Development and Application of High-Precision Algorithm for Non-Target Identification of Organohalogens Based on Ultrahigh-Resolution Mass Spectrometry

Qing-Long Fu[1], Manabu Fujii[1]*, Eunsang Kwon[2]

[1] Department of Civil and Environmental Engineering, Tokyo Institute of Technology, Ookayama, Tokyo, Japan.

[2] Research and Analytical Center for Giant Molecules, Graduate School of Science, Tohoku University, Sendai, Japan.

ABSTRACT

The brominated and/or chlorinated organic compounds (referred to as organohalogens) are frequently detected in natural and engineered environments. However, the ultrahigh-resolution mass spectrometry (UHR-MS)-based non-target identification of the organohalogens remains challenging due to the presence of vast number of halogenated and non-halogenated organic molecules in the same aqueous sample. In this study, a new algorithm, namely NOMDBP Code, was developed, based on natural organic matter (NOM) chemistry, to simultaneously identify organohalogens and non-organohalogens from the UHR-MS spectra of natural and engineered waters. In addition to isotopic pattern extraction, for the first time, three optional filter rules (namely selection of minimum non-oxygen heteroatoms, inspection of newly formed halogenated disinfection byproducts [X-DBPs] and precursors) were incorporated in our code, which can accurately identify DBPs-associated peaks and further elucidate the X-DBPs generation and transformation mechanisms. The formulae assignment rate against previously reported 2,815 unique organohalogens and their 11,583 isotopologues was determined to be >97%. Application of our algorithm to disinfected NOM indicated that oxygen-containing X-DBPs species accounted for a majority of X-DBPs. Further, brominated X-DBPs (Br-DBPs) during disinfection process were characterized by higher degree of unsaturation compared to chlorinated X-DBPs (Cl-DBPs). Our algorithm also suggested that, in addition to electrophilic substitution and electrophilic addition reactions, the decomposition/transformation is another important mechanism in Br-DBPs formation. Results of this study highlight the superior potential of this code to efficiently detect yet-

unknown organohalogens (including organohalogens with non-oxygen heteroatoms) in a non-target manner and identify their generation mechanism during the disinfection process.

INTRODUCTION

Halogenated organic compounds (organohalogens, in this study referred to as chlorine and/or bromine-containing organic matters formed via natural and artificial processes) are widely present in the natural and engineering environments. In urban wastewater and drinking water treatment systems, chemical disinfection has been often used to inactivate pathogens and micropollutants. Natural organic matters (NOM), being ubiquitous and chemically active and very diverse in aquatic environments, however, can react with hydrolyzed products of disinfectants (*e.g.* HOCl), yielding toxic halogenated disinfection byproducts (X-DBPs, X= Cl and/or Br) [1-4]. To determine the DBPs molecular species, ultrahigh-resolution mass spectrometry (UHR-MS) including Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) was for the first time introduced in 2012. Since then, approximately 2,800 unique high-molecular weight X-DBPs formulae have been identified in freshwater NOM, seawater, and lysed cyanobacterial cells after disinfection [1,3,5-8], though complete picture of X-DBPs including "missing fraction of X-DBPs" remains inconclusive yet [9].

To detect organohalogens using mass spectrometry, researchers typically employ isotopic pattern representing unique peak combinations of multiple isotopes with different masses and natural abundances (Meusel et al., 2016). Given the high natural abundance of Cl (75.78% for ^{37}Cl and 24.22% for ^{35}Cl) and Br isotopologues (50.69% for ^{79}Br and 49.31% for ^{81}Br), the isotopic pattern of Cl and Br (*e.g.* chlorinated X-DBPs [Cl-DBPs], brominated X-DBPs [Br-DBPs], and Cl- and Br-containing X-DBPs [Cl-Br-DBPs]) is a distinctive signal in organohalogens detection with high levels of confidence. In the past two decades, therefore, isotopic patterns had been successfully employed to selectively detect X-DBPs

peaks from UHR-MS spectra [10,11] as well as to identify the X-DBPs and other organic pollutants [2,12-18], non-halogenated metabolites [19] and some unknown biomolecules [20]. Further, the mass spectrometry with combination of precursor ion scans (PIS) has been proposed to selectively detect Cl-DBPs and Br-DBPs [10,11]. While the PIS-based methods are powerful and rigorous in identification of X-DBPs, the analysis using collision-induced dissociation as well as measurements for non-halogenated compounds are typically entailed to identify the precursor and product ions, which requires the additional measurement cost.

To comprehensively understand the X-DBPs occurrence and generation mechanism in natural and engineered waters, high-throughput non-target analysis using UHR-MS is attracting attention. While it remains challenging to detect a number of organohalogen peaks in a non-target manner and concurrently assign the correct formulae to the thousands of UHR-MS peaks for X-DBPs and other molecules, several algorithms have been developed by integrating the molecular formula assignment algorithm with the isotopic pattern detection for Cl-DBPs. For example, chemical formulae assignment software such as ProfileAnalysis software [13] and SigmaFit™ algorithm [15] were integrated with isotope detection algorithm and, by using the precompiled formula library, applied to the measured peaks of emerging organic contaminants, drugs (and their metabolites) and transformation products. Further, an isotopic pattern algorithm as well as pre-constructed X-DBPs isotopic library has also been incorporated into a recently developed software, namely Formularity [21], though its application to the Br-DBPs and Cl-Br-DBPs compounds may be currently limited due to the presence of only five Br-DBPs formulae in the isotopic database. The isotopic patterns of Cl and Br were also incorporated into the newly announced ICBM-

OCEAN, and assignment performance and accuracy for organohalogens can be examined in future studies [22].

In contrast to the library-based methods, the SIRIUS software (widely used for metabolites identification) utilizes isotopic patterns to filter multiple potential solutions for the given mass to only a few candidates, and directly determine the X-DBPs with multiple Cl or Br atoms [19]. However, monohalogen formulae (*i.e.*, containing sole Cl or Br) as well as Cl-Br-DBPs are not incorporated in SIRIUS, regardless of their frequent observation in chlorinated samples [3-5,17,19,20]. The compatibility of this biomolecule-customized SIRIUS to identify X-DBPs molecules therefore needs to be investigated. Moreover, due to the lack of ad hoc filtering rules (*e.g.* O/C and, H/C ratios) for NOM molecules in SIRIUS, a manual 'post-processing' step was necessary to assign chemical formulae to all chlorinated NOM peaks [19].

While there are several algorithms developed, the interpretation of UHR-MS spectra containing thousands of molecular peaks can be largely affected by the analytical resolution and algorithm performance capable of distinguishing the organohalogen peaks from non-halogenated NOM peaks and concurrent assignment of chemical formulae. In particular, if the disinfection of natural waters is a concern, formula assignment of halogenated and non-halogenated organic molecules (including yet-unknown organohgalogens with not only CHO-type but also including non-oxygen heteroatoms) should be consistent with NOM chemical properties (as a precursor of X-DBPs) as well as the associated X-DBPs generation mechanisms. Therefore, in this study, a new method based on the halogen isotopic patterns with combination of precise formula assignment algorithm considering NOM chemistry is developed and applied to the naturally and anthropogenically

halogenated waters. A significant revision was made toward the TRFu algorithm, which had been developed to efficiently and automatically assign chemical formulae to the UHR-MS peaks of NOM at high accuracy [23]. By incorporating several algorithms to detect isotopic pattern as well as to filter and assign DBPs formulae, a novel algorithm for X-DBPs analysis, namely NOMDBP Code, was developed. Formula assignment accuracy of this updated algorithms was evaluated using the reported DBPs formulae [1,3,5,6] and their simulated ^{37}Cl, ^{81}Br, and ^{13}C isotopic formulae. The developed NOMDBP Code was subsequently applied to the artificially chlorinated waters and natural water samples to screen known and unknown organohalogens among thousands of compounds in a non-target manner.

