Molecular characteristics of saturated halogenated disinfection byproducts in chlorinated urban eutrophic lake water elucidated by ultrahigh-resolution mass spectrometry

Shixi Wu [a], Qing-Long Fu [a, *], Manabu Fujii [b], Jibao Liu [b], Ahmed Elreedy [b, c], Xin Yang [d], Pingqing Fu [e], Yanxin Wang [a]

[a] *MOE Key Laboratory of Groundwater Quality and Health, School of Environmental Studies, China University of Geosciences, Wuhan 430074, China*

[b] *Department of Civil and Environmental Engineering, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro-Ku, Tokyo 152-8550, Japan*

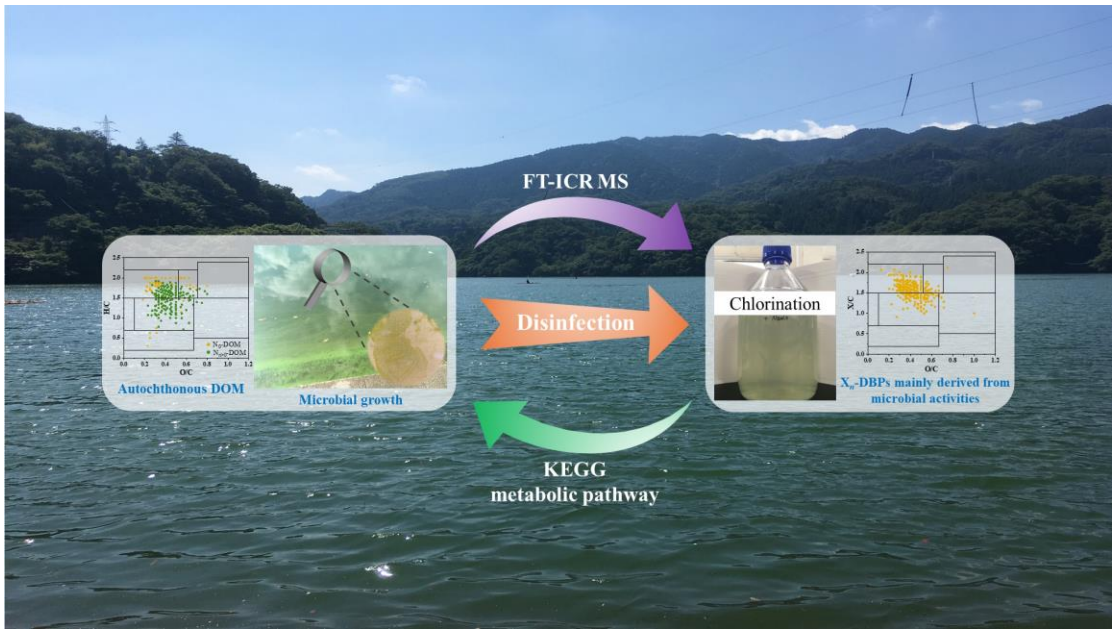
[c] *Institute of Technical Microbiology, Hamburg University of Technology, Hamburg 21073, Germany*

[d] *School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China*

[e] *Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin, 300072, China*

[*] *Corresponding author: Qing-Long Fu (E-mail address: fuqinglong@cug.edu.cn)*

Graphical abstract



Highlights

- The *Microcystis* growth revealed the slight algal bloom of the natural urban eutrophic lake.
- Microbial growth enhanced precursors for the nitrogenous X_n-DBPs in urban eutrophic lake.
- A novel data interpretation paradigm was proposed for X_n-DBPs research.
- X_n-DBPs mainly derived from microbes were highly saturated, aliphatic, bioavailable.
- Lipid metabolism is involved in saturated X_n-DBPs species from autochthonous precursors.

