	10	6
		Qualifying	101	55	70	23	10	6
Internal Standard	4.25	Quantifying	145	97	40	16	10	6

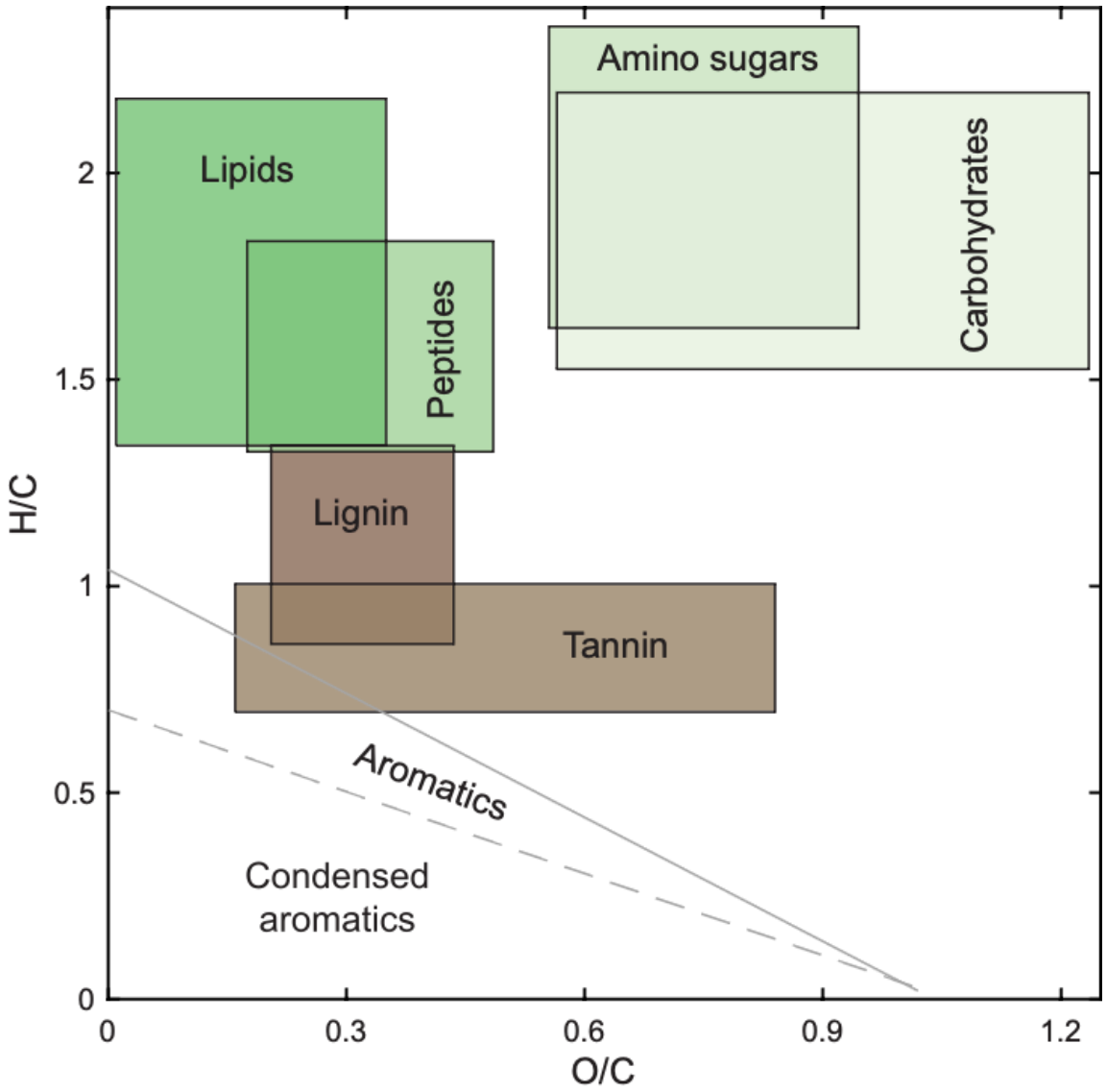

Figure S1: Example van Krevelen Diagram. The different regions (lipid-, peptide-, amino sugar-, carbohydrate-, lignin- and tannin-like) as determined by Laszakovits and MacKay (2021) are highlighted in the boxes.[10] Aromatics and condensed aromatics as determined by the aromaticity derived by Koch *et al.* (2006 & 2016) are shown as lines, where below the line the formula is classified as either aromatic (below solid line) or condensed aromatic (below dashed lines).[6, 7] The boxes and lines are included on every van Krevelen diagram presented to improve readability and easily make comparisons.

