A New Formula Assignment Algorithm for the Deuterium Labeled Ultrahigh-Resolution Mass Spectrometry: Implications to the Formation Mechanism of Halogenated Disinfection Byproducts

Qing-Long Fu¹, ², Manabu Fujii²*, Akari Watanabe², Eunsang Kwon³

¹ School of Environmental Studies, China University of Geoscience, Wuhan 430074, China.
² Department of Civil and Environmental Engineering, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-Ku, Tokyo 152-8550, Japan.
³ Research and Analytical Center for Giant Molecules, Graduate School of Science, Tohoku University, 6-3 Aoba, Aramaki, Aoba-Ku, Sendai 980-8578, Japan.

*E-mail: fujii.m.ah@m.titech.ac.jp.
ABSTRACT

The ultrahigh-resolution mass spectrometry (UHR-MS) coupled with isotope labeling is of increasing attentions in elucidating the transform mechanisms of dissolved organic matter (DOM). However, there is a paucity of automated formula assignment algorithm applicable to halogenated disinfection byproducts (X$_n$-DBPs), particularly for iodinated organic compounds, and deuterated DOM. Herein, for the first time, we have developed a novel formula assignment algorithm based on deuterium-labeled UHR-MS, namely FTMSDeu, and the algorithm was applied to determine precursor molecules of X$_n$-DBPs and evaluate the relative contribution of electrophilic addition and electrophilic substitution reactions in X$_n$-DBPs formation according to the hydrogen/deuterium exchange of DOM molecules. Furthermore, tandem mass spectrometry with homologous-based network analysis was employed to validate the formula assignment accuracy (41%) of FTMSDeu for iodinated disinfection byproducts (I$_n$-DBPs). And the remaining I$_n$-DBPs compounds were assigned with the empirical rule of minimum number of non-oxygen heteratoms. The electrophilic substitution accounted for 82%-98%, 71%-89%, and 43%-45% of X$_n$-DBPs formation for X$_n$-DBPs containing chlorine, bromine, and iodine, respectively, manifesting the dominant role of electrophilic substitution in chlorine disinfection under conditions of low bromine and iodine concentrations. The absence of presumed X$_n$-DBPs precursors in some treatments in this study also suggests that X$_n$-DBPs formation include secondary reactions (e.g., oxidation, hydrolysis) in addition to electrophilic addition and/or substitution of halogens. These findings highlight the significance of isotopically labeled UHR-MS techniques in revealing the transformation of DOM in natural and engineered systems.
INTRODUCTION

Natural organic matter (NOM) is a ubiquitous mixture of complex organic compounds from the abiotic and biotic degradation of living organic matter \(^1\), playing important roles as precursors for halogenated disinfection byproducts (X\(_n\)-DBPs, where X represents halogen atoms including chlorine [Cl], bromine [Br], and iodine [I], and \(n\) is the number of halogen atoms) in terrestrial and aquatic environments \(^2\,^3\). Due to the extremely diverse nature of NOM, it was challenging to elucidate their chemical composition at the molecular level until the employment of ultrahigh-resolution mass spectrometry (UHR-MS), particularly the Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) \(^4\,^5\). Since its first application to NOM study \(^6\), FTICR-MS has been widely adopted to characterize the complexity of NOM in the last two decades through the development of automated molecular formula assignment methods such as the in-house code from Kujawinski and Behn \(^7\), MassCal \(^8\), Formularity software \(^9\), MFAssignR \(^10\), ICBM-OCEAN \(^11\), and the TRFu code \(^12\). Moreover, FTICR-MS with stable isotope labeling has initiated the new possibilities of quantifying the number of labile H and O and structural information such as ether O atoms, carboxyl and hydroxyl functional groups in individual molecules of NOM \(^1\,^13\,^15\), further refining compound aromaticity \(^16\), and relation of NOM molecular structures with optical property \(^17\). However, data interpretation of stable isotope-labeled UHR-MS spectra remains challenging, and only a few methods have been developed to address the formula assignment for stable isotope-labeled UHR-MS spectra. For example, the Transhums software is capable of solving the formulae to NOM molecules labeled with deuterium (D) and \(^18\)O, respectively \(^1\,^14\), but it only considers C, H, and O atoms in the formulae calculation.

The H atoms involved in the acid-base reactive moieties of NOM molecules (\(e.g.,\) carboxyl and hydroxyl groups) can be readily exchanged by D in solution at diffusion-limited rate (referred
as to “labile H”). 13,18. In contrast, when H atoms are incorporated into the structural backbone (referred as to “backbone H”), the H/D exchange generally require activation of the molecules such as acid-, base-, or metal-mediated catalysis and chemical ionization at high temperature 19.

Given the pronounced acidic nature of acid-base functional groups in NOM molecules 20,21 and the fact that the H/D exchange rate for the labile H is much faster than for the backbone H 19, it would be reasonable to assume that the carboxyl and hydroxyl groupd containing labile H account for the majority of the H/D exchange sites for NOM in the D2O system. The resultant labile D in deuterated NOM molecules could be reversibly exchanged with H in the H2O system.

Xn-DBPs are inevitably formed by the interaction between halogens and NOM during chlorination (e.g., NaClO treatment) and of great concern in water and wastewater treatment due to their toxic effects on human and aquatic organisms 22-24. Chlorination of natural and engineered waters containing bromide and iodide may unintentionally yield brominated and iodinated byproducts (Brn-DBPs and In-DBPs, respectively), which are more toxic than chlorinated byproducts (Cln-DBPs) 24. The Cl and Br atoms in hypochlorite (OCl⁻) and hypobromite (OBr⁻) have strong electrophilic properties, and readily react with the abundantly present unsaturated functional groups in NOM molecules mainly via electrophilic substitution and electrophilic addition, followed by secondary reactions such as oxidation, hydrolysis and decarboxylation 25-28. While electrophilic substitution is considered to be major pathway in the formation of Xn-DBPs (compared to electrophilic addition) during disinfection process due to its higher reaction rate 26,28, it is still challenging to quantify their relative contributions primarily because of the high complexity of NOM and associated reactions as well as lack of techniques to identify such system at molecular level. For example, there is a difference in chemical formula of products between electrophilic substitution (-nH+nX) and addition (+nOH+nX) for the identical precursor. Thus the
absolute atom difference between their corresponding precursors of a given Xₙ-DBPs was integer
number of H₂O, which was a typical formula building block for NOM molecules (Figure S1)²⁹-³¹. Furthermore, when disinfection is performed in the D₂O system with NOM, the electrophilic substitution and addition could be distinguished according to the number of D involved in products, as shown in Eqs. 1-3.