Abstract

Lake eutrophication affects the molecular composition of aquatic dissolved organic matter (DOM) and halogenated disinfection byproducts (X_n-DBPs). However, the effects of autochthonous DOM on the X_n-DBPs formation during disinfection of natural eutrophic water from the perspective of biological metabolism are still poorly revealed. Herein, the natural urban eutrophic lake (UEL) water with slight eutrophication was employed to elucidate the discrepancies in X_n-DBPs formation between autochthonous and allochthonous DOM based on the ultrahigh-resolution mass spectrometry. The increased operational taxonomic units, microbial cell density and relative abundance of genus *Microcystis* in the collected samples indicated the slight algal bloom of the lake. The number and its proportion of nitrogenous X_n-DBPs in chlorinated UEL water samples were significantly larger ($p < 0.05$) than those for chlorinated SRNOM. Microbes dominated by *Microcystis* contributed largely to releasing autochthonous DOM due to the microbial cell lysis and microbe-derived X_n-DBPs species upon disinfection. The X_n-DBPs species mainly derived from microorganisms were highly saturated, reduced, bioavailable, nitrogenous, and toxic but lowly oxidized and aromatic than terrestrially derived X_n-DBPs species. Moreover, for the first time, the connection between microbial lipid metabolism and X_n-DBPs species exclusively identified in chlorinated UEL water indicated the considerable contribution of lipid metabolites to saturated X_n-DBPs species. This study has reported novel data interpretation paradigm for X_n-DBPs research, deepening our understanding towards the formation mechanisms of microbe-derived X_n-DBPs species from the view of microbial metabolic pathways.

Keywords: Eutrophication, Disinfection byproducts, Dissolved organic matter, Fourier transform ion cyclotron resonance mass spectrometry, Microbial metabolism

1. Introductions

The water quality of lake, a major source of drinking water, is crucial to human health and aquatic ecology. Disinfection with oxidative reagents is strongly recommended for drinking water treatment to inhibit pathogenic bacterial growth [1]. However, the frequent occurrence of eutrophication not only deteriorates the lake water quality [2, 3] but also yields more microbe-derived organic compounds in the water with resultant high diversity of halogenated disinfection byproducts (X_n-DBPs, X represents the halogen and *n* is the atom number of halogen) during the drinking water disinfection [2-6].

Aquatic dissolved organic matter (DOM) is a highly complex mixture of various autochthonous and allochthonous molecules [7]. Autochthonous DOM (microbial metabolites such as peptides and proteins) is mainly derived from normal metabolic activities and released from cell lysis of algal microorganisms, including cyanobacteria [2, 5, 8]; while allochthonous DOM typically originates from terrestrial input and is mainly composed of humic substances including lignin-like compounds [7]. In recent three decades, the Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), providing ultrahigh mass resolution and the highest mass accuracy, has been the state-of-the-art technology in elucidating the molecular composition of DOM and X_n-DBPs [9, 10]. The increasing lake trophic status will release and accumulate autochthonous DOM characterized by more CHON compounds, higher aliphatic content, and lower oxidation levels than allochthonous molecules [2, 3]. For example, DOM molecules positively correlated with total nitrogen were in the compositional space of high H/C and low O/C values in the van Krevelen diagram [2]. The DOM molecules identified only in cyanobacterial bloom water and the cyanobacterial extract contained more CHO, CHON, and CHONS compounds with high H/C values than those identified only in the bloom-free water [5]. Therefore, elucidating the molecular composition of aquatic DOM using FT-ICR MS is expected to provide valuable insights for the drinking water treatment of high trophic status

lake water.

Algal organic matter (AOM) released from the living and/or lytic algal cells caused by disinfectants could react with disinfectants, forming harmful X_n-DBPs species [4, 11]. For example, cyanobacterial intracellular organic matter contributes to the formation of carbonaceous and nitrogenous X_n-DBPs during chlorination and chloramination [12]. Only 66 regulated and priority unregulated X_n-DBPs were identified in algae-impacted waters during chlorination with more than two-folds of total concentrations of X_n-DBPs and 2-3-fold increase in the toxicity of X_n-DBPs than those for the algae-free water [13, 14]. However, compared with few reported regulated/known X_n-DBPs in these studies, a comprehensive analysis of numerous unknown X_n-DBPs chemodiversity during disinfection of natural algae-containing waters is unknown. In addition, autochthonous DOM exhibits distinct effects on forming X_n-DBPs and their molecular characteristics, chemical components, and resultant toxicity. For example, AOM generally contains more N-containing compounds (partially amino acids), thus increasing the molecular diversity of the more toxic N_n-DBPs species during disinfection than natural organic matter (NOM) [11, 15]. AOM significantly contributes to forming organic chloramines in chlorinated drinking water due to its abundant N-containing precursors for X_n-DBPs formation [16]. Furthermore, O- and N-rich AOM precursors released from *Microcystis aeruginosa* by algaecides were of higher unsaturation and more preferable to X_n-DBPs precursors than wildfire ash-derived DOM [17]. The chlorination of pre-ozonized AOM substantially increased the intensity and diversity of unknown X_n-DBPs species [18]. Moreover, the chlorination of AOM molecules yielded X_n-DBPs species with high O/C and low H/C values, mainly from phenolic and unsaturated aliphatic precursors [19]. Similarly, The fatty acid-type X_n-DBPs with low unsaturation were the addition reaction products of reactive chlorine species with aliphatic DOM precursors from solid-phase extracted *M. aeruginosa* [20]. The chlorination of plasma oxidated algal-cultured medium yielded 2,486 N-free and 1,984 N-containing X_n-DBPs species, with