METHODS AND MATERIALS

X-DBP formulae library for assignment test

In this study, 2,815 X-DBPs compounds (Table S1) previously validated by isotopic patterns from halogenated Suwannee River fulvic acid (SRFA) [1], treated drinking water [3], aquaculture seawater [6], and lysed cyanobacterial cells [5] were retrieved from literature. Further, their isotopologues were generated with Eqs. (S1) and (S2) as described in Content SI1. Therefore, DBP formulae library in this study (totally 14,398 peaks) includes the reported 2,815 unique formulae (2,303, 471, 33, 6 and 2 for Cl-, Br-, Cl-Br-, Cl- I-, and Br-I-formulae, respectively, labeled as “parent formulae”), 4,306 isotopologue formulae generated for ^{37}Cl and/or ^{81}Br (labeled as “major isotopologue”) and 7,277 isotopologue formulae generated for ^{13}C and/or ^{34}S (labeled as “minor isotopologue”). The FT-ICR-MS

spectrum of SRFA containing 9,753 peaks was also combined with the X-DBP formulae library. The SRFA were used as a representative of non-halogenated NOM sample to examine the possible influence of non-halogenated NOM peaks on the assignment performance of the NOMDBP Code. The X-DBP formula library built for the assignment test were descripted in Content SI2.

Sample preparation

In the laboratory, X-DBPs compounds were generated in 3.0 mg-C/L Suwannee River NOM (SRNOM [1R101N] purchased from International Humic Substances Society [IHSS]) solutions amended with 90 mg/L $NaHCO_3$ and 5.0 mg-Cl_2/L NaClO (free chlorine concentration was determined by colorimetry using *N*, *N*-diethyl-*p*-phenylenediamine [DPD] with a Pocket Colorimeter II, SIBATA, Japan). In the test where Br-DBPs were generated, potassium bromide (KBr) were additionally amended at concentration of 2.0 mg-Br/L by following the method detailed elsewhere [2]. After adjusting pH ~7.5 by concentrated hydrochloric acid (HCl, Ultrapur, Kanto Chemical, Japan), the halogenated water samples were incubated for a week at room temperature in the dark.

Moreover, a seawater sample was collected in Tokyo Bay, Japan (Odaiba Seaside Park at 35.62986 °N, 139.77405 °E), and filtered with 0.45 μm membrane immediately after transport to laboratory. Similarly, freshwater samples from the upstream (35.74427 °N, 139.31692 °E) and downstream (35.58394 °N, 139.67025 °E) of Tama River in Tokyo, Japan (abbreviated by Tama-up and Tama-down, respectively) were also collected and filtered. The natural water samples were subjected to the extraction and purification of

dissolved organic matter (DOM) using the solid phase extraction (SPE) with the method described elsewhere [24]. Briefly, filtered water samples were acidified with concentrated HCl to pH 2 and fed through the styrene divinylbenzene polymer cartridge (Bond Elut PPL, Agilent) activated with 6 mL methanol (LC/MS grade). Then, the cartridge was rinsed with the 20 mL HCl-acidified ultrapure water to completely remove salts, followed by N_2-drying for 5 min. Then, DOM was eluted with 1 cartridge volume of methanol with gravity flow. Subsequently, these samples were diluted with the identical volume of ultrapure water.

In order to evaluate the possible formation of Cl-adduct ions during the extraction and electrospray ionization (ESI) processes, Suwannee River Fulvic Acid II (SRFAII) and Nordic Reservoir NOM (NRNOM) (2S101F and 1R108N purchased from IHSS, respectively) were subjected to the identical SPE procedures (that include the treatment using HCl) prior to the FT-ICR-MS measurement. Separately, DOM solutions (that omit SPE treatment) were prepared at final concentration of 250 mg-C/L by dissolving the desired amount of Upper Mississippi River NOM (MRNOM), Suwannee River Fulvic Acid (SRFA), Nordic Lake Fulvic Acid (NLFA) and Pony Lake Fulvic Acid (PLFA) (1R110N, 1S101F, 1R105F, and 1R109F purchased from IHSS, respectively) into ultrapure water, followed by dilution with the identical volume of methanol. Finally, all DOM samples were filtered with 0.22 μm membrane and kept at 4 ℃ in the dark prior to the measurement.

2.3 FT-ICR-MS analysis

A Bruker SolariX 9.4T FT-ICR-MS interfaced with negative ion mode electrospray ionization was applied to characterize DOM samples in Tohoku University, Sendai, Japan. Instrumental conditions were set as follows: -4.5kV capillary voltage; 150 μL/h direct infusion rate; 2 MWord data size; 1,000 average scan; 0.1 s ion accumulation; 150-1,500 mass to charge ratio (*m*/*z*) range [25]. The resolving power was ~ 200,000 at *m/z* = 401. All spectra were externally calibrated with sodium iodine ion clusters (*m/z*= 276-3,125) and internally recalibrated with known CHO-homologous series of freshwater DOM (*m/z*= 200-600 with >200 mass calibration points) to obtain a mass accuracy <1.0 ppm. For further identification of organohalogen in the Tokyo Bay seawater, parent ions of seawater DOM ions at the nominal mass of 431 (2 Da mass windows for ion isolation) were fragmented in the quadrupole via collision induced dissociation with argon gas. The fragmentation spectra were recorded in the same FT-ICR-MS instrument with 100 scans, 2 MWord data size, 15.0 V collision voltage, and 1 s ion accumulation time.

Organohalogen peaks were detected from the Cl and Br isotopic patterns and formulae were assigned simultaneously with non-halogenated DOM compounds using our Matlab-based in-house NOMDBP Code (see details in next section). The parent peaks with relative abundance (RA) ⩾ 0.5% were only considered in the analysis, where the parent peak was defined by the mono-isotopic peak for the most abundant isotopologue (in terms of RA) among the multi-isotopic peaks with same halogenated chemical formula but different isotopic composition. Formulae were assigned to the peaks with signal to noise ratio ≥4 under the following conditions: $0.3 \leq X/C \leq 2.25$ (where X = H+ Cl+ Br), $0 < O/C \leq 1.15$ ($0 \leq O$ when $C \geq 5$), double-bond equivalent (DBE) ≥0, $1 \leq {}^{12}C \leq 50$, $0 \leq {}^{13}C \leq 2$, $0 \leq {}^{18}O \leq 1$, $N \leq 5$, ${}^{32}S \leq 3$, $0 \leq {}^{34}S \leq 1$, and $P \leq 1$, ${}^{35}Cl \leq 5$, ${}^{37}Cl \leq 5$, ${}^{79}Br \leq 5$, ${}^{81}Br \leq 5$. Some molecular

characteristics including O/C, H/C, X/C, modified aromaticity index (AI_{mod}), DBE, DBE/C, nominal oxidation state of carbon (NOSC) were also calculated in the algorithms. Since Cl-DBPs compounds are formed most likely via the electrophilic substitution, where hydrogen atoms are usually displaced by halogen atoms [3,26], van Krevelen diagrams of O/C versus *X*/C were employed to visually characterize the putative DBPs precursors in this study.

Description of the NOMDBP Code

The NOMDBP Code is an updated version of our Matlab-based TRFu Code [23]. Besides all fundamental algorithms for NOM characterization in the TRFu Code, the following new algorithms have been incorporated into the NOMDBP Code: *i.e.*, (i) parent peak determination for Cl-DBPs, Br-DBPs, and Cl-Br-DBPs based on the distinctive pattern of isotopologue peaks, (ii) selection of the optimum Cl-DBPs, Br-DBPs, and Cl-Br-DBPs formulae, and (iii) assignment of DBPs major and minor isotopologue formulae. Briefly, as depicted in Fig. 1, after input of UHR-MS data and calculation conditions and subsequent extraction of dataset one by one, all theoretically possible solutions (formulae) are calculated for each peak. For a given peak, if all solutions are Cl- or Br-free, these solutions are directly filtered with the optimum formula selection rules detailed previously [23]. Otherwise, the multi-isotopic patterns of all Cl- and/or Br-containing formulae are inspected in succession. Once the multi-isotopic pattern is detected (with acceptable error range of *m/z* and RA, relative to theoretical values, being 0.0035 and 30%, respectively), mono-halogen formulae are optionally filtered with the rule of minimum non-oxygen heteroatoms (Optional Filter Rule 1, given the dominance of CHO-type X-DBPs in

literature). In addition to the optional filter that inspect the presence of organohalogen peaks in the sample where halogenation treatment was omitted (Optional Filter Rule 2), another optional filter, namely DBPs precursor inspection (Optional Filter Rule 3), can also be executed in the NOMDBP Code, which improves the formulae assignment accuracy for organohalogens, particularly, those containing more than four carbon atoms.