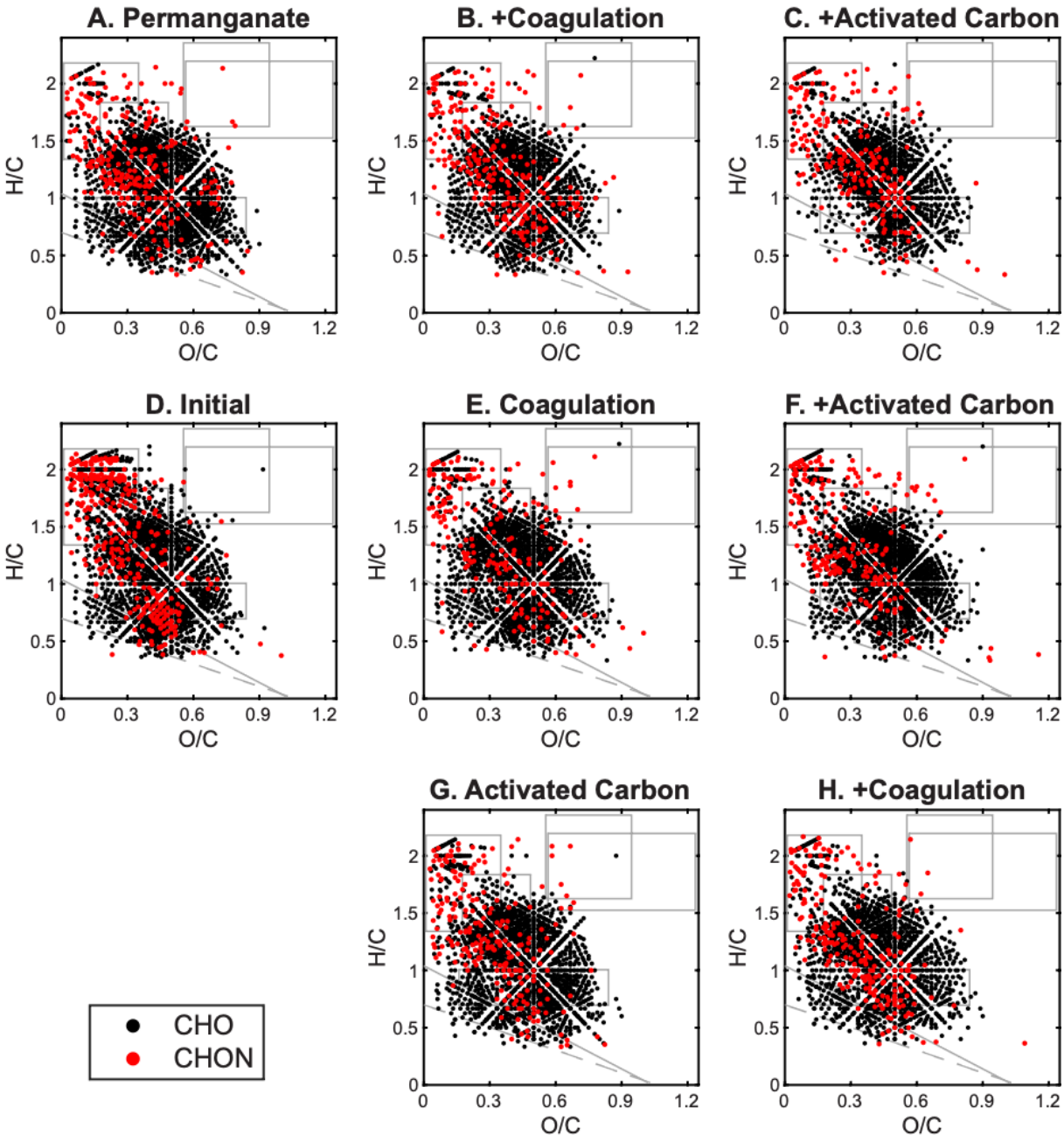

Figure S2: Overview van Krevelen diagrams of SRFA after treatments. Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is ordered to mimic the experimental overview found in Figure 1 in the main text. A. Formulae present after permanganate oxidation; B. Formulae present after coagulation following permanganate oxidation; C. Formulae present after activated carbon sorption following coagulation and permanganate oxidation; D. Formulae present in initial SRFA sample; E. Formulae present after coagulation; F. Formulae present after activated carbon sorption following coagulation; G. Formulae present after activated carbon sorption; H. Formulae present after coagulation following activated carbon sorption.