the most abundant X_n-DBPs species of $C_{11}H_nO_mCl_x$ and $C_{18}H_nN_mO_zCl_x$, respectively [21]. However, all these FT-ICR MS-based X_n-DBPs studies were performed only using pure cultured cyanobacterial cells (*e.g.*, *M. aeruginosa*) disinfected in a well-controlled laboratory environment. Effects of naturally occurring algal microorganism growth in field eutrophic lake water on the X_n-DBPs formation remain largely unknown at the molecular level.

Algae maintain their metabolic activities by consuming and releasing organic molecules in surrounding waters [5, 22]; these metabolites have been recognized to affect X_n-DBPs formation during disinfection of algae-containing water profoundly. For example, fatty acid molecules were identified as the algae-derived DOM precursors for $Cl_{1\text{-}5}$-DBPs species during chlorinating *M. aeruginosa*-containing water [20]. However, the specific microbial metabolites involved in forming X_n-DBPs species are still poorly known. The Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolome database is a valuable tool to reveal possible metabolic pathways by linking DOM molecules in field eutrophic waters to the metabolic pathways of algal microorganisms. For example, the "biosynthesis of other secondary metabolites" revealed by the KEGG metabolic pathways was the dominant DOM molecule in Taihu Lake [5]. Therefore, the detailed investigation of the relationship between X_n-DBPs precursors and microbial metabolic pathways is expected to contribute to an in-depth understanding of the microbial metabolic source of X_n-DBPs precursors during disinfection of natural eutrophic source water.

The main objectives of this study were to (1) investigate the effect of the microbial growth dominated by cyanobacteria on the molecular composition and chemodiversity of DOM and X_n-DBPs in natural waters and (2) to link the microbial metabolic pathways with precursor compounds and resultant X_n-DBPs species by integrating the ultra-high resolution mass spectrometry (FT-ICR MS) and KEGG database. The discrepancy between autochthonous DOM and allochthonous DOM in X_n-DBPs formation was also examined.

2. Materials and Methods

2.1 Sample preparation

The investigated lake (Lake A) is a significant drinking water source for residents in satellite towns of the Kanto area, Japan, and has been listed as a typical urban eutrophic lake (UEL) (https://www.env.go.jp/en/water/wq/lakes/issue.html, **Figures S1** and **S2**). Therefore, the raw UEL water from Lake A, as a representative natural and eutrophic water, was periodically sampled during the typical algal bloom season (August 2020). Specifically, four raw UEL water samples were collected from Lake A on August 07, 14, 21, and 28, 2020. The physicochemical variables of all water samples were tabulated in **Table S1**. Granular active carbon (GAC), a promising sorption material for DOM removal [23], was employed to simulate the chlorination scenario in the drinking water treatment. Specifically, the UEL water collected on 28th August (UEL28) was treated using GAC at room temperature in the dark (UEL28+GAC), followed by filtration with a 0.45 μm membrane. A part of each water was prefiltered with a 0.45 μm membrane to remove cyanobacterial cells (typically 0.5 to 60 μm) for further disinfection (UEL+ClOPF). In contrast, another portion of the UEL water was disinfected without prefiltration (UEL+ClOF). All raw water with or without prefiltration was subjected to chlorination using 5.0 mg/L NaClO (expressed as free chlorine concentration determined by the colorimetric method of *N*, *N*-diethyl-*p*-phenylenediamine with a pocket colorimeter) without adjusting pH to simulate the chlorination scenario in the drinking water treatment plant. All chlorinated samples were incubated for 7 days at room temperature under the dark and then filtered through a 0.45 μm membrane for the DOM extraction. Three natural water samples without eutrophication (referred to as background samples) were collected from the urban lake or river geographically close (8 - 27 km distance) to the Lake A in later autumn or winter (**Content S1**) to represent the terrestrially dominated waters with ignorable autochthonous components of water eutrophication. The collected background samples were filtered through a 0.45 μm