After the Cl- and/or Br-containing formulae are confirmed, the optimum formula for the peak is selected with priority of (1) maximum number of Cl + Br (formulae with more Cl+ Br are more reliable due to the presence of more ^{37}Cl and/or ^{91}Br peaks in the UHR-MS spectra), and (2) maximum number of oxygen (note that precursors with high O/C can be preferentially chlorinated than low O/C by disinfection [1,3]). Otherwise, all Cl- or Br-containing formulae are deleted, and the remaining Cl- and Br-free formulae are further filtered with the optimum formula selection rules or set as unassigned peak. Then, after combining all peaks together, ^{37}Cl and ^{81}Br formulae are reassigned on the basis of multi-isotopic patterns. Subsequently, ^{13}C isotopic formulae are reassigned according to the ^{12}C-^{13}C isotopic pattern. Similarly, ^{34}S isotopic formulae are further assigned to daughter candidates with the ^{32}S-^{34}S isotopic pattern. Finally, the formulae result and summary are automatically exported as Microsoft Excel files. Then, UHM-MS data of next sample are successively extracted and aforementioned calculations were repeated.

Notably, due to the close mass differences between ^{32}S and ^{34}S ($\Delta m/z$=1.995796), ^{35}Cl and ^{37}Cl ($\Delta m/z$=1.997050), and ^{79}Br and ^{81}Br ($\Delta m/z$=1.997953), multi-isotopic peaks with the same number of heavy isotope are generally overlapped in the UHR-MS spectra. For example, the ^{37}Cl-isotopic peak of $C_9H_{18}O_6{}^{37}Cl_1{}^{79}Br_1$ will be completely overlapped by the ^{81}Br-isotopic peak of $C_9H_{18}O_6{}^{35}Cl_1{}^{81}Br_1$ due to its higher abundance compared to

$C_9H_{18}O_6{}^{37}Cl_1{}^{79}Br_1$. For such case, the intensities of overlapped peaks in the UHR-MS spectra are mixed and theoretically equals to the sum intensity of all overlapped isotopologues. In this study, therefore, formulae of the overlapped peaks generally represent the mono-isotopic formulae with the highest intensities.

Data analysis

The accuracy ratio was defined by the number of true positive (correctly assigned) formulae relative to the total number of X-DBPs examined. The one-way analysis of variance was employed to compare different datasets at the significant level of $p<0.05$ using MATLAB 2019b (MathWorks, U.S.). The principle component analysis (PCA) was performed by MATLAB using the intensity weighted DOM molecular parameters (*m/z*, H/C, O/C, S/C, Cl/C, Br/C, AI_{mod}, DBE, NOSC, and X/C, Table S2) for all identified organohalogens. Each parameter was standardized by subtracting its mean value, and subsequently dividing the standard deviation using the function, zscore, in MATLAB in order to weigh the contribution of each parameter to the principal components equally. All assigned formulae were categorized with the criteria described elsewhere [23].

RESULTS AND DISCUSSION

Formula assignment performance

The isotopic pattern-based formula assignment performance of NOMDBP Code was examined by using following three peak datasets: (i) the previously reported 2,815 X-DBPs peaks (parent peaks + major isotopologue peaks); (ii) all the 14,398 X-DBPs peaks (parent

peaks + major and minor isotopologue peaks) and (iii) all the 14,398 X-DBPs peaks + 9,753 non-DBP peaks (*i.e.*, SRFA, which was used to examine the effect of non-DBP peaks on the assignment performance). The assignment ratios were determined to be 99.0%, 98.0%, and 97.3% for the first, second, and third scenario, respectively, generally demonstrating the excellent assignment performance of the NOMDBP Code for X-DBPs formulae. The slight decreasing trend in assignment ratios from the first to third scenarios potentially suggests the adverse effect of some other adjacent peaks (that are close to minor X-DBPs isotopologue peaks) on the X-DBPs peak selection and following formulae assignment. Generally, assignment performance was found to be better when the intensity for the relevant isotopic peaks were higher.

To obtain further insights into the reason of false assignment, the false positive assignment of X-DBPs formulae was carefully examined. Inspection of assigned formulae indicated that false positive X-DBPs formulae were generally detected due to the false assignment of X-DBPs formulae rather than false detection of isotopic patterns. For example, different formulae ($C_{11}H_{14}O_6N_4Cl_2$ and $C_{17}H_{10}O_2N_2Cl_1Br_1$) were assigned to the published peaks of $C_{13}H_{19}O_7{}^{81}Br_1$ and $C_{14}H_{11}O_8{}^{81}Br_1$, respectively, because formulae with multiple halogen atoms were, in this case, preferentially selected with the Selection Rule 1 (*i.e.*, maximum Cl +Br, see Fig. 1). When all the solutions include multiple halogens, formulae with minimum non-oxygen heteroatoms can be selected by the NOMDBP Code (see Selection Rule 3 in Fig. 1), leading some false positive assignments especially for peaks with high *m/z* (*e.g.*, *m/z*>600 and Cl+ Br=2 in Fig. 2A). Some examples include false assignment of $C_{19}H_{46}O_{16}N_6Cl_2$, $C_{27}H_{58}O_{10}N_6Cl_2$ and $C_{28}H_{60}O_{10}N_6Cl_2$ (in the X-DBP

formulae library) to less non-heteroatom formulae, $C_{27}H_{50}O_{13}S_1Cl_2$, $C_{35}H_{62}O_7S_1Cl_2$ and $C_{36}H_{64}O_7S_1Cl_2$, respectively.

The assignment accuracy is generally mass-dependent, and decreases largely with increasing *m/z* values due to the ascending number of possible formulae, particularly for *m/z* > 600 [23]. In this study, consistently, the false positive ratio of X-DBPs parent formulae gradually increased from 0% to 3.5% for *m/z* from 200 to 600 in the second scenario dataset, followed by an obvious increase over 600 *m/z* (Fig. 2C). Among all DBPs solutions for a given parent peak, the formula with higher Cl+ Br was more reliable, because the formula was validated with higher number of ^{37}Cl- and/or ^{81}Br- isotopic peaks than formula with less Cl+ Br. The parent formula with higher Cl+ Br was therefore selected at higher priority in the NOMDBP Code (Selection Rule 1 in Fig. 1), which resulted into the decreasing false positive ratio with increasing number of Cl+ Br (Fig. 2B). These results demonstrate reasonably high level of formula assignment accuracy by the NOMDBP Code for X-DBPs compounds, particularly for X-DBPs formulae with lower mass or multiple number of Cl and/or Br.

It should be noted that the assignment performance of NOMDBP Code was evaluated based on the detection of isotopic patterns and subsequent formulae assignment for detected X-DBPs peaks considering NOM chemical properties. The three optional filer rules (Fig. 1) were disabled in the assignment performance test, primarily because the filtration processes such as precursor inspection and new peak appearance are unnecessary in the test using the published peak library. Nonetheless, as discussed in detail below, executing such optional filter rules is critical in detecting X-DBPs related peaks, assigning the correct formula and identifying the major mechanism of X-DBPs generation from

thousands of organohalogen and non-organohalogen peaks present in the artificially disinfected and natural waters.