\[
\begin{align*}
OX^- + D_2O & \leftrightarrow DOX + OD^- \\
C_xH_yO_z + nDOX & \xrightarrow{\text{Electrophilic substitution}} C_xH_yD_zO_zX_n + nHOD \\
C_xH_yO_z + nDOX & \xrightarrow{\text{Electrophilic addition}} C_xH_yD_zO_zX_n + nHOD
\end{align*}
\]

where XO⁻ and CₓHₓOₓ represent hypohalite ions and NOM molecules, respectively. In case of the labile D in Xₙ-DBPs formed via electrophilic substitution, D will be readily replaced by H when solutions were subjected to H₂O (particularly such reaction can be facilitated under acidic condition [e.g., pH ~2]). In contrast, electrophilically added D in Xₙ-DBPs (i.e., Eq 3) will remain intact under the identical condition.

Recently, UHR-MS techniques have enabled high-throughput non-target screening of Xₙ-DBPs species, including hundreds to approximate three thousand DBPs species. In our previous study, accuracy of formula assignment for Clₙ-DBPs and Brₙ-DBPs was improved up to 97% by accounting for distinct isotopic patterns of Cl and Br in addition to three optional rules. However, an automatic formula assignment algorithm for Iₙ-DBP is not yet available, partly because there is only one naturally occurring stable isotope of iodine (i.e., ¹²⁷I). Furthermore, UHR-MS coupled with D isotope labeling approach is expected to be a valuable tool to improve the accuracy of molecular assignment of Xₙ-DBPs including Iₙ-DBPs. This technique could also be
useful to quantify the contribution of electrophilic substitution and addition in the formation of $X_{\sigma^{-}}$ DBPs and to trace their direct precursors. However, an effective formula assignment method is still required to automatically analyze non-oxygen heteroatoms-containing molecules (e.g., $X_{\sigma^{-}}$ DBPs) for the isotope-labeled UHR-MS spectra of dissolved organic matter (DOM) in natural and engineered environments.

The main objectives of this study were (i) to develop a new formula algorithm to assign formulae to NOM and $X_{\sigma^{-}}$-DBPs labeled with D (where X is Cl, Br or I), and (ii) to apply the developed algorithm to quantify the contribution of reaction mechanisms (i.e., electrophilic substitution and addition) for $X_{\sigma^{-}}$-DBPs at individual molecular level. The relevant results will provide valuable insights into algorithm development for UHR-MS spectra labeled with other isotopes such as $^{13}$C and $^{18}$O and elucidate further details in the formation mechanisms of $X_{\sigma^{-}}$-DBPs.

**METHODS AND MATERIALS**

**Sample preparation.** The Suwannee River NOM (SRNOM [2R101N] purchased from International Humic Substances Society) was prepared at concentration of 50 mg-C/L in 10 mL D$_2$O (99.8%, Kanto Chemical, Japan). The SRNOM solution was then chlorinated with 50 mg/L NaClO (Kanto Chemical, Japan) in the absence and presence of 5.0 mM potassium bromide (>99.0%, Sigma-Aldrich, USA) or 1.0 mM potassium iodide (>99.0%, Sigma-Aldrich, USA). Thus, following three samples were prepared: i.e., (i) ClO$^-$ + NOM in D$_2$O (referred to as “Treatment A”), (ii) ClO$^-$ + Br$^-$ + NOM in D$_2$O (“Treatment B”), and (iii) ClO$^-$ + I$^-$ + NOM in D$_2$O (“Treatment C”). All samples were then incubated for a week at room temperature under the dark condition. The chlorination reactions were terminated by adding excess Na$_2$SO$_3$ (>99.0%, Kanto...
Chemical, Japan). Due to the limited availability of D$_2$O, concentrations of aforementioned chemicals were set at approximately ten times the dose of ClO$^-$ typically used in water treatments and ten times the environmentally relevant concentrations of dissolved organic carbon, Br$^-$ and I$^-$. The pH values for Treatments A, B, and C were determined to be 8.11, 8.93, and 7.81 at the beginning of the treatment, and 6.20, 6.30, and 5.08 at the end of the treatment (i.e., after one week), respectively.

After the treatment, the samples were diluted to 250 mL with ultrapure water (Milli-Q, ≥15 MΩ·cm), followed by the solid-phase extraction (SPE) for dissolved organic matter (DOM) using the method reported elsewhere. Briefly, all diluted samples were acidified with concentrated HCl (Ultrapure Regent, Kanto Chemical, Japan) at pH ~2 and then gravitationally passed through Bond Elut PPL cartridges (1g and 6 mL, Agilent) which were activated and rinsed with 12 mL methanol (MeOH, LC-MS grade, Kanto Chemical, Japan) and 6 mL Milli-Q water, respectively. The cartridge was then rinsed with 20 mL HCl (pH ~2.0) and 6 mL Milli-Q water to desalt and remove residual Cl$^-$, respectively, followed by complete drying using N$_2$ gas (99.9% gas purity). DOM was finally eluted with 6 mL MeOH and diluted twofold with Milli-Q water. Separately, two SRNOM standard solutions (200 mg-C/L) were prepared by dissolving SRNOM in Milli-Q H$_2$O and deuterium oxide (D$_2$O), respectively (referred to as H-SRNOM and D-SRNOM) and used to examine exchange of labile H/D in NOM molecules. The H-SRNOM and D-SRNOM solutions were further diluted twofold with MeOH and MeOD (99.5% D, Sigma-Aldrich, USA), respectively.