membrane for the DOM extraction immediately after transferring to the laboratory. In addition to UEL water DOM with autochthonous and allochthonous components, the Suwannee River NOM (SRNOM, 2R101N, purchased from International Humic Substances Society) was employed as the representative of terrestrially derived DOM [24], to tentatively discern the different effects of autochthonous and allochthonous DOM on the X_n-DBPs formation. The FT-ICR MS data of the two SRNOM standards (SRNOM1 and SRNOM2, respectively) and their chlorinated samples were collected from our previous studies [25, 26]. Briefly, 2 L SRNOM1 (3.0 mg-C/L) and SRNOM2 (50.0 mg-C/L) were chlorinated with 5.0 mg/L and 50 mg/L NaClO, respectively. The residual chlorine in the SRNOM1 was 3.58 ± 0.02 mg/L after 7-day incubation (**Figure S3**). The chlorinated SRNOM solutions were incubated for 7 days at room temperature in the dark without adjusting the solution pH, then filtered through a 0.45 μm membrane for the solid-phase extraction (SPE). The sample treatments are detailed in **Table S2**. (Zhang et al., 2014; Zhang and Yang, 2018)

The DOM from all samples was purified with SPE extraction according to the procedures described elsewhere [27]. Hydrochloric acid (HCl) was used to acidify sample to evaluate SPE recovery and avoid the adverse effects of high concentration of non-violate sulfurate on the FT-ICR MS instrument. Briefly, all filtered samples were acidified to pH ≈ 2 using concentrated HCl and then gravitationally passed through the SPE cartridges (Bond Elut PPL cartridges (1 g and 6 mL, Agilent, USA) rinsed and activated with 6 mL of methanol (LC-MS grade, Kanto Chemical, Japan). The cartridges were desalted using 20 mL HCl (pH ≈ 2), followed by 6 mL ultrapure water, and then thoroughly dried using N_2 gas (≥ 99.9% purity). DOM molecules were finally eluted with 6 mL methanol.

The centrifugal concentrator (CVE-2200, EYELA, Japan) was used to completely dry the aliquot (0.2 mL) of each SPE-extracted DOM, followed by redissolving in ultrapure water. A total organic carbon (TOC) analyzer (TOC 4100, Shimadzu, Japan) was employed to determine the TOC of all samples. The TOC

concentration and SPE recovery of DOM were averaged at 1.55 mg/L and 64.04%, respectively. All eluted DOM samples were diluted with an identical volume of ultrapure water (1: 1, *v*: *v*) for FT-ICR MS measurement.

2.2 DNA extraction and microbial community analysis

The microbial DNA extraction and microbial community analysis were described in **Content S2**.

2.3 FT-ICR MS measurement

The FT-ICR MS spectra of 16 SPE-extracted DOM samples were measured by a 9.4 Telsa FT-ICR MS (SolariX XR, Bruker Daltonics Inc., Germany) in the negative electrospray ionization mode at Tohoku University, Sendai, Japan, with the following instrument parameters tabulated in **Table S3**: 150-1000 mass-to-charge ratio (*m*/*z*), 450 average scans, 2MWord data size, -4.5 kV capillary voltage, 150 μL/h direct infusion rate, and 1 ms ion accumulation time [25]. All spectra were externally calibrated with ion clusters using the NaI solution before FT-ICR MS measurement and then internally calibrated with known CHO-homologous series to ensure the mass accuracy of entire spectra within 1.0 ppm.

2.4 Data analysis

Given the few structural isomers for X_n-DBPs with three C atoms (C_3-DBPs) and their difficulties in the tandem mass spectrometry measurement, the toxicities of C_3-DBPs were evaluated based on their retrieved structures from the PubChem in this study. The ecotoxicity and predicted physical-chemical properties (probability for rapid biodegradation and biotransformation half-life) of some typical highly saturated C_3-DBPs species including all their isomers were assessed by the quantitative structure-activity relationship method using Ecological Structure Activity Relationships software [28, 29] and EPI Suite™ software by U.S. EPA (Version 4.11) [30], respectively (**Content S6**). Based on the Globally Harmonized System of Classification and Labeling of Chemicals (GHS), the predicted toxicity for organic chemicals was divided into four levels: not harmful ($LC_{50}/EC_{50} \geq 100$ mg/L), harmful

(100 mg/L > LC_{50}/EC_{50} ≥ 10 mg/L), toxic (10 mg/L > LC_{50}/EC_{50} ≥ 1 mg/L), and very toxic (1 mg/L ≥ LC_{50}/EC_{50}).