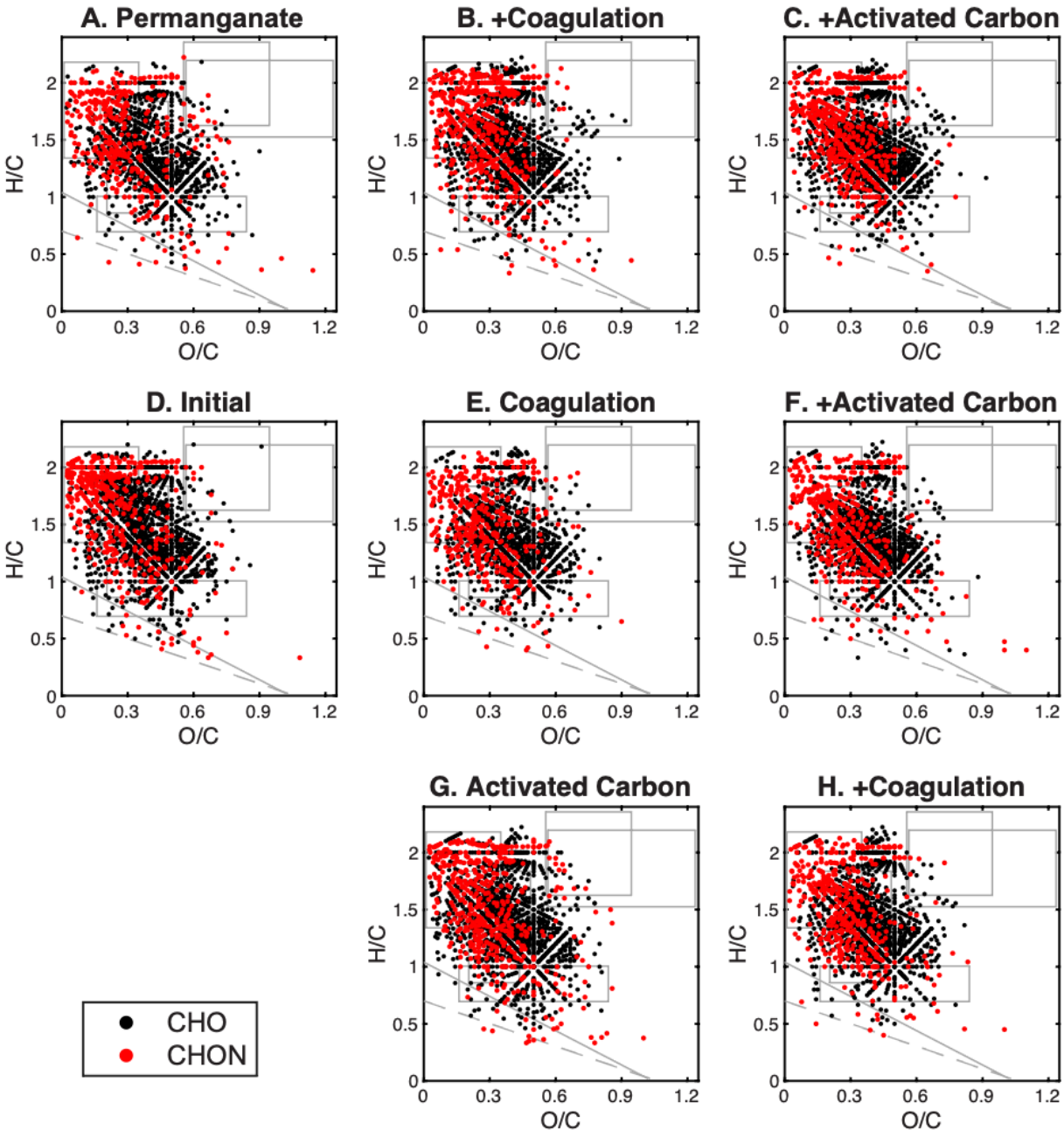

Figure S3: Overview van Krevelen diagrams of EfOM after treatments. Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is ordered to mimic the experimental overview found in Figure 1 in the main text. A. Formulae present after permanganate oxidation; B. Formulae present after coagulation following permanganate oxidation; C. Formulae present after activated carbon sorption following coagulation and permanganate oxidation; D. Formulae present in initial EfOM sample; E. Formulae present after coagulation; F. Formulae present after activated carbon sorption following coagulation; G. Formulae present after activated carbon sorption; H. Formulae present after coagulation following activated carbon sorption.