Additional chlorination treatment (Treatment D) was performed to examine the applicability of newly developed algorithm (namely FTMSDeu) to the formula assignment of L$_n$-DBPs by using FTICR-MS/MS and network analysis. To this end, the sample was prepared at concentrations of
2.5 mg-C/L for SRNOM, 50 mg/L for ClO⁻ and 200 mg/L for I⁻, and incubated for a week at room

temperature under the dark. In this sample, high I⁻ concentration was employed to generate high-

intensity for Iᵦ-DBPs in the FTICR-MS/MS analysis. The samples were subjected to SPE-based

DOM extraction by using aforementioned procedure.

All samples were stored in the dark at 4°C and filtered through a 0.22 μm PVDF membrane

prior to FTICR-MS measurements.

**FTICR-MS measurement.** All samples were measured by the FTICR-MS instrument

equipped with a 9.4 T superconducting magnet system (Solarix XR, Bruker) and electrospray

ionization (negative ion mode, -ESI) at Tohoku University, Sendai, Japan. All FTICR-MS spectra

were measured with the following instrumental conditions: -4.5 kV capillary voltage; 150 μL/h

direct infusion rate; 2 megaword time-domain data size; 450 average scans; 1 ms ion accumulation;

150 -1,500 mass-to-charge ratio (m/z) range, and > 200,000 resolving power (m/z = 399). Parent

ions for the Treatment D at nine nominal masses (267, 311, 373, 407, 445, 477, 485, 559, and 6230

were isolated at 1 Da mass windows and fragmented in the quadrupole using the collision-induced

dissociation by argon gas. The fragmentation spectra were recorded in the same FTCR MS

instrument with 100 average scans and 2 megaword time-domain data size. The collision voltage

and ion accumulation time were adjusted to obtain optimal fragmentation spectra (Table S1). Prior

to the measurement, the FTICR-MS instrument was rinsed by the deuterated solvent (MeOD +

D₂O) for D-SRNOM sample and normal solvent (MeOH + H₂O) for the other samples to prevent

the possible exchange of H/D between DOM molecules and residual solvents in the instrument ¹⁸.

All FTICR-MS and FTICR-MS/MS spectra were externally calibrated with ion clusters using the

NaI solution before measurement and internally recalibrated with known CHO-homologous series
of freshwater DOM to achieve a mass accuracy < 1.0 ppm for the entire spectrum during post-data processing.\textsuperscript{12,40}

**Algorithm description.** The FTMSDeu algorithm was developed based on our previous NOMDBP Code\textsuperscript{40} by incorporating D in the formula assignment and extending the formula assignment capability to Cl- and Br-free solutions containing I (referred to as Org-I\textsubscript{n} hereafter). The FTMSDeu algorithm is executed with the flow depicted in Figure 1. Briefly, after inputting a calibrated UHR-MS spectral information (\(m/z\), intensity, signal-to-noise ratio [S/N]), all chemically possible solutions are calculated for each \(m/z\) according to following calculation conditions: \textit{i.e.}, (i) mass error tolerance (typically 1.0 ppm), (ii) maximum number of element, (iii) maximum number of D, (iv) DBE minus O rule\textsuperscript{12}, and (v) nitrogen rule. Then, unlikely solutions are filtered based on the \(^{13}\text{C}\)-isotopic pattern with the acceptable intensity error tolerance (30 % relative to the theoretical value). For a given \(m/z\), if all filtered solutions are halogen-free, then the optimum solution is selected in the first scenario (typically suitable for NOM) with the precedence of (i) minimum number of N+S+P, (ii) minimum number of S+P, (iii) D\(\leq\)O rule (optional rule only for D-labeled UHR-MS spectra), and (iv) minimum error. Otherwise, all filtered solutions are inspected in the second scenario, where organohalogen formulae containing non-oxygen heteroatoms must have sufficient intensity (\textit{e.g.}, S/N\(\geq\) 10), and Org-I\textsubscript{n} formulae must be restricted to \(m/z\) in the range from its nominal value minus 0.4 to plus 0.02 (namely the empirical I\textsubscript{n}-DBPs mass distribution rule). Then, the effective candidates of organohalogen formula are determined including (i) Org-I\textsubscript{n}, (ii) organohalogen formulae solely containing single Cl or Br (Org-Cl\textsubscript{1} or Org-Br\textsubscript{1}), and (iii) organohalogen formulae with multiple numbers of Cl+Br using our previously proposed rules (\textit{i.e.}, precursor and new peak appearance inspection\textsuperscript{40} and Cl and Br isotopic pattern validation). The optimum formula for this given \(m/z\) among non-halogen and
organohalogen formulae is subsequently selected with the priority of (i) maximum number of Cl+Br (only for organohalogen candidates with multiple numbers of Cl+Br), (ii) minimum number of N+S+P, (iii) minimum number of S+P, and (iv) minimum error. Once all monoisotopic peaks are assigned to unequivocal formulae, all unassigned peaks and assigned peaks in both scenarios are combined to assign isotopic formulae for $^{13}$C, $^{18}$O, $^{34}$S, $^{37}$Cl, and $^{81}$Br based on their isotopic patterns of natural abundances with an acceptable intensity error of 30%. Some important molecular parameters (e.g., (H+D)/C, O/C, X/C, DBE, AI mod, and NOSC) are also calculated and exported together with formula results.