The FT-ICR MS data analysis, statistical analysis, and KEGG data analysis were described in **Content S3**.

3. Results and Discussion

3.1 Bacterial community structure during algal bloom

The number of OTUs (11042 - 13289) and microbial cell density ($10^{3.78}$ - $10^{6.08}$ cell/mL) gradually increased during the algal bloom season in Lake A, suggesting elevating levels of microbes, including algae, during the sampling period (**Figure 1A**). The abundance of the four sample species at the phylum level illustrated in **Figure 1B** reveals that Proteobacteria (relative abundance of 36.07%), Firmicutes (21.92%), and Cyanobacteria (15.86%) were the three most prevalent phyla in samples collected on 7th August (UEL07). Contrasting to the considerably decreased relative abundance of *Firmicutes* after three weeks of natural algae growth, the relative abundance of *Cyanobacteria* increased to 53.55% in sample UEL28. *Proteobacteria* and *Cyanobacteria* were the most abundant during algal blooms in Lake A. Furthermore, it can be clearly shown that the bacterial communities changed with the eutrophication processing at the family and genus levels (**Figure 1**). For example, *Exiguobacterium* sp. was dominant in the UEL07 sample, accounting for 20.43% of the total abundance, but was decreased to 0.02% in the UEL28 sample. Compared to the sample UEL07, the relative abundance of *Microcystis* in UEL28 increased from 0.00% to 47.32%, suggesting that *Microcystis* was the key contributor to algal bloom in Lake A. The equal abundance noticed for the family *Microcystaceae* in UEL28 (*i.e.*, 47.32%) also indicates that no genera other than *Microcystis* from such a family were identified. These results demonstrated that *Microcystis* was the dominant microbial species responsible for the cyanobacterial bloom in Lake A from August 07 to 28, 2020, which was consistent with the previous finding that *Microcystis* were the predominant species in Lake A during cyanobacterial seasonal succession in late

August [31].

M. aeruginosa, the major species of the genus *Microcystis*, are frequently detected in bloom-forming cyanobacteria in natural and engineered waters [19, 21]. The AOM molecules from *M. aeruginosa* had been recognized as reactive X_n-DBPs precursors upon chlorination because of their higher average C number and lower average oxygen number than those for terrestrial DOM [17]. The concentrations of N_0-DBPs (N-free X_n-DBPs) and $N_{n>0}$-DBPs (N-bearing X_n-DBPs) species were found to increase with the algae growth due to the discrepancy in organic matter composition at different growth stages of *M. aeruginosa* [32]. The high residual chlorine and extended growth periods of algal cells promoted the formation of X_n-DBPs species with more chlorine [32, 33]. The prevalence of cyanobacteria, including *M. aeruginosa* demonstrates the autochthonous and allochthonous DOM molecules in Lake A, suggesting the possible formation of more toxic $N_{n>0}$-DBPs and/or $Cl_{n\geq2}$-DBPs species in the UEL water.