Organohalogen characteristics

Freshwater samples

Generally, CHO and lignin-like compounds were the predominant fractions for freshwater DOM and same trend were observed in this study for all freshwater DOM from Tama River and IHSS (Content SI3). It was noteworthy that some peaks in FT-ICR-MS spectra followed ^{35}Cl-^{37}Cl and/or ^{79}Br-^{81}Br isotopic patterns; thus such peaks were considered as organohalogens regardless of the absence of chlorination treatment (note that we further confirmed the presence of organohalogens by applying the MS/MS technique to the selected peak in a target manner, as described below). As shown in Table S3, organohalogen formulae accounted for 0.35%-3.64% of total formulae. It is worth noting that, all of these formulae contained only one halogen, and 80% formulae had a low abundance (*e.g.*, RA< 5%) and 95% formulae contained non-oxygen heteroatoms. The presence of Cl-containing compounds detected in non-chlorinated freshwater can be due to (i) naturally chlorinated compounds, (ii) adducts of chloride ions remaining during the acidification and desalination steps of DOM extraction using HCl and (iii) unknown sources [3]. Due the low concentration of bromide in freshwater (typically ranging from trace level to ~0.5 mg/L [2]), the sources of Br-containing compounds remain unknown. The false positive ratios for monohalogens were relatively higher than formulae containing multiple number of halogen. Thus, some formulae bearing one halogen for non-chlorinated

freshwater may be generated by the false positive assignment of non-halogenated peaks (that resemble the Cl or Br isotopic peak pattern) in the FT-ICR-MS spectra composed of several thousands of peaks.

The optional DBPs precursor inspection rule (Optional Filter Rule 3 in Fig. 1), was executed to further validate the presence of halogenated compounds in the natural samples. The Filter Rule 3 is proposed according to assumption that X-DBPs are formed via electrophilic substitution where H atoms in X-DBPs precursors are substituted by the identical number of halogen atoms [3,26]. After executing the Optional Filter Rule 3, the total number of organohalogen formulae in the non-halogenated freshwater samples decreased substantially ($p<0.05$) to 0-0.35% relative to total assigned formulae. As depicted in Fig. S2, the precursors-inspected natural halogen compounds generally had high H/C and moderate O/C ratios (H/C=1.3-1.6, O/C=0.3-0.5) and were dominated by lignin-like and protein-like molecules.

While suspicious halogens could be eliminated by the optional precursor inspection rule (*i.e.*, Filter Rule 3), the identification of Cl-halogens can be affected by Cl-adduct ions potentially formed during the ESI negative mode ionization via interaction of DOM molecules and remaining Cl^- from the SPE-extraction using HCl [27]. In this study, the potential adduct of Cl-molecules were observed for the SPE-extracted SRNOM and NRNOM, while not for the SPE-extracted SRFAII (Fig. S4). These results suggest that the formation of Cl-DOM adducts is controlled not only by the presence of remaining Cl^-, but also by the types of DOM. In order to avoid adduct formation, formic acid has been suggested in the SPE extraction instead of HCl [16-18,27,28]. However, the employment of other type of acid, formic acid, in SPE extraction (instead of HCl) may lead to the occurrence of

formate adduct ions in ESI negative mode [29-31], interfering the formula assignment of non-halogenated DOM molecules. Therefore, an additional SPE cartridges washing with ultrapure water is alternatively suggested to completely remove remaining Cl^- or formate ions in the case of simultaneously analyzing halogenated and non-halogenated DOM compounds. Because rigorous formula assignments are central in X-DBPs characterization, only the X-DBPs formulae treated with precursor inspection (Filter Rule 3) were discussed for freshwater DOM in the following sections.

Seawater sample

Regarding the general trend of Tokyo Bay seawater DOM, the lignin-like compounds and CHO compounds were the predominant fractions of non-halogenated compounds (accounting for 75% and 51% of all intensity, respectively, Fig. 3A), which is similar to the results of freshwater samples (*e.g.* Tama River upstream DOM and SRNOM in Figs. S5 and S6). However, the percentage of CHOS compounds (*i.e.*, compounds only containing C, H, O, and S, 20.1%) was significantly higher ($p<0.05$) in the seawater DOM than that of all freshwater DOM examined (3.4%±3.2%), indicating the higher level of dissolved organic sulfur (DOS) in the coastal seawater compared to natural freshwater *(e.g.* Tama River upstream DOM) and potential importance of such DOS in the coastal sulfur cycle [32].

The high abundance of CHOS compounds can also be attributed to the sulfur-containing surfactants in reclaimed water from municipal wastewater treatment plants (WWTPs). Indeed, some typical surfactants (*e.g.* $C_{11}H_{14}(CH_2)_nO_5S_1$ and $C_{12}H_{14}(CH_2)_nO_5S_1$, n= 1- 4) in the Eawag surfactants suspect list for wastewater [33] were detected in the

seawater with RA values ranging from 17.2% to 100% (Fig. S7), highlighting the significant impacts of anthropogenic activities to coastal water qualities. The seawater sampling station Odaiba in the Tokyo Bay typically receives WWTPs effluents from the Shibaura and Ochiai Water Reclamation Center.

Regarding the organohalogens, 18 unique organohalogens containing 1, 3, and 4 atoms of Cl were identified in the formulae in the Tokyo Bay seawater DOM (Table S4). Sixteen organohalogens were non-oxygen heteroatom-free compounds, while the other 2 organohalogens contained sulfur and phosphorus. The most abundant organohalogen in the Tokyo Bay seawater was identified as $C_{12}H_{20}O_8Cl_4$, followed by $C_{12}H_{19}O_8Cl_3$ (RA = 24.2% and 8.5% for the parent peak, respectively, Fig. S8). The formula, $C_{12}H_{20}O_8Cl_4$, was further validated by the FT-ICR-MS/MS technique. As depicted in Fig. 3B, the parent ion at *m/z* 430.98398 ($[M]^-$) was observed to continuously lose two HCl, yielding two conspicuous fragment ions at *m/z* 395.0073 and 359.03065 ($[M\text{-}HCl]^-$ and $[M\text{-}2HCl]^-$, respectively). While typical NOM fragment ions are formed via CO_2, CO and/or H_2O release from parent ions in the MS/MS spectra [34-36], these fragments were not identified in the parent ion *m/z* 430.98398, suggesting the absence of acidic (*e.g.*, carboxyl) functional group in the structure of the parent ion. The $[M\text{-}HCl]^-$ ion intensity was approximately ten times higher than that for the $[M\text{-}2HCl]^-$ ion, suggesting the dominant loss of one HCl in the parent ion *m/z* 430.98398.

The adduct Cl is loosely bound to precursors, which is in contrast to Cl directly bound to carbon skeletons in DOM structure. Provided that the ion at *m/z* 430.983997 ($[C_{12}H_{19}O_8Cl_3+Cl]^-$) is the Cl-adduct ion of $[C_{12}H_{19}O_8Cl_3\text{-}H]^-$ at *m/z* 395.007286 for the Tokyo Bay seawater DOM, this organohalogen ($C_{12}H_{19}O_8Cl_3$) was assigned to the artificial

sweetener, sucralose [37], which had been detected in Hong Kong coastal water (average level =0.06 μg /L) [38], Swiss lake surface water and wastewater effluent (average level= 44-3,641 ng/L) [39], and U.S. coastal and marine waters (average level= 2.0-394 ng/L) [40]. Due to its highly persistent property under typical operational conditions of municipal WWTPs [41], sucralose has been employed as a chemical tracer for domestic wastewater [42]. Thus, both sulfur-containing surfactants and sucralose in the Tokyo Bay are most likely originated from WWTPs effluent or potentially combined sewer overflow.