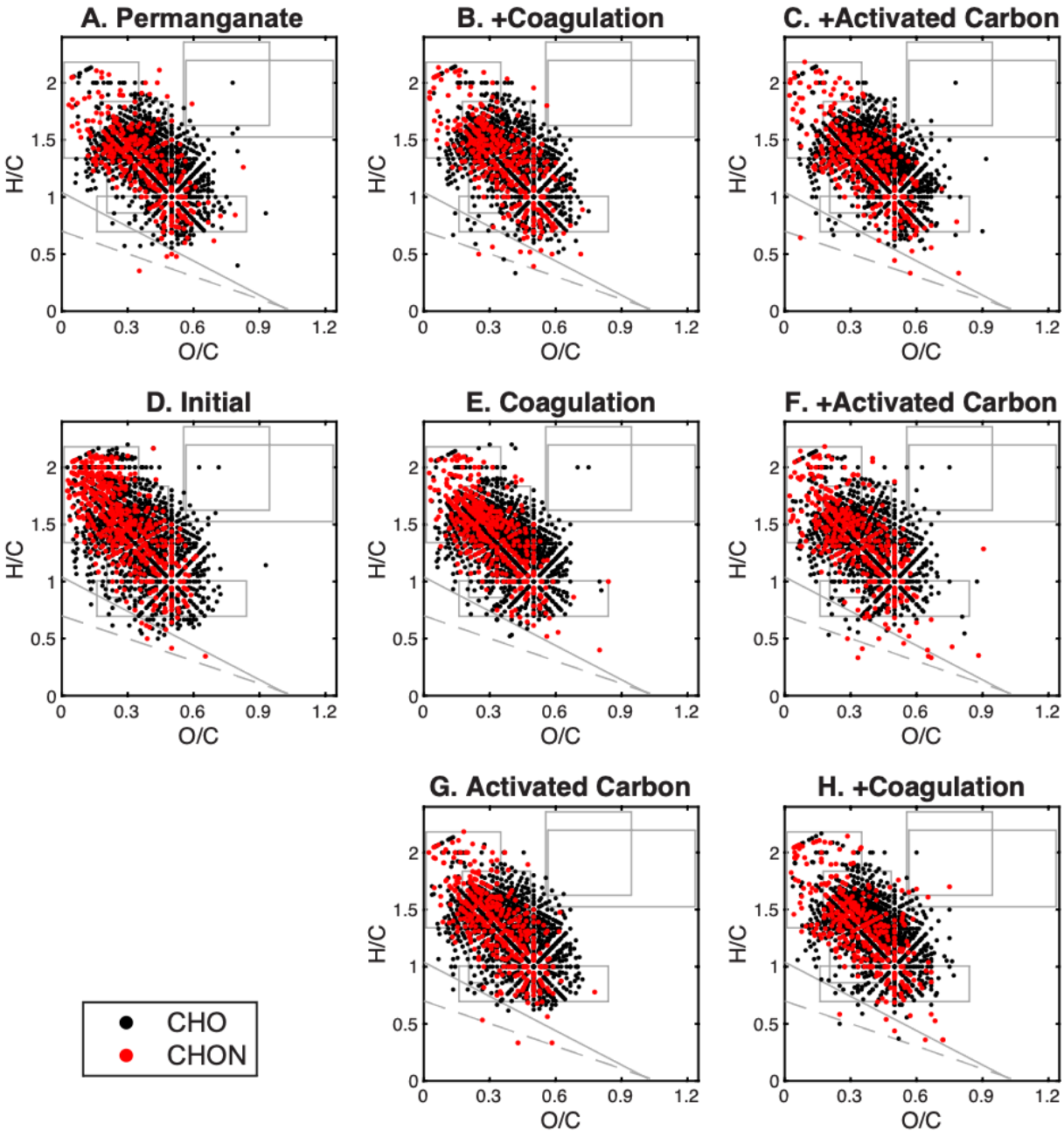

Figure S4: Overview van Krevelen diagrams of AOM after treatments. Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is ordered to mimic the experimental overview found in Figure 1 in the main text. A. Formulae present after permanganate oxidation; B. Formulae present after coagulation following permanganate oxidation; C. Formulae present after activated carbon sorption following coagulation and permanganate oxidation; D. Formulae present in initial AOM sample; E. Formulae present after coagulation; F. Formulae present after activated carbon sorption following coagulation; G. Formulae present after activated carbon sorption; H. Formulae present after coagulation following activated carbon sorption.