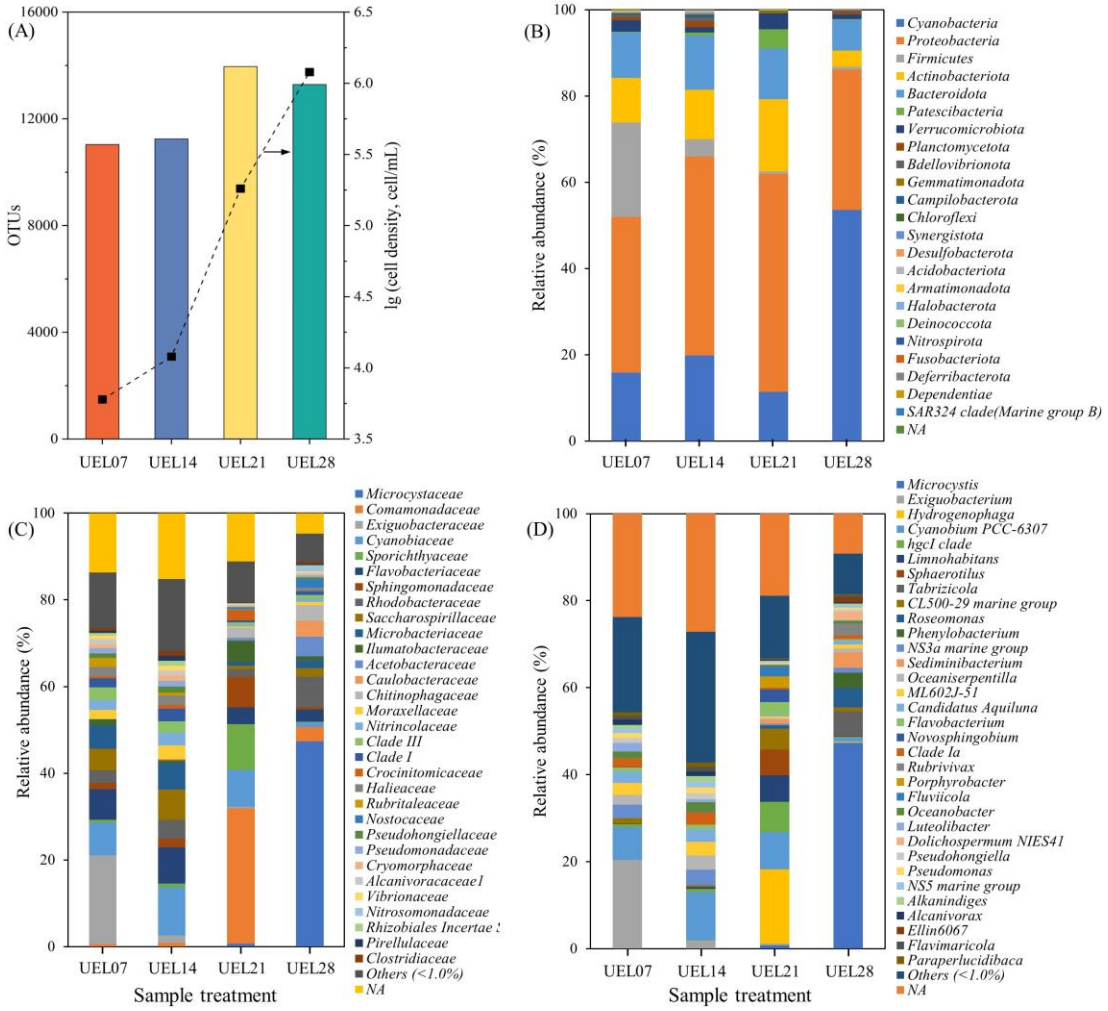


Figure 1. The OTUs and the cell density of bacteria in samples (A) and the microbial relative abundance at phylum (B), family (C), and genus (D) levels. The analysis was based on 16S rRNA gene sequencing of the microbiota of the four raw UEL water samples.

3.2 Molecular composition of DOM and X_n-DBPs species affected by microbial growth

In addition to the similar overall FT-ICR MS spectra profile (**Figure S4**), minor differences were observed for UEL water DOM samples collected during microbial growth with the respects of number and intensity proportion of different molecular compositions (**Figure S5**). This finding was further supported by the visual similarity between UEL28 and other raw UEL water samples in the van Krevelen diagram based

on the inter-sample ranking analysis, suggesting minor temporal-spatial variations in DOM molecular composition (**Figure S6**). Similar results were also identified for the intensity-weighted values of molecular parameters, including m/z_{iw}, H/C_{iw}, O/C_{iw}, $AI_{mod,\ iw}$, DBE_{iw}, and $NOSC_{iw}$ (**Figure S7**). However, the gradually elevating N/C_{iw} value (**Figure S7E**) demonstrated the increase of N-containing compounds in the UEL water caused by microbial growth. Compared to the SRNOM samples, a series of N-containing compounds (*e.g.*, $C_{10}H_7N_1O_4$ and $C_{15}H_{25}N_1O_7S_1$) with a gradual increase in their peak intensities were detected in the UEL water samples during microbial growth (**Figure S8**), which was consistent with the N-rich nature of AOM [20]. The minor difference in the overall molecular composition of UEL water DOM during the sampling period in this study could be attributed to the dominated allochthonous molecules in the slightly eutrophicated Lake A. In contrast, the microbial growth resulted in the continuous release of autochthonous molecules into the Lake A water.