Chlorinated SRNOM samples

X-DBPs compounds were newly formed by the chlorination treatment of DOM sample (*e.g*. new peaks clearly observed in Figs. 4 and S9). Although precursor inspection is a useful filter rule for the accurate detection of X-DBPs in disinfected waters, it should be noted that precursors for the "secondary X-DBPs" (formed by the decomposition of precursor via multiple degradation pathways) [43,44] were difficult to be validated in the halogenated samples due primary to their multiple formation mechanisms. It has been reported that the major of carbonyl DBPs contains ketoacid structure, which are generated most likely due to the phenyl ring-cleavage of halogenated macromolecular phenols, such as lignin-derived moieties in SRNOM [45]. For example, regardless of the successful validation by isotopic patterns, the FT-ICR-MS peaks for $C_8H_7O_3Br_1$, derived from the cleavage of the phenyl ring in lignin-derived compounds [45], failed to be assigned to any Br-DBP formulae due to the absence of direct precursors in the chlorinated SRNOM if only considering precursors inspection. Similarly, no X-DBPs formulae were assigned to $C_7H_4O_2Br_2$ peaks in the same sample when only precursor inspection was considered.

Such issue can be, at least partially, solved in the NOMDBP Code by introducing the new rule (Optional Filter Rule 2 in Fig. 1), where the presence of target X-DBPs peaks in the control spectrum (*e.g.* spectrum of water prior to the disinfection treatment) is inspected. The Optional Filter Rule 2 only works for disinfected samples when the control spectrum is available. New peaks validated successfully by isotopic patterns in the chlorinated DOM are most likely due to the presence of newly formed X-DBPs compounds, which are thus considered as reliable X-DBPs formulae in this study. As visually depicted in Figs. S10-S11, X-DBPs formulae, $C_1H_2Br_2$, $C_2H_2O_2Br_2$, $C_7H_3O_3Br_3$, and $C_{10}H_7O_5Cl_1$, could be successfully assigned to the secondarily formed DBPs compounds with the new peak rule. However, the execution of this optional rule should be avoided for the formula assignment of naturally occurring organohalogens due to the lack of defined control samples.

In this study, X-DBPs compounds in chlorinated SRNOM samples were examined by executing all optional filter rules. Totally, 146 and 240 unique X-DBPs (excluding isotopologues) tabulated in Table S4 were identified in the chlorinated SRNOM in the absence and presence of bromide, respectively. Consistent with previous findings [1,2], only few Br-DBPs with multiple Br atoms were observed in chlorinated SRNOM with bromide (*e.g.*, 23 and 1 formulae for Br-DBPs containing 2 and 3 bromine atoms, respectively). The less number of Br-DBPs formulae with multiple bromine atoms (compared to mono-halogenated X-DBPs) could be attributed to the passivating role of bromine, which can impede their successive receive of another bromine by the mono-brominated X-DBPs [46]. Similarly, the absence of Cl-DBPs with multiple chlorine atoms in chlorinated SRNOM (in the absence of bromide addition) can be caused by the decreased chlorination rate for

the mono-chlorinated DBPs (which is formed by incorporating chlorine into the reactive precursors through electrophilic aromatic substitution) [28].

While Cl-DBPs and Br-DBPs accounted for ~99% of all X-DBPs compounds in the SRNOM sample with "ClO^-" and "ClO^- + Br^-" treatments, respectively, the proportion of total X-DBPs compounds in the "ClO^-" treatment was ~61% of that for the "ClO^-+Br^-" treatment. The high abundance of X-DBPs species for the SRNOM sample chlorinated in the presence of bromide can be consistent with the fact that free bromine, generated from the rapid reaction of free chlorine with bromide, was more effective in X-DBPs formation than free chlorine [26,47]. The previous study also indicated that Br-DBPs formation is strongly related to bromide concentrations in waters to be disinfected [46]. In contrast to the predominant formation of Cl-DBPs during chlorination of river and reservoir waters typically containing low levels of bromide, Br-DBPs formation is reported to dominate for the other types of environmental waters such as groundwater and coastal seawater containing relatively high Br concentrations [17,18,28]. Disinfection treatment of waters containing high level of bromide (*e.g.* >0.28 mg/L in groundwater) tends to yield more X-DBPs and results into the predominance of highly toxic Br-DBPs species [48], posing severe risks to human and aquatic creatures [28]. These results suggest that an additional X-DBPs removal is necessary in the case of disinfection treatments for groundwater and seawater.

Most DBPs molecules were identified together with precursors of the electrophilic substitution in this study. Similarly, the electrophilic addition reaction accounted for 1 Cl-DBPs and 2 Br-DBPs compounds in this study (Table S4), indicating the reliability of detection for DBPs compounds generated in these reaction pathways. Nonetheless, 24 Br-DBPs compounds (Table S4) were found to miss their precursors assuming electrophilic

substitution or electrophilic addition reactions. Although the precursors were absent in the spectrum of chlorinated SRNOM, these 24 newly formed Br-DBPs were identified in the chlorinated sample using the Optional Filter Rule 2. This phenomenon may be associated with secondary formation of X-DBPs, and suggest the potentially critical role of secondary reaction in X-DBPs formation. For example, three typical secondary X-DBPs (CH_2Br_2, $C_2H_2O_2Br_2$, and $C_8H_7O_3Br_1$), often identified as dibromomethane, dibromoacetic acid, and 3-bromo-4-hydroxy-5-methoxybenzaldehyde, respectively [16], were detected as secondary products via the pathways including hydrolysis in water treatment systems and swimming pools [43,49-52]. Compared to the absence of secondary formed species in the "ClO^-" treatment, 24 of secondary Br-DBPs species in the "ClO^-+Br^-" treatment signifies the more complex formation mechanisms of Br-DBPs than Cl-DBPs. Furthermore, 15 X-DBPs formulae with one phosphorus were identified in this study. Among them, 11 formulae were validated by the precursor inspection and the remaining 4 were assigned to newly formed compounds (without direct precursors). Since these compounds were not recorded in chemical database such as PubChem, the MS/MS fragment analysis should be employed to further validate formula assignment accuracy and elucidate their structures that are essential to their toxicity evaluation.

Non-halogenated byproducts are often formed during the disinfection treatment, revealing the mechanism of X-DBPs formation [44,53,54]. In addition to non-halogenated byproducts with low molecular weights detected with conventional chromatographic and mass techniques, the NOMDBP Code is capable of simultaneously assigning formulae to non-halogenated byproducts with larger molecular weights in the FT-ICR-MS spectra. Totally, 102 and 243 newly formed and chlorination-derived putative non-halogenated

byproducts were identified in this study (Table S5), including two reported non-halogenated byproducts decomposed from X-DBPs, $C_8H_6O_4$ and $C_8H_6O_5$, [44]. The newly formed $C_8H_6O_5$ was absent in non-chlorinated SRNOM. In case of $C_8H_6O_4$, the peak for this molecule was detected in non-chlorinated SRNOM, the chlorination treatment showed substantially high RA in both "ClO^-" and "ClO^-+Br^-" treatments (*e.g.*, by 2.7 and 3.6 folds, respectively). Therefore, the comparison of formulae results exported from the NOMDBP Code for chlorinated and non-chlorinated samples provides additional information in elucidating the formation and transformation of halogenated and non-halogenated products during disinfection process.

Moreover, the comparison of X-DBPs and non-DBPs organohalogens discussed in Content SI4 reveals higher degree of unsaturation for Br-DBPs compared to Cl-DBPs, signifying that the reactive hypobromite preferentially reacts with highly unsaturated lignin-like precursors under the conditions examined. Environmental implication discussed in Content SI4 highlights the powerful potential of the NOMDBP Code in in non-target screening of some unique and yet-unknown X-DBPs species with non-oxygen heteroatoms, as well as non-halogenated aromatic byproducts, in drinking water and WWTPs effluent.

CONCLUSIONS

In this study, the NOMDBP Code has been successfully developed and applied to simultaneously identify halogenated and non-halogenated organic compounds from the UHR-MS spectra of natural and engineered waters. Compared to the conventional PIS-based methods for X-DBPs identification, our automated algorithm could eliminate the

additional UHR-MS measurement cost for non-halogenated molecules measurement, minimize the time-consuming post data analysis procedure, and also efficiently output high quality formula results with the assignment accuracy up to 97.3% for reported X-DBPs compounds. In addition to the precise detection of multi-isotopic patterns, three optional filter rules proposed in this study are designed, for the first time, according to the chemical views of X-DBPs generation and thus are favorable to identifying organohalogens, practically for disinfected waters. The NOMDBP Code is, therefore, notable to produce the true positive solutions for X-DBPs species containing both oxygen and non-oxygen heteroatoms, highlighting its superior potential compared to the conventional X-DBPs studies [2,18,28].