**Figure 1.** The FTMSDeu algorithm flow. Red words indicate new functions compared with our previous NOMDBP Code.
It is worth noting that some false positive solutions can be caused by the incorporation of D for Org-Xₙ formula assignment due to the close mass difference of C₂H₃O₄ versus D₆Br₁ (Δm/z=0.18 mDa), C₄H₅O₅ versus D₃N₃Br₁ (Δm/z=0.02 mDa), C₁D₃ versus ¹³C₁H₅ (Δm/z = 0.17 mDa), and C₄D₁₀ versus H₂O₃ (Δm/z= 0.22 mDa). For example, peaks at m/z= 306.945864 and 307.952127 have multiple Org-Xₙ solutions within 1.0 ppm mass error (C₉H₉O₇Br₁ versus C₇H₆D₆O₃Br₂ and C₉H₈D₁O₇Br₁ versus C₇H₅D₇O₃Br₂, respectively). However, due to the obviously different isotopic patterns for the Org-Br₁ and Org-Br₂ formulae, as exemplified in Figure S2, the true positive formulae (i.e., C₉H₉O₇Br₁, and C₉H₈D₁O₇Br₁) can be assigned to m/z 306.945864 and 307.952127, respectively. The isotopic pattern (in this case for Br) is, therefore, an effective tool to solve the formula assignment issue of C₂H₃O₄ versus D₆Br₁. Analogously, for m/z= 313.056496, the candidate formula C₉H₁₁D₅O₄N₃Br₁ can be declined due to the absence of ⁸¹Br isotopic peak in the identical UHR-MS spectrum (Figure S3), and the true positive formula, C₁₃H₁₄O₉ can be ultimately assigned, which is further validated by the minor error (3%) for intensity ratio between measured intensity ratio of ¹³C isotope and its theoretical value.

The utilization of ¹³C isotopic pattern can also solve the formula assignment issue of C₁D₃ versus ¹³C₁H₅, when the monoisotopic peaks have sufficiently high intensity. For example, if ¹³C₁C₁₉H₇D₇O₅ was assigned for m/z= 341.124221 (RA=7.59%) in the UHR-MS spectrum of D-SRNOM, there must be a distinct monoisotopic peak at m/z= 340.1207855 with a theoretical relative abundance (RA) of 35.09% (Figure S4). This formula is, therefore, found to be a false positive solution for m/z= 341.124221. However, there are still two candidate formulae without non-oxygen heteroatom C₂₀H₂D₁₀O₅, and C₁₆H₂₂O₈ calculated for this ion (i.e., C₄D₁₀ versus H₂O₃). The carboxylic and hydroxylic functional groups (-COOH and -OH) are the major moieties containing labile H in NOM molecules¹³, suggesting that the number of D in NOM
molecules is less than the number of O under the D₂O system. Also, the hydroxylation (e.g., UV
irradiation treatment) can be important mechanism that incorporates labile or non-labile OH into
aromatic molecules 45-47, and the number of D becomes less than O number for NOM molecules
under the hydroxylation with D₂O. Therefore, the D ≤ O rule (D number ≤ O number) is
incorporated in the FTMSDeu algorithm to assign formulae for D-labeled UHR-MS spectra of
NOM. By introducing the D ≤ O rule, the true positive formula, C₁₆H₂₂O₈, is finally assigned to
$m/z= 341.124221$. The D≤ O rule is also supported by the fact that NOM is rich in refractory
carboxyl-rich alicyclic molecules (CRAM) with the compositional space of DBE/C = 0.30- 0.68,
DBE/H = 0.20- 0.95, and DBE/O = 0.77- 1.75 21. For $m/z= 341.124221$, C₂₀H₂D₁₀O₅ is far from
the restricted area of CRAM, while C₁₆H₂₂O₈ (DBE/C=0.38, DBE/H=0.28, and DBE/O=0.75) is
close to its empirical area border. A formula assignment flow was exemplified in Figure S5 for the
FTICR-MS spectra for Treatment B, D-SRNOM in D₂O, and Treatment D (the parent ions at $m/z= 306.945864$, 341.121157, and 432.942712, respectively).

Data analysis. Formula assignment was conducted by the FTMSDeu algorithm using the
following calculation conditions: $S/N \geq 6$ and $\geq 10$ for non-halogenated and halogenated
monoisotopic formula, respectively; $0.3 \leq (H+Cl+Br+I)/C \leq 2.25$ and $0 < O/C \leq 1.2$ for molecule
with C ≥ 5, (H+Cl+Br+I)/C ≤ 4 and 0 ≤ O/C ≤ 1.2 for molecule with C ≤ 4; an integer value ≥ 0
for double bond equivalent (DBE); 1≤ ¹²C ≤ 50; 0 ≤ D ≤ 10 for chlorinated or non-chlorinated
SRNOM in D₂O and D = 0 for H-SRNOM; ¹³C ≤ 2; ¹⁸O ≤ 1; -10 ≤ DBE-O ≤ 10; ¹⁴N ≤ 5; ³²S ≤ 3;
³³S ≤ 1; P ≤ 1; ³⁵Cl ≤ 5; ³⁷Cl ≤ 5; ⁷⁹Br ≤ 5; ⁸¹Br ≤ 5; and I=0 and ≤ 5 for all chlorinated treatments
without and with I-, respectively. One H was assumed to be lost during the negative ESI process
for all treatments, expect for D-SRNOM, in which one D was lost. Thus, one H or D was added to
calculate the neutral formula for the relevant samples. The assigned formulae were classified into
eight biochemical groups in the van Krevelen diagram based on the reported criteria. Xₙ-DBPs precursor herein was defined as the molecule or molecule moiety that forms Xₙ-DBPs via electrophilic substitution and/or electrophilic addition. The precursor of a given Xₙ-DBPs formula \((C_xH_yO_zD_kX_l)\) was estimated as \(C_xH_{y+l-k}O_{z-k}\) according to stoichiometric changes of electrophilic substitution and electrophilic addition. The relative contribution of electrophilic substitution and electrophilic addition for a given Xₙ-DBPs molecule (ContSub₁ and ContAdd₁, respectively) and all Xₙ-DBPs molecules (ContSub₂ and ContAdd₂, respectively) were quantified by Eqs. (S1)-(S4) in Content S1. The DBE, modified aromaticity index (AImod), the nominal oxidation state of carbon (NOSC), and the intensity-weighted values of molecular parameters were calculated with Eqs. (S5)-(S8) in Content S2. The homologous series of all Iₙ-DBPs were also inspected with an in-house algorithm based on (i) Iₙ-DBPs formulae validated by FTICR-MS/MS, (ii) common NOM formula building blocks (i.e., H₂, H₂O, C, CH₂, CO₂, and CO), and (iii) building blocks representing electrophilic substitution of iodination (i.e., mass of I minus H, I-H) and electrophilic addition of iodination (i.e., mass of I plus H, I+H). The network diagram was visualized by Gephi software. Principal component analysis (PCA) was conducted with MATLAB using the molecular parameters tabulated in Table S2.