The different overall profile of FT-ICR MS spectra for UEL water samples from those for two raw SRNOM samples (**Figure S4**) could be mainly attributed to the autochthonous molecules in the natural eutrophic water in this study. The UEL water DOM contained more CHON-class and saturated compounds but fewer CHO-class and tannin-like compounds than SRNOM (**Figure S5**). Indeed, formulae in the ranks 1-6 (relatively high mass peak intensities) of SRNOM2 featured more relative oxygen-rich and hydrogen-poor components (*e.g.*, tannin-like CHO-type compounds) compared to the UEL water DOM (**Figure S6**). The UEL water DOM peaks were distributed mainly at a lower m/z (~150 to ~600), yielding lower m/z_{iw} values than those of SRNOM (~250 to ~800) (**Figure S4** and **S7B**). Furthermore, the UEL water DOM had higher H/C_{iw} and N/C_{iw} but fewer formulae and lower O/C_{iw}, $AI_{mod,\ iw}$, DBE_{iw}, and $NOSC_{iw}$ than those of SRNOM (**Figure S7**). These results suggest the dominance of more saturated and aliphatic but less oxidized and aromatic compounds in UEL water than in the SRNOM, which could be attributed to autochthonous compounds [3, 34] and contributed considerably to the formation of aliphatic $X_{n>0}$-DBPs

[20, 21].

In this study, the conventional H/C in the van Krevelen diagram was replaced by (H+X)/C to visually characterize the X_n-DBPs and their putative precursors [25]. The X_n-DBPs species identified only in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample were considered the X_n-DBPs mainly derived from microbial activities. Furthermore, Cl-containing molecules identified both in the raw SRNOM and the corresponding chlorinated SRNOM samples were considered as the possible Cl-adduct resulting from samples acidification by HCl. Only 15 and 7 possible Cl-adducts were found in the chlorinated SRNOM1 and SRNOM2 samples, accounting for < 3% and < 1% of the number of all X_n-DBPs in the chlorinated SRNOM1 and SRNOM2 samples, respectively. These results suggested that the potential Cl-adduct artifact induced by HCl had minor impact on the main conclusion of this study. In contrast to the decreasing number of chlorinated organic compounds (from 107 to 59 on 07 and 28 August, respectively, **Figures 2** and **S9**), an insignificant difference ($p > 0.05$) was observed in the number-average molecular characteristics, including *m*/*z*, DBE, (H+X)/C, NOSC, Cl/C, and AI_{mod} among all UEL water samples without chlorination from 07 August to 28 August (**Figure S10**). Compared with the raw UEL water without chlorination, chlorination resulted in a significant ($p < 0.05$) increase in the intensities and number of chlorinated organic compounds, including X_n-DBPs, in all UEL water samples, suggesting the occurrence of X_n-DBPs species induced by chlorination. Additionally, the X_n-DBPs number in chlorinated UEL water samples with prefiltration decreased from 1117 (excluding isotopic formula, accounting for 17.00% and 27.44% of the number and intensity of all assigned peaks, respectively, 07 August) to 480 (6.22% and 9.13%, 14 August), followed by an increase to 693 (10.58% and 35.06%) on 28 August (**Figures 2** and **S9**). While the X_n-DBPs number in chlorinated UEL water samples without prefiltration decreased from 1147 (20.52% and 34.94% of number and intensity percentages, respectively, 07 August) to 494 (8.42% and 10.53% of number and intensity percentages, respectively, 21 August),

followed by an increase to 1097 (16.22% and 30.39% of number and intensity percentages, respectively) on 28 August. Similar results were also observed for the number and proportion of saturated X_n-DBPs in chlorinated UEL water samples with and without prefiltration (**Figures 2** and **S9**), suggesting that the identified saturated X_n-DBPs species were mainly derived from microbial growth. This conclusion is further supported by the consistent changing tendency of the number and proportion of $N_{n>0}$-DBPs species in chlorinated UEL water samples without prefiltration with the relative abundance of *Cyanobacteria* and the N/C_{iw} value in the raw UEL water samples during the sampling period (**Figures 1B**, **S7E**, and **S11**). Moreover, the number of CHON and CHONS newly assigned in the UEL+ClOPF increased from 705 on 07 August to 1329 on 28 August and 391 on 07 August to 851 on 28 August, respectively. Similarly, the number of CHON and CHONS newly assigned in the UEL+ClOF increased from 260 on 07 August to 897 on 28 August and 279 on 07 August to 390 on 28 August, respectively (**Figure S12**). These results are consistent with previous findings that the increased X_n-DBPs species by the microbial growth in the UEL water [32, 33] could be attributed to the release of autochthonous precursors [17].