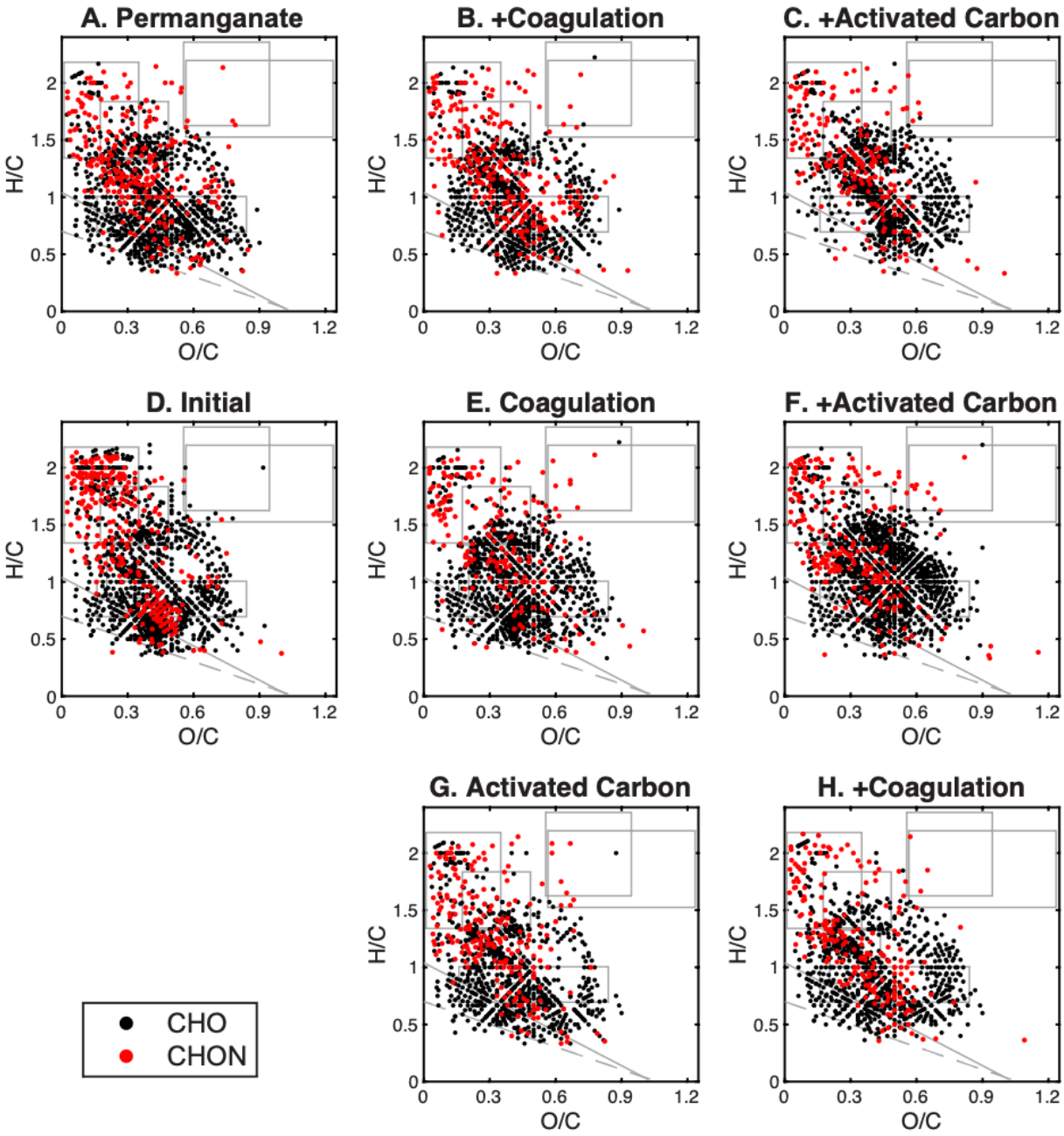

Figure S5: van Krevelen diagrams of SRFA suspect precursors. Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is ordered to mimic the experimental overview found in Figure 1 in the main text for all the chlorinated samples. A. Suspect precursors in the permanganate oxidized sample; B. Suspect precursors in the sample after coagulation following permanganate oxidation; C. Suspect precursors in the sample after activated carbon sorption following coagulation and permanganate oxidation; D. Suspect precursors in the initial SRFA sample; E. Suspect precursors in the sample after coagulation; F. Suspect precursors in the sample after activated carbon sorption following coagulation; G. Suspect precursors in the sample after activated carbon sorption; H. Suspect precursors in the sample after coagulation following activated carbon sorption.