RESULTS AND DISCUSSION

Labile H in SRNOM. While H-SRNOM and D-SRNOM samples shared a similar spectral profile in the overall UHR-MS spectra (Figure S6A), the spectrum for D-SRNOM was more complicated than that for H-SRNOM due to deuteration of labile H in SRNOM molecules. The discrepancy of peak intensity was more apparent for D-SRNOM at even nominal masses than that for odd nominal mass. H-SRNOM peaks at the even nominal masses had lower intensities than
those for D-SRNOM (Figures S6B and S6C). The former peaks were attributed mostly to the $^{13}$C-isotopologues and to lesser extent to the compounds containing even number of N, while the higher intensity of latter peak was assigned to peaks for deuterated compounds with an odd number of D. Consistent with the previous observation 13, the presence of multiple numbers of labile H in SRNOM (Figure S7) resulted in about twofold increase in the number of assigned peaks for D-SRNOM compared to that for H-SRNOM with 2 to 6 number of labile H (Figure S8).

Nonetheless, it is noteworthy that, during negative-ESI ionization, a few D-SRNOM molecules that have lost one labile H are hard to be distinguished from more abundant D-SRNOM molecules and are considered as molecules losing one labile D for the number estimation of liable D. This was also supported by the relatively small intensity (10%) of molecules that have mostly lost one labile D during negative-ESI ionization (Figure S8). Lignin-like and tannic-like compounds accounted for 67.7% and 28.5% of D-SRNOM molecules, respectively, and generally had more labile D than other types of D-SRNOM molecules (Figure S9). The number of labile D linearly increased with increasing average values of O number and O/C ratios for D-SRNOM molecules ($R^2=0.944$ and 0.968, respectively, Figure S10). This result further supported the hypothesis that carboxyl and hydroxyl functional groups were the predominant contributors of labile H (or D) for SRNOM 16-18. Furthermore, Figure S10A revealed the presence of O-containing function groups irrelevant to labile H (such as the carbonyl or ether group) 13 for D-SRNOM molecules with number of labile D being no more than seven.

**Identification of I$_n$-DBPs.** Regarding I-containing compounds, 1,436 unequivocal I$_n$-DBPs formulae were identified by our FTMSDeu algorithm for the Treatment D. Also, unique Cl$_m$I$_n$-DBPs was also detected and validated with the Cl isotopic pattern (C$_2$H$_3$O$_2$Cl$_1$I$_2$ in Figure S11). In the FTICR-MS/MS spectra of parent ions at nine selected nominal masses, the distinct I' peak at
m/z=126.9050165 was detected, confirming the presence of organo-iodine compounds in these nominal masses (Table S1) \(^{48,49}\). Fragment ions with neutral losses of I radical (I·, 126.904468 Da) and HI (127.912293 Da) were also identified in these FTICR-MS/MS spectra (mass error tolerance <1.0 ppm, Table S1, and Figure S12). For example, nearly all parent ions with low m/z (e.g., <560) had lost a single or multiple numbers of I· or HI.

There were 411 nodes and 388 edges identified in the homologous series inspection of all I\(_n\)-DBPs in the Treatment D, revealing that 411 unique I\(_n\)-DBPs compounds had direct homologous connections to FTICR-MS/MS-validated I\(_n\)-DBPs formulae. As exemplified in Figure S13, another 13 I\(_n\)-DBPs formulae were supported by C\(_7\)H\(_4\)O\(_2\)I\(_2\), which was validated by FTICR-MS/MS, and typical blocks including CO\(_2\), H\(_2\)O, CO, C, CH\(_2\), and +I-H. Furthermore, 178 unequivocal I\(_n\)-DBPs were computed under the calculation conditions. Totally, 589 I\(_n\)-DBPs compounds identified in the aforementioned two scenarios were considered to be highly reliable and accounted for 41.0% and 59.2% of the total number and intensities, respectively, for all 1,436 assigned I\(_n\)-DBPs formulae in the Treatment D. The equivocal solutions for 300 other peaks were caused by the close mass differences of H\(_4\)I\(_1\) versus C\(_3\)O\(_2\)S\(_1\)P\(_1\) and H\(_4\)I\(_1\) versus C\(_4\)O\(_1\)S\(_1\)35Cl\(_1\) (\(\Delta m/z=0.11\) and 0.07 mDa, respectively). However, the solutions of C\(_3\)O\(_2\)S\(_1\)P\(_1\) and C\(_4\)O\(_1\)S\(_1\)35Cl\(_1\) were rejected in our algorithm as a result of (i) the absence of detectable 37Cl-isotopic peaks if presence \(^{40}\), (ii) deficiency of non-oxygen heteroatom in NOM precursors \(^{29}\) and (iii) the selected I\(_n\)-DBPs formula having more moderate degree of saturation than C\(_3\)O\(_2\)S\(_1\)P\(_1\)-containing solution. For example, the non-oxygen heteroatom-free I\(_n\)-DBPs formula, C\(_9\)H\(_9\)O\(_7\)I\(_1\) with moderate degree of saturation (DBE=5), was attributed to peak at m/z 354.932139 rather than the non-oxygen heteroatom-containing unsaturated formulae C\(_{13}\)H\(_5\)O\(_8\)S\(_1\)Cl\(_1\) and C\(_{12}\)H\(_5\)O\(_9\)S\(_1\)P\(_1\) (DBE=11). For similar reason, I\(_n\)-DBPs formulae with a minimum number of non-oxygen heteroatom were
assigned to the other 547 peaks, and most of them (>91%) had $m/z$ values >500. Therefore, in addition to reliable Cl$_n$-DBPs and Br$_n$-DBPs formulae, our FTMSDeu algorithm can automatically assign I$_n$-DBPs formulae with high accuracy, providing an useful tool for non-targeted screen of halogenated organic compounds in the complex organic mixtures.