ASSOCIATED CONTENT

Supporting Information

Supporting contents, tables, and figures include: theoretical isotopic pattern calculation; X-DBPs formula library description; freshwater DOM characteristics; comparison of X-DBPs and non-DBPs; environmental implication; X-DBPs formulae and their mass distribution and van Krevelen diagram; DOM molecular parameters for PCA analysis and PCA results; Effect of precursor inspection on Cl- and Br-containing formulae identification; formula and RA of identified organic halogens and their relevant putative precursors; putative non-halogenated byproducts; van Krevelen diagrams for naturally halogenates, formulae assigned by the TRFu Code and the NOMDBP Code, non-halogenated formulae in freshwater samples, and Cl-DBPs and Br-DBPs; representative putative Cl-adducts and typical surfactants; expanded FT-ICR-MS spectra.

AUTHOR INFORMATION

Corresponding Author

*E-mail: fujii.m.ah@m.titech.ac.jp. Tel: +81-3-5734-3687. Fax: +81-3-5734-3577.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

This study was financially supported by the Japan Society for the Promotion of Science (JSPS, Grants No. P17374, 17H04588, 19H02271). The authors also appreciate the technical support from Hiroyuki Momma in the FT-ICR-MS measurement and Prof. David Bastviken from Linköping University and Prof. Dongmei Zhou from Nanjing University for their valuable discussions in manuscript editing.

References

(1) Zhang, H. F.; Yang, M. *Sci. Total Environ.* **2018**, *627*, 118-124.
(2) Zhang, H. F.; Zhang, Y. H.; Shi, Q.; Zheng, H. D.; Yang, M. *Environ. Sci. Technol.* **2014**, *48*, 3112-3119.
(3) Lavonen, E. E.; Gonsior, M.; Tranvik, L. J.; Schmitt-Kopplin, P.; Kohler, S. J. *Environ. Sci. Technol.* **2013**, *47*, 2264-2271.
(4) Zhang, H. F.; Zhang, Y. H.; Shi, Q.; Hu, J. Y.; Chu, M. Q.; Yu, J. W.; Yang, M. *Environ. Sci. Technol.* **2012**, *46*, 4396-4402.
(5) Gonsior, M.; Powers, L. C.; Williams, E.; Place, A.; Chen, F.; Ruf, A.; Hertkorn, N.; Schmitt-Kopplin, P. *Water Res.* **2019**, *155*, 300-309.
(6) Wang, J.; Hao, Z. N.; Shi, F. Q.; Yin, Y. G.; Cao, D.; Yao, Z. W.; Liu, J. F. *Environ. Sci. Technol.* **2018**, *52*, 5662-5670.
(7) Shen, X.; Shao, Z. J.; Xian, Q. M.; Zou, H. X.; Gao, S. X.; Zhang, J. Q. *Water Res.* **2010**, *44*, 974-980.
(8) Richardson, S. D.; Temes, T. A. *Anal. Chem.* **2018**, *90*, 398-428.

(9) Richardson, S. D. *Trac-Trends in Analytical Chemistry* **2003**, *22*, 666-684.
(10) Zhang, X. R.; Talley, J. W.; Boggess, B.; Ding, G. Y.; Birdsell, D. *Environ. Sci. Technol.* **2008**, *42*, 6598-6603.
(11) Yang, M. T.; Zhang, X. R.; Liang, Q. H.; Yang, B. *Water Res.* **2019**, *158*, 322-337.
(12) Avagyan, R.; Westerholm, R. *Talanta* **2017**, *165*, 702-708.
(13) Chen, W. L.; Cheng, J. Y.; Lin, X. Q. *Sci. Total Environ.* **2018**, *637*, 253-263.
(14) Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S. *Environ. Sci. Technol.* **2015**, *49*, 12333-12341.
(15) Ojanpera, S.; Pelander, A.; Pelzing, M.; Krebs, I.; Vuori, E.; Ojanpera, I. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1161-1167.
(16) Luek, J. L.; Schmitt-Kopplin, P.; Mouser, P. J.; Petty, W. T.; Richardson, S. D.; Gonsior, M. *Environ. Sci. Technol.* **2017**, *51*, 5377-5385.
(17) Ziegler, G.; Gonsior, M.; Fisher, D. J.; Schmitt-Kopplin, P.; Tamburri, M. N. *Environ. Sci. Technol.* **2019**, *53*, 8006-8016.
(18) Gonsior, M.; Schmitt-Kopplin, P.; Stavklint, H.; Richardson, S. D.; Hertkorn, N.; Bastviken, D. *Environ. Sci. Technol.* **2014**, *48*, 12714-12722.
(19) Bocker, S.; Letzel, M. C.; Liptak, Z.; Pervukhin, A. *Bioinformatics* **2009**, *25*, 218-224.
(20) Meusel, M.; Hufsky, F.; Panter, F.; Krug, D.; Muller, R.; Bocker, S. *Anal. Chem.* **2016**, *88*, 7556-7566.
(21) Tolic, N.; Liu, Y.; Liyu, A.; Shen, Y. F.; Tfaily, M. M.; Kujawinski, E. B.; Longnecker, K.; Kuo, L. J.; Robinson, E. W.; Pasa-Tolic, L.; Hess, N. J. *Anal. Chem.* **2017**, *89*, 12659-12665.
(22) Merder, J.; Freund, J. A.; Feudel, U.; Hansen, C. T.; Hawkes, J. A.; Jacob, B.; Klaproth, K.; Niggemann, J.; Noriega-Ortega, B. E.; Osterholz, H.; Rossel, P. E.; Seidel, M.; Singer, G.; Stubbins, A.; Waska, H.; Dittmar, T. *Anal. Chem.* **2020**, *92*, 6832-6838.
(23) Fu;, Q.-L.; Fujii;, M.; Riedel, T. *Anal. Chim. Acta* **2020**, *1125*, 247-257.
(24) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. *Limnol. Oceanogr. Meth.* **2008**, *6*, 230-235.
(25) Cao, D.; Huang, H. G.; Hu, M.; Cui, L.; Geng, F. L.; Rao, Z. Y.; Niu, H. Y.; Cai, Y. Q.; Kang, Y. H. *Anal. Chim. Acta* **2015**, *866*, 48-58.
(26) Brezonik, P.; Arnold, W. *Water chemistry: an introduction to the chemistry of natural and engineered aquatic systems.* ; Oxford University Pres.: New York, 2011.
(27) Andersson, A.; Lavonen, E.; Harir, M.; Gonsior, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Kylin, H.; Bastviken, D. *Environ. Sci.Wat. Res. Technol.* **2020**, *6*, 779-794.
(28) Andersson, A.; Harir, M.; Gonsior, M.; Hertkorn, N.; Schmitt-Kopplin, P.; Kylin, H.; Karlsson, S.; Ashiq, M. J.; Lavonen, E.; Nilsson, K.; Pettersson, A.; Stavklint, H.; Bastviken, D. *Environ. Sci. Wat. Res. Technol.* **2019**, *5*, 861-872.
(29) Ji, Q. C.; Xu, X. H.; Ma, E.; Liu, J.; Basdeo, S.; Liu, G. W.; Mylott, W.; Boulton, D. W.; Shen, J. X.; Stouffer, B.; Aubry, A. F.; Arnold, M. E. *Anal. Chem.* **2015**, *87*, 3247-3254.
(30) Verkh, Y.; Rozman, M.; Petrovic, M. *Chemosphere* **2018**, *200*, 397-404.
(31) Abbassi-Ghadi, N.; Jones, E. A.; Gomez-Romero, M.; Golf, O.; Kumar, S.; Huang, J. Z.; Kudo, H.; Goldin, R. D.; Hanna, G. B.; Takats, Z. *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 255-264.