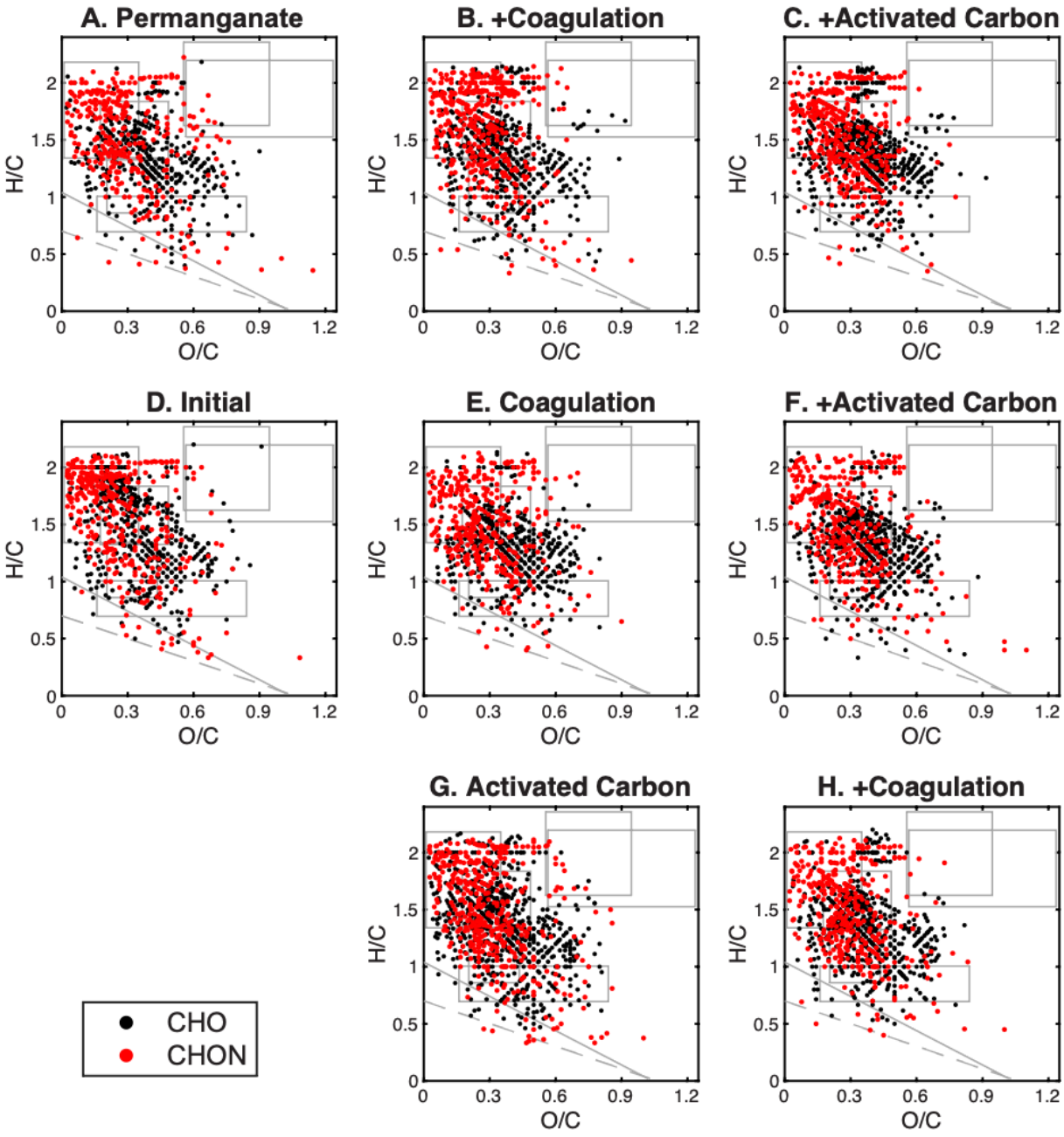

Figure S6: van Krevelen diagrams of EfOM suspect precursors. Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is ordered to mimic the experimental overview found in Figure 1 in the main text for all the chlorinated samples. A. Suspect precursors in the permanganate oxidized sample; B. Suspect precursors in the sample after coagulation following permanganate oxidation; C. Suspect precursors in the sample after activated carbon sorption following coagulation and permanganate oxidation; D. Suspect precursors in the initial EfOM sample; E. Suspect precursors in the sample after coagulation; F. Suspect precursors in the sample after activated carbon sorption following coagulation; G. Suspect precursors in the sample after activated carbon sorption; H. Suspect precursors in the sample after coagulation following activated carbon sorption.