The enlarged UHR-MS spectrum in Figure 2 indicated the applicability of the I$_n$-DBPs mass distribution rule (i.e., I$_n$-DBPs ions locate in the mass window of nominal value minus 0.4 to plus 0.02) due to the significant mass deficiency of $^{127}$I isotope compared with its nominal mass (126.904468-127=-0.095532) and the mass window of NOM ions from nominal value to nominal value plus 0.3. It should be noted that some I$_n$-DBPs formula containing non-oxygen heteroatoms such as sulfur and nitrogen (namely, CHOSI and CHONI, respectively) were identified in the UHR-MS spectrum for the treatment D. The typical CHOSI (C$_2$H$_1$O$_3$S$_1$I$_3$) was detected at $m/z$ 484.670894 (Figure 2) and was confirmed to contain sulfo group (Figure S12D) which can be attributed to sulfotriiodoethylene. Sulfur-bearing X$_n$-DBPs has been proposed to be generated by the bromination of surfactant degradation products in seawater 50 and wastewater 51, and chlorination of CHOS compounds in secondary effluent 52. Because of the low abundance of CHOS molecules in SRNOM (6.5% of total intensity), CHOSI species herein can be generated by the reaction with dehalogenation agent (Na$_2$SO$_3$) used to terminate the chlorination. Analogous to halomethane ions (e.g. [CHBr$_2$]$^+$ and [CHClBr$_2$]$^+$ detected in the Treatment B 40), the newly discovered oxygen-free [C$_3$N$_2$I$_3$]$^+$ ion in Figure 2 (which was validated by FTICR-MS/MS, Table S1) is unlikely to be ionized in negative ESI mode and therefore this molecule is most likely generated by the dissociation of C$_3$N$_2$I$_3$ moiety from the large parent CHONI molecule during ionization process in ESI section 40.
Figure 2. The enlarged UHR-MS spectrum for the Treatment D at the nominal masses of 445 and 485.

Characteristics of Xₙ-DBPs formed in the D₂O system. As illustrated in Figure 3, the deuterated Xₙ-DBPs species were detected with high resolution by the FTICR-MS technique. The measured UHR-MS spectra for deuterated and non-deuterated DBPs species containing Cl and/or Br were highly close to their theoretical spectra (Figures 3, S14 and S15), suggesting the high accuracy of our FTMSDeu algorithm in assigning both deuterated and non-deuterated Xₙ-DBPs formulae. The FTMSDeu algorithm had identified 1,573, 1,025, and 1,623 unique Xₙ-DBPs species in Treatments A, B, and C, respectively. Compared to Clₙ-DBPs, Br- or I- containing Xₙ-DBPs species tended to be deuterated during the disinfection process, suggesting that electrophilic
addition contributed to larger extent in the formation of Br- or I-bearing $X_n$-DBPs species than
electrophilic substitution. HOCl could readily oxidize Br$^-$ and I$^-$ to HOBr (which is more reactive
with NOM molecules)$^{53}$, and thermodynamically unstable HOI ($5\text{HOI} = 2\text{I}_2 + \text{IO}_3^- + \text{H}^+ + 2\text{H}_2\text{O}$,
which is less likely to participate in $X_n$-DBPs formation)$^{54}$, respectively. Therefore, $Br_n$-DBPs and
$Cl_m$-DBPs were observed to be the predominant $X_n$-DBPs species (accounting for 74.5% and 94.6%
of all DBPs species in Treatments B and C, respectively, Figure S16). The formula number and
intensity for $Br_n$-DBPs were substantially larger than those for $Cl_m$-DBPs in Treatment B (Figure
S16), suggesting that $Br_n$-DBPs are effectively formed via oxidation of Br$^-$ by active chlorine
species followed by reactions of HBrO with DOM molecules. Unlike our previous results which
showed that more $X_n$-DBPs was yielded in the SRNOM+ClO$^-$+Br$^-$ treatment than the
SRNOM+ClO$^-$ treatment$^{40}$, the proportion of total $X_n$-DBPs intensity in Treatment B of this study
was approximately 80% of that in Treatment A. This discrepancy could be related to different
initial solution pH values (8.9 and 7.5 in this and previous study, respectively). As an effective
disinfectant, ClO$^-$ is the predominant chlorine species in Treatment B, but this Cl form is six orders
of magnitude less reactive with Br$^-$ compared to HClO, yielding BrO$^-$ at a much slower rate for
$Br_n$-DBPs formation$^{25}$. $X_1$-DBPs and $X_2$-DBPs compounds were the dominant $X_n$-DBPs species
in all three treatments (Figure S17) due to the passivating effect of halogen atom for successive
reception of halogen atoms during $X_n$-DBPs formation$^{55}$. Furthermore, $X_n$-DBPs species could be
well distinguished by PCA analysis with three components (Figure S18) using the intensity-
weighted molecular parameters tabulated in Table S2. The $Br_n$-DBPs and $Cl_mI_n$-DBPs in
Treatments B and C were separated from other $X_n$-DBPs species by principle component 1, which
was related to the intensity-weighted number of carbon ($C_{nw}$). Principle component 3 associated
with the degree of unsaturation and number of D was capable of differentiating Cl$_n$-DBPs in Treatment A and Treatment C from others.

![Theoretical spectrum](image)

**Figure 3.** Representative measured and theoretical UHR-MS spectra of C$_5$H$_3$O$_4$Cl$_3$ and C$_5$H$_2$D$_1$O$_4$Cl$_3$. The measured UHR-MS spectrum for Treatment A.