(32) Ksionzek, K. B.; Lechtenfeld, O. J.; McCallister, S. L.; Schmitt-Kopplin, P.; Geuer, J. K.; Geibert, W.; Koch, B. P. *Science* **2016**, *354*, 456-459.
(33) Schymanski, E. L.; Singer, H. P.; Longree, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Vidal, C. R.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 1811-1818.
(34) Witt, M.; Fuchser, J.; Koch, B. P. *Anal. Chem.* **2009**, *81*, 2688-2694.
(35) Cortes-Francisco, N.; Caixach, J. *Anal. Bioanal. Chem.* **2015**, *407*, 2455-2462.
(36) Zark, M.; Dittmar, T. *Nat. Commun.* **2018**, *9*, 3059.
(37) Hu, R. K.; Zhang, L. F.; Hu, J. Y. *Sci. Total Environ.* **2017**, *603*, 8-17.
(38) Sang, Z. Y.; Jiang, Y. A.; Tsoi, Y. K.; Leung, K. S. Y. *Water Res.* **2014**, *52*, 260-274.
(39) Berset, J. D.; Ochsenbein, N. *Chemosphere* **2012**, *88*, 563-569.
(40) Mead, R. N.; Morgan, J. B.; Avery, G. B.; Kieber, R. J.; Kirk, A. M.; Skrabal, S. A.; Willey, J. D. *Mar. Chem.* **2009**, *116*, 13-17.
(41) Torres, C. I.; Ramakrishna, S.; Chiu, C. A.; Nelson, K. G.; Westerhoff, P.; Krajmalnik-Brown, R. *Environ. Eng. Sci.* **2011**, *28*, 325-331.
(42) Tran, N. H.; Gin, K. Y. H.; Ngo, H. H. *Sci. Total Environ.* **2015**, *538*, 38-57.
(43) Sun, X. F.; Chen, M.; Wei, D.; Du, Y. G. *J. Environ. Sci.* **2019**, *81*, 52-67.
(44) Jiang, J.; Han, J.; Zhang, X. *Environ. Sci. Technol.* **2020**, *54*, 1646-1656.
(45) Liu, X.; Liu, R.; Zhu, B.; Ruan, T.; Jiang, G. *Environ. Sci. Technol.* **2020**.
(46) Hao, Z. N.; Shi, F. Q.; Cao, D.; Liu, J. F.; Jiang, G. B. *Environ. Sci. Technol.* **2020**, *54*, 1668-1676.
(47) Allard, S.; Tan, J.; Joll, C. A.; von Guntenit, U. *Environ. Sci. Technol.* **2015**, *49*, 11105-11114.
(48) Jiang, G.; Li, X. *A New Paradigm for Environmental Chemistry and Toxicology: From Concepts to Insights*; Springer: Singapore, 2020.
(49) Gonsior, M.; Mitchelmore, C.; Heyes, A.; Harir, M.; Richardson, S. D.; Petty, W. T.; Wright, D. A.; Schmitt-Kopplin, P. *Environ. Sci. Technol.* **2015**, *49*, 9048-9055.
(50) Moreno-Andres, J.; Peperzak, L. *Chemosphere* **2019**, *232*, 496-505.
(51) Manasfi, T.; De Meo, M.; Coulomb, B.; Di Giorgio, C.; Boudenne, J. L. *Environ. Int.* **2016**, *88*, 94-102.
(52) Zwiener, C.; Richardson, S. D.; De Marini, D. M.; Grummt, T.; Glauner, T.; Frimmel, F. H. *Environ. Sci. Technol.* **2007**, *41*, 363-372.
(53) Fielding, M.; Farrimond, M. *Disinfection by-products in drinking water: current issues*; Elsevier: Cambridge, 1999.
(54) Diana, M.; Felipe-Sotelo, M.; Bond, T. *Water Res.* **2019**, *162*, 492-504.

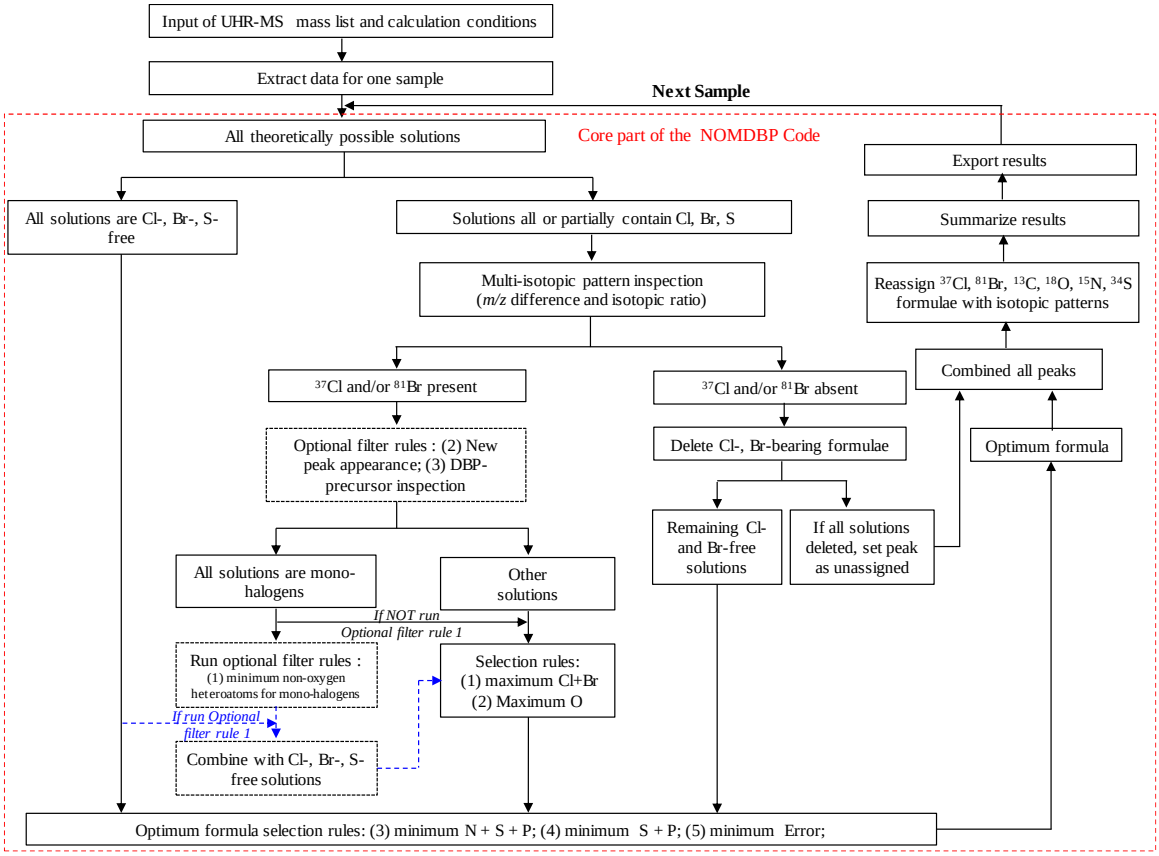

Fig. 1. Flow of the NOMDBP Code.