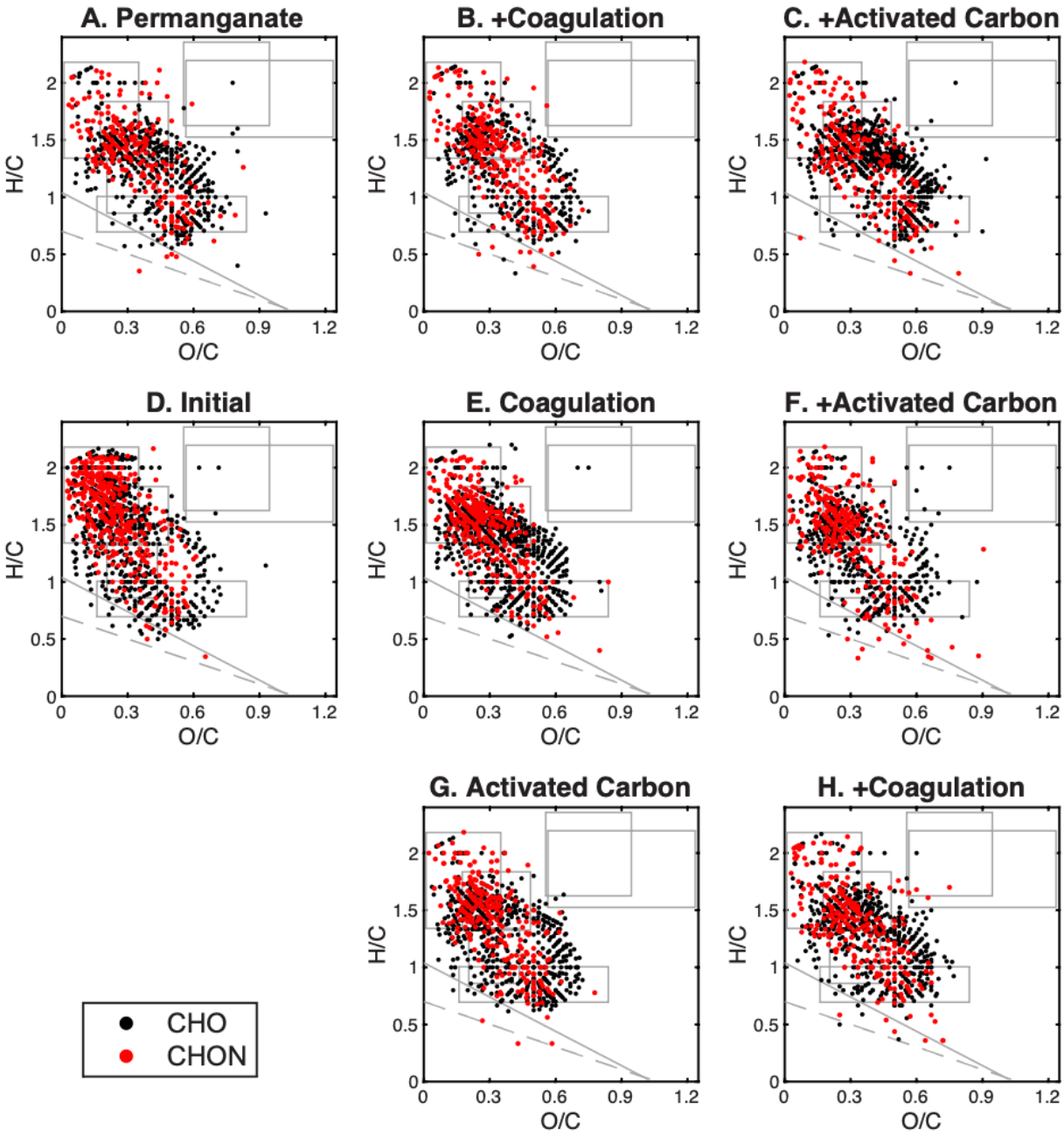

Figure S7: van Krevelen diagrams of AOM suspect precursors. Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is ordered to mimic the experimental overview found in Figure 1 in the main text for all the chlorinated samples. A. Suspect precursors in the permanganate oxidized sample; B. Suspect precursors in the sample after coagulation following permanganate oxidation; C. Suspect precursors in the sample after activated carbon sorption following coagulation and permanganate oxidation; D. Suspect precursors in the initial AOM sample; E. Suspect precursors in the sample after coagulation; F. Suspect precursors in the sample after activated carbon sorption following coagulation; G. Suspect precursors in the sample after activated carbon sorption; H. Suspect precursors in the sample after coagulation following activated carbon sorption.