The X_n-DBPs species in all chlorinated UEL water samples exhibited similar compositional space in the van Krevelen diagram (**Figures 2** and **S9**). However, differing from the dominated lignin-like X_n-DBPs species in the chlorinated SRNOM, the majority of X_n-DBPs species mainly derived from microbial activities are of a saturated nature (**Figures 2** and **S9**) [20]. Indeed, 62 - 233 of X_n-DBPs species with $(H+X)/C \geq 1.5$, accounting for 74.70% - 92.96% number of all X_n-DBPs species, were identified exclusively in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample. The greater extent of unsaturation and oxidation of DOM molecules correlated positively with their larger positive (DBE-0.5O)/C and NOSC values, respectively [35]. In this study, the majority of X_n-DBPs species mainly derived from microbial activities in all disinfected UEL water samples were reduced compounds, accounting for 38.19% - 60.24% number of unsaturated compounds and 6.02% -

32.16% of saturated compounds (**Figure S13**)[35]. The chlorinated UEL28 with prefiltration had significantly lower ($p < 0.05$) *m*/*z*, DBE, NOSC, and AI_{mod} but higher ($p < 0.05$) (H+X)/C than other chlorinated UEL water samples with prefiltration (**Figure S10**). For chlorinated UEL water samples without prefiltration, no significant difference ($p > 0.05$) was observed in these molecular characteristics between UEL28 and other UEL water samples. These results suggested the difference in the effects of chlorination with and without prefiltration on the molecular characteristics and composition of X_n-DBPs species mainly derived from microbial activities in UEL water.

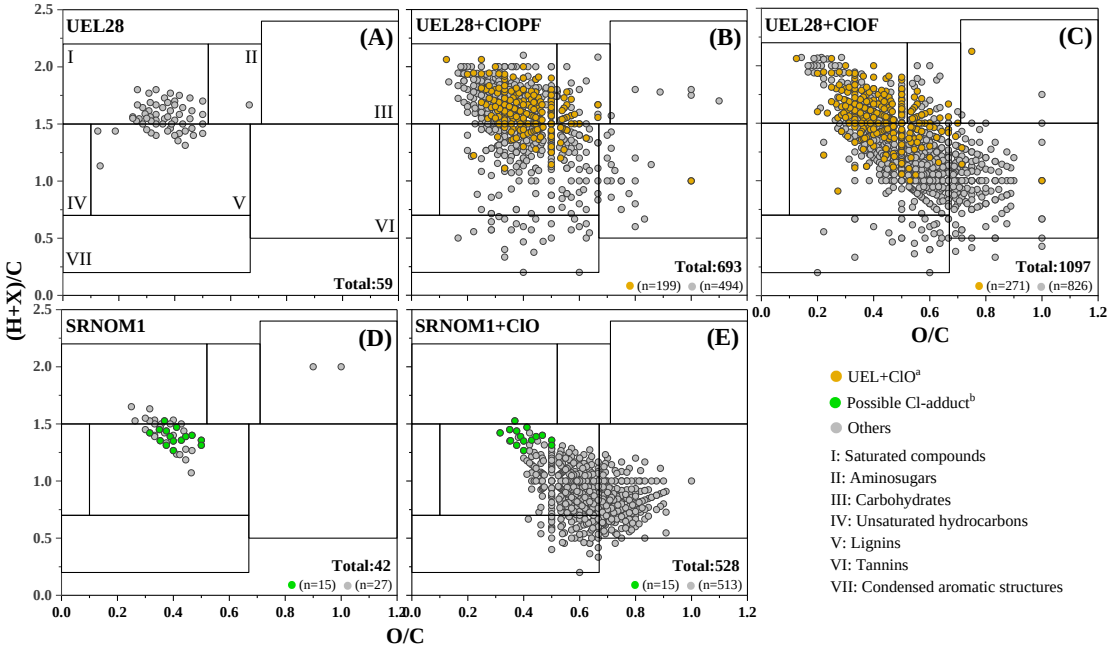


Figure 2. The van Krevelen diagram of X_n-DBPs for UEL28 (A), UEL28+ClOPF (B), UEL28+ClOF (C), SRNOM1 (D), and SRNOM1+ClO (E) samples. UEL28 indicated the raw UEL water samples were collected on August 28, 2020. ClOPF and ClOF indicate the chlorinated UEL water samples with and without prefiltration, respectively. SRNOM1 and SRNOM1+ClO indicated 1R108N and chlorinated 1R108N samples, respectively. UEL+ClO[a]: X_n-DBPs identified exclusively in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample. Possible Cl-adduct [b]: Cl-containing molecules identified both in the raw SRNOM2 and chlorinated SRNOM2 samples.

3.3 Effects of cell lysis on the X_n-DBPs formation

Effects