**Relative contribution for X$_n$-DBPs formation.** The relative contributions of electrophilic addition and electrophilic substitution to X$_n$-DBPs formation in all treatments were calculated based on the intensity of all identified X$_n$-DBPs formulae (Table 1). Consequently, the electrophilic substitution was estimated to be the predominant mechanism for Cl$_n$-DBPs formation in Treatments A-C, contributing to 81.8% - 98.4% of Cl$_n$-DBPs formation, while the occurrence of HClO addition on unsaturated moieties of DOM molecules were less likely under the conditions examined due to the relatively lower rate of this pathway. The PCA result also supported that Cl$_n$-DBPs showed higher degree of unsaturation and fewer number of D than the other X$_n$-DBPs.
species (Figure S19). Compared to Clₙ-DBPs, the relative contribution of electrophilic addition increased by approximately 20% for Br-containing Xₙ-DBPs species (Brₙ-DBPs and ClₘBrₙ-DBPs, see Treatment B in Table 1), but electrophilic substitution was still the major contributor in Br-containing Xₙ-DBPs formation. In contrast, electrophilic addition played more critical roles in I-containing Xₙ-DBPs species (Iₙ-DBPs and ClₘIₙ-DBPs accounting for 56.7% and 55.2%, respectively) formation than electrophilic substitution. The increasing in the relative contribution of electrophilic addition for Clₙ-DBPs, Brₙ-DBPs, and Iₙ-DBPs is consistent with the increase in the dissociation constants of hypohalous acid (Figure 4, R²=0.90, p<0.05, note that pKₐ are 7.50, 8.63 and 10.4 for HClO, HBrO, and HIO, respectively) and decrease in the electronegativity of halogens (Figure 4, R²=0.82, p<0.05, electronegativity are 3.16, 2.96, 2.66, and 3.51 for Cl, Br, I, and O, respectively). This observation was supported by the proposed reaction steps of electrophilic addition, where X⁺ atom in hypohalous acid initially transferred to the unsaturated carbon bound to provide a halonium ion, forming transitional halohydrin products, followed by subsequent addition of remaining OH⁻. The lower halogen electronegativity is favorable for halohydrin formation, and consequently formed Xₙ-DBPs more via electrophilic addition. Moreover, OCl⁻, OBr⁻ and HOI were the initially dominant hypohalous acid species (80.3%, 66.6%, and 99.8%, respectively) under the condition examined (Figure S19). Therefore, the electrophilic addition potential of HXO could be in order of HIO > HBrO > HClO, suggesting that reaction pathways were determined by the halogen electrophilicity and acidity of reactive hypohalous acid species. Furthermore, the contribution of electrophilic addition for Xₙ-DBPs species is consistent with their degree of toxicity: i.e., Iₙ-DBPs > Brₙ-DBPs > Clₙ-DBPs. This finding highlights the importance of electrophilic addition in terms of formation of toxic Xₙ-DBPs species.
Table 1. The relative contribution of electrophilic addition and electrophilic substitution to X-n-DBPs formation in different treatments.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>X-n-DBPs Species</th>
<th>X-n-DBPs Peaks</th>
<th>Relative Contribution (%)</th>
<th>Electrophilic Addition</th>
<th>Electrophilic Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment A (D₂O+ClO⁻)</td>
<td>Clₙ-DBPs</td>
<td>4223</td>
<td>1.6</td>
<td>98.4</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>4223</td>
<td>1.6</td>
<td>98.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment B (D₂O+ClO⁻+Br⁻)</td>
<td>Brₙ-DBPs</td>
<td>2051</td>
<td>11.2</td>
<td>88.7</td>
<td></td>
</tr>
<tr>
<td>ClₙBrₙ-DBP</td>
<td>309</td>
<td>28.8</td>
<td>71.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>2620</td>
<td>12.9</td>
<td>87.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment C (D₂O+ClO⁻+I⁻)</td>
<td>Iₙ-DBPs</td>
<td>73</td>
<td>56.7</td>
<td>43.4</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>3661</td>
<td>3.1</td>
<td>96.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Graph](image_url)

- Exponential fitting for electronegativity
- Exponential fitting of pKₐ
**Figure 4.** Relationship between relative contributions of electrophilic addition for Cl$_n$-DBPs, Br$_n$-DBPs, and I$_n$-DBPs and $pK_a$ values of HOCl, HOBr, and HOI, respectively.

**X$_n$-DBPs precursors.** Putative precursors of X$_n$-DBPs compounds were determined by assuming that X$_n$-DBPs was formed via electrophilic addition and substitution reactions. In the calculation, only stoichiometric changes associated with electrophilic reactions (but not secondary reactions) were considered. Similar profiles were observed for van Krevelen diagram from three treatment conditions (*i.e.*, Treatments A-C, Figures 5A-5C), suggesting that the majority of X$_n$-DBPs compounds were derived from the halogenation of similar NOM precursors *(Figure 5D)* with active hypohalous acid species. Treatments A and C shared 1,622 X$_n$-DBPs (84.5% and 89.5% number of all X$_n$-DBPs species, respectively) and 930 X$_n$-DBPs precursors (88.9% and 87.9% number of total X$_n$-DBPs precursors, respectively). This result combined with van Krevelen diagram profile generally indicates that the majority of X$_n$-DBPs was originated from the electrophilic substitution of reactive chlorine with lignins- and tannins-like DOMs molecules with O/C=0.4-0.9 and H/C=0.5-1.5. Since the electrophilic substitution is dominant for Cl$_n$-DBPs in Treatments A and C (the contributions were to more than 97%), it can be reasonable to visually characterize Cl$_n$-DBPs precursors by replacing H/C with (H + Cl)/C in the van Krevelen diagram of chlorinated waters containing low concentrations of Br$^-$ (*e.g.*, fresh groundwater)\(^{40}\). The plotted molecules on the $y$-axis (*i.e.*, O/C = 0) could be either X$_n$-DBPs species where halogen was electrophilically added to oxygen-free NOM molecules (*e.g.*, C$_3$H$_7$D$_3$O$_2$Cl$_3$, C$_4$H$_5$D$_2$O$_2$Br$_2$, and C$_6$H$_2$D$_2$O$_2$Cl$_1$I$_1$ in Treatments A, B, and C, respectively) or electrophilically substituted oxygen-free X$_n$-DBPs species (*e.g.*, halomethane in the Treatment B). Putative N-containing precursors were only identified in the Treatments B and C *(Figure S20A)* and were responsible for more toxic N-containing X$_n$-DBPs species than N-free X$_n$-DBPs compounds \(^{26}\). Putative precursors were
sometime not detected in some treatments (Figure S20B and Table S3), which could be due to the secondary reactions including oxidation and hydrolysis of electrophilically added and/or substituted Xₙ-DBPs species and thus considered as putative precursor moieties for secondary Xₙ-DBPs formation. The fact that 613 Xₙ-DBPs precursors were identified in all three treatments suggests that non-selectivity of NOM molecules toward different reactive hypohalous acid species in Xₙ-DBPs formation. Moreover, considerable proportions (32.0%-55.0%) of putative precursors were involved in formation of multiple (different) unique Xₙ-DBPs species in the three treatments (Figure S20C), suggesting that the formation of Xₙ-DBPs is highly complex and diverse even when originating from the same NOM molecule. Compared to Treatment A, more unsaturated hydrocarbon and unsaturated lignins-like precursors were exclusively scattered in the left area of van Krevelen diagrams (O/C=0-0.3 and H/C=0.3-1.5) for the other two treatments. These precursors are not preferentially removed by typical treatments such as granular activated carbon adsorption and metal coagulation in drinking water treatment and are responsible for formation of more toxic Brₙ-DBPs and Iₙ-DBPs, highlighting the necessity of removing these precursors in drinking water systems.
Figure 5. The van Krevelen diagrams of estimated $X_n$-DBPs precursors for Treatment A (A), Treatment B (B), and Treatment C (C); and the Venn diagram of all estimated unique $X_n$-DBPs precursors (D). RA(%) is the relative abundance of $X_n$-DBPs monoisotopic peaks.