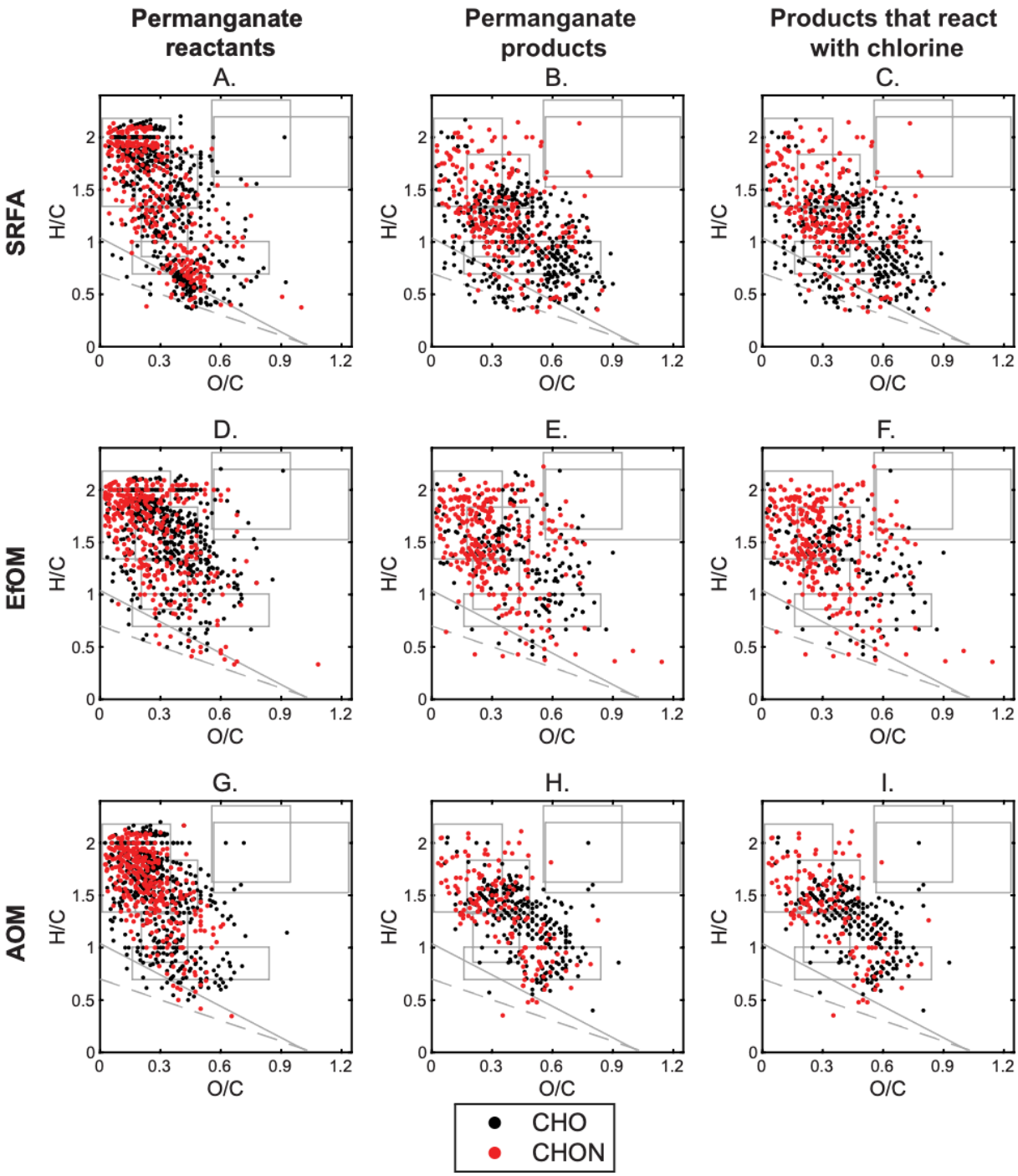

Figure S8: van Krevelen diagrams of permanganate oxidation reactants (first column), oxidation products (second column), and products that reacted with chlorine (third column). Data is presented as CHO containing formulae (black circles) and nitrogen-containing (CHON) formulae (red circles) with regions as grey boxes or lines presented in Figure S1. The data is shown for SRFA (first row, A-C), EfOM (second row, D-F), and AOM (third row, G-I). There is significant overlap between permanganate oxidation products and those that react with chlorine.

Table S1: Solution water quality parameters. The pH values, cation, and anion concentrations for each model water background are summarized below.

Ion (mM)	SRFA	EfOM	AOM
Na^+	2.3	2.9	2.8
Ca^{2+}	1.4	0.14	0.14
Mg^{2+}	1.2	2.6	2.6
K^+	0.34	0.099	0.16
NH_4^+	0.003	0.003	0.003
Fe^{3+}	0.012	0.012	0.012
HCO_3^-	2.3	2.8	2.8
CO_3^{2-}	-	0.014	0.043
Cl^-	2.9	0.38	0.38
SO_4^{2-}	1.2	2.6	2.6
NO_3^-	0.003	0.003	0.003
$HPO4_2^-$	0.005	0.005	0.005
pH	7	8	8.5

References

(1) 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. *Marine Chemistry* **2008**, *110*, 140−152. DOI: 10.1016/j.marchem.2008.04.008.

(2) Stubbins, A.; Spencer, R. G. M.; Chen, H.; Hatcher, P. G.; Mopper, K.; Hernes, P. J.; Mwamba, V. L.; Mangangu, A. M.; Wabakanghanzi, J. N.; Six, J. Illuminated darkness: molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnology and Oceanography* **2010**, *55*, 1467-1477. DOI: 10.4319/lo.2010.55.3.1467.

(3) Luek, J. L.; Schmitt-Kopplin, P.; Mouser, P. J.; Petty, W. T.; Richardson, S. D.; Gonsior, M. Halogenated Organic Compounds Identified in Hydraulic Fracturing Wastewaters Using Ultrahigh Resolution Mass Spectrometry. *Environmental Science & Technology* **2017**, *51* (10), 5377-5385. DOI: 10.1021/acs.est.6b06213.

(4) Herzsprung, P.; Hertkorn, N.; von Tumpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Molecular formula assignment for dissolved organic matter (DOM) using high-field FT-ICR-MS: chemical perspective and validation sulphur-rich organic components (CHOS) in pit lake samples. *Analytical and Bioanalytical Chemistry* **2016**, *408*, 2461-2469. DOI: 10.1007/s00216-016-9341-2).

(5) Herzsprung, P.; Hertkorn, N.; von Tumpling, W.; Harir, M.; Friese, K.; Schmitt-Kopplin, P. Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view. *Analytical and Bioanalytical Chemistry* **2014**, *406* (30), 7977-7987. DOI: 10.1007/s00216-014-8249-y.

(6) Koch, B. P.; Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry* **2016**, *30* (1), 250-250. DOI: 10.1002/rcm.7433.

(7) Koch, B. P.; Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry* **2006**, *20* (5), 926-932. DOI: 10.1002/rcm.2386.

(8) Van Krevelen, D. W. Graphical-statistical method for the study of structure and reaction process of coal. *Fuel* **1950**, *29*, 269-284.

(9) Kim, S.; Kramer, R. W.; Hatcher, P. G. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Analytical Chemistry* **2003**, *75* (20), 5336-5344.

(10) Laszakovits, J. R.; MacKay, A. A. Data-Based Chemical Class Regions for Van Krevelen Diagrams. *Journal of the American Society for Mass Spectrometry* **2021**, *33* (1), 198-202. DOI: 10.1021/jasms.1c00230.