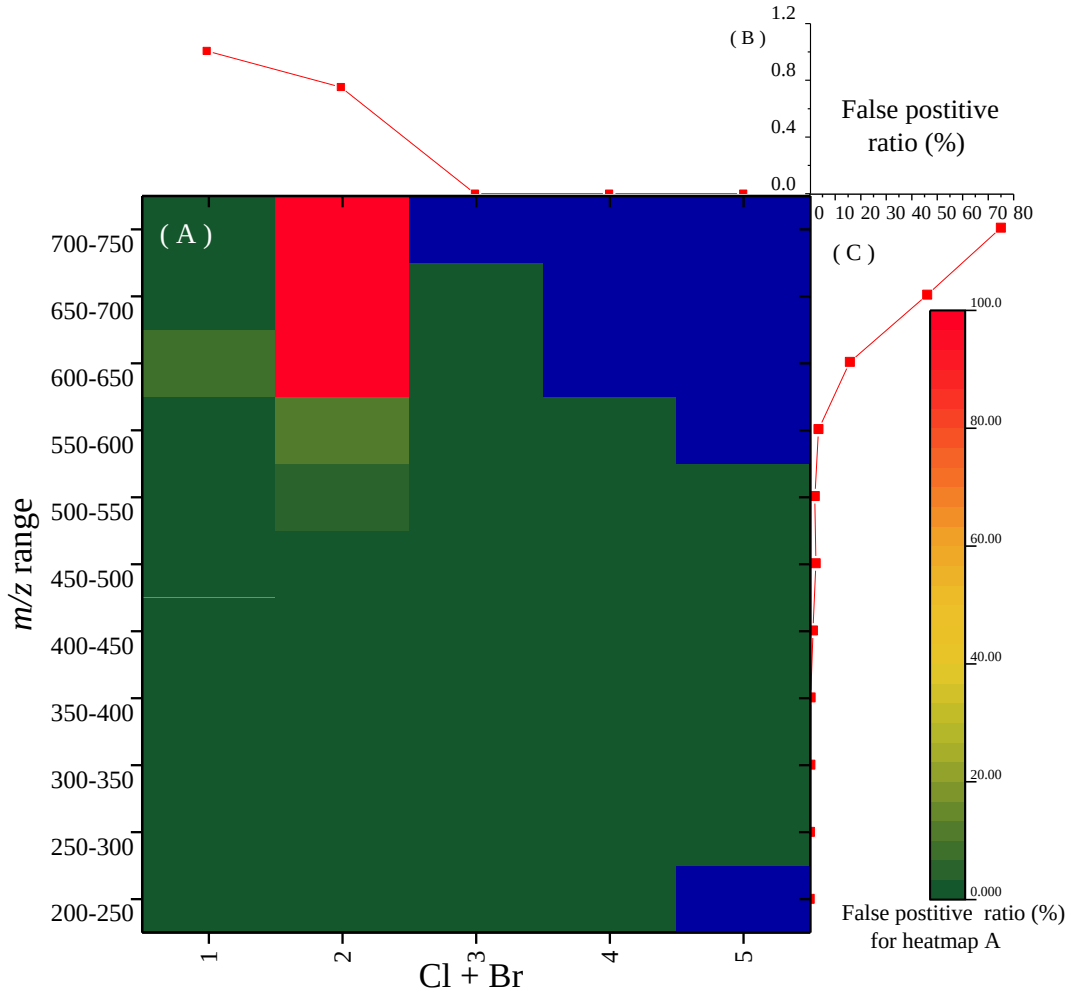

Fig. 2. The false positive (inaccuracy) ratio of all X-DBP parent formulae as functions of *m/z* ranges and Cl +Br number (A), and false positive ratio for Cl +Br number (B) and *m/z* range (C). Blue color in heatmap A represents the absence of DBP parent formulae in the relevant *m/z* range and Cl+Br number.

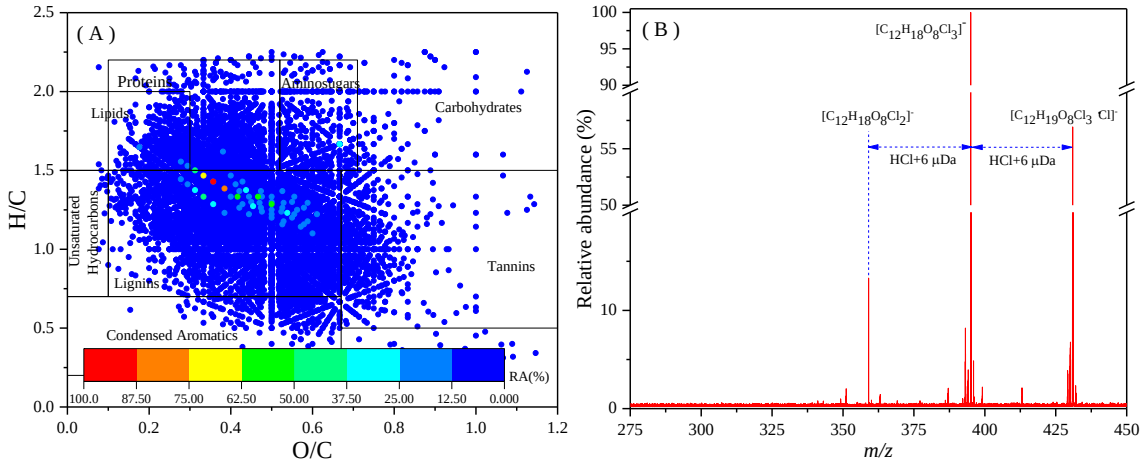

Fig. 3. The van Krevelen diagrams of the Tokyo Bay seawater DOM (Odaiba sample) (A), and FT-ICR-MS/MS fragmentation spectra of the parent ion at *m/z* 431 ($[C_{12}H_{19}O_8Cl_4]^-$) (B).

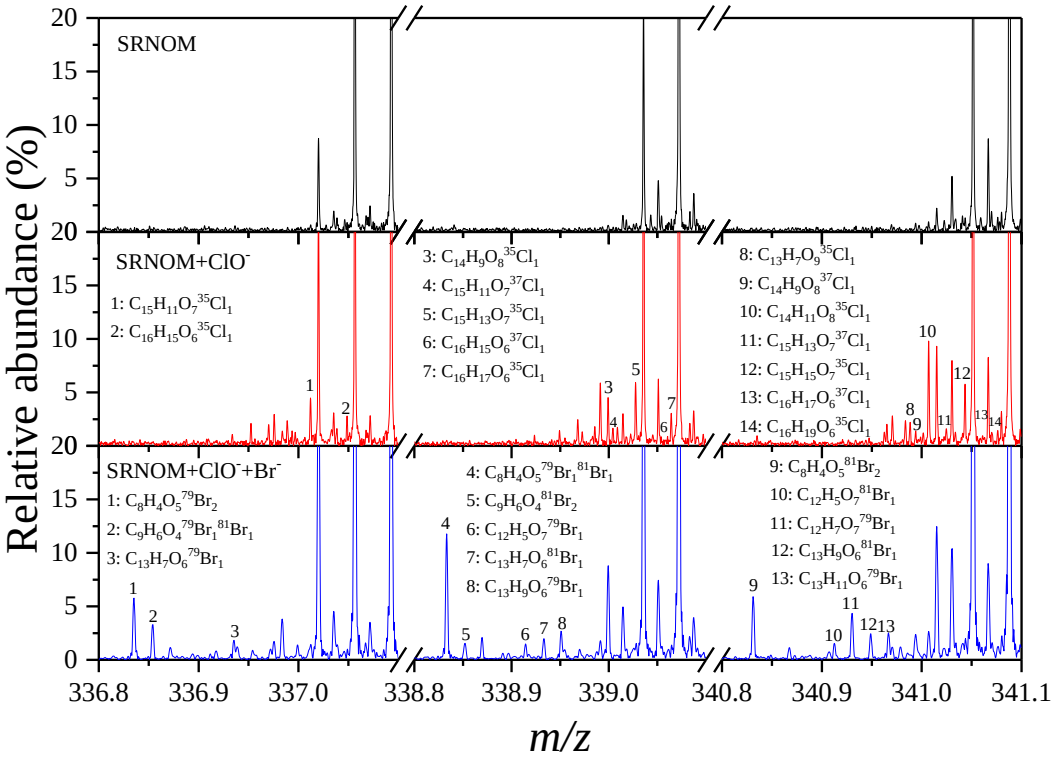

Fig. 4. Enlarged FT-ICR-MS spectra of chlorinated and non-chlorinated SRNOM in the presence and absence of bromide at nominal mass 337, 339, and 341.