**Limits and Future direction**

The main limitation of this study is the spontaneous exchange between -OD added to the backbone structure of NOM (as the backbone structure) and surrounding H$_2$O molecule during SPE extraction process. Such loss of D during the post-treatment (in case D$_2$O is not used for all the chemicals used) potentially underestimate the contribution of electrophilic addition to $X_n$-DBPs formation. The electrophilical addion of -OD to backbone structures will increase their...
saturation degree, yielding less labile -OD than that with higher degree of unsaturation (i.e., phenolic -OD). Moreover, given that OCl\(^-\), OBr\(^-\), and HOI are dominant active hypohalous acid species at the initial pH employed, this issue may be minor for Treatment A but non-negligible for Treatment C because H/D exchange in skeleton generally occurs in the presence of catalysis (or under high temperature) \(^{19}\). This issue, however, can be satisfactorily solved by using D-\(^{18}\)O dual-isotope labeling, where D and \(^{18}\)O are added in precursors together with halide atoms for electrophilic addition, and only halide atoms substitute with H atoms in precursors for electrophilic substitution.

\(X_n\)-DBPs electrophilically added in the aromatic skeleton are characterized by identical numbers of \(^{18}\)O and halide atoms; and \(X_n\)-DBPs electrophilically added in unsaturated side chains contain identical numbers of D, \(^{18}\)O, and halide atoms. Thus, UHR-MS techniques coupled with D/\(^{18}\)O isotope labeling could be useful in elucidating \(X_n\)-DBPs formation mechanisms. At the same time, further updates of our FTMSDeu algorithm are necessary for UHR-MS spectra labeled with \(^{18}\)O, \(^{13}\)C, and other isotopes with development of associated new filtering rules. For example, both C\(_6\)H\(_6\)O\(_2\)Cl\(_2\) and C\(_7\)H\(_6\)O\(_3\)\(^{18}\)O\(_1\)N\(_2\)Cl\(_2\) are within the 1.0 ppm mass error tolerance for the peak at \(m/z\) = 252.967 559 due to the close mass difference between CH\(_2\)O\(_2\) and N\(_2\)\(^{18}\)O\(_1\) (\(\Delta m/z\) = 0.17 mDa).

CONCLUSIONS

In this study, for the first time, the FTMSDeu algorithm was successfully developed for D-labeled UHR-MS spectra and employed to automatically assign chemical formulae for organoiiodine (filtered with newly proposed I\(_n\)-DBPs mass distribution rule), organochlorine, and organobromine. Its assignment accuracy for organo-iodine compounds was further validated with FTICR-MS/MS technique and homologous-based network analysis. It was found that the number of labile D in SRNOM molecules linearly increased with their O content and O/C ratios, suggesting...
that labile D is attributed to the O-containing active functional groups (e.g., -COOH and -OH).

The relative contributions of electrophilic addition and substitution were dependent on the halogen species involved in the reactions, and the solution pH and pK_a values for hypohalous acids, as well as type of halogen, could be important parameters. Under the conditions examined in this study, the electrophilic substitution was the predominant mechanism for Cl_n-DBPs and Br-containing X_n-DBPs, while electrophilic addition becomes significant in the formation I-containing X_n-DBPs.

The secondary reactions of electrophilically added and/or substituted X_n-DBPs species were indirectly supported by few putative precursor moieties. The UHR-MS technique coupled with isotope labeling was of significant importance in revealing the transformation of DOM in natural and engineered systems such as water and wastewater treatments. Overall, our FTMSDeu algorithm has laid valuable basis for further developing formula assignment for UHR-MS spectra labeled with the other isotopes.

ASSOCIATED CONTENT

Supporting Information

Calculation of contribution and intensity weighted DBE, AI_{mod} and NOSC; FTICR-MS/MS spectra and ions; molecular parameters for PCA analysis; commonly estimated X_n-DBPs precursors; representative measured and theoretical UHR-MS spectra; broadband and expended UHR-MS spectra; relative intensity percentage of D-SRNOM molecules; Van Krevelen diagrams; Relationship between labile D number and average oxygen number and O/C; network plot for C_{7}H_{4}O_{2}I_{2}; unique formula number, intensity proportions, PCA results of all X_n-DBPs; Speciation distribution of hypohalous acid.
ACKNOWLEDGEMENTS

This study was financially supported by the Japan Society for the Promotion of Science (NO. 17H04588, 19H02271), the Special Funding for Discipline Construction from China University of Geosciences (NO. 162301212611) and the National Natural Science Foundation of China (No. 42107484). The authors appreciate the technical support from Hiroyuki Momma in FTICR-MS measurement.

REFERENCES

(45) Xiang, Y. Y.; Fang, J. Y.; Shang, C. Water Res. 2016, 90, 301-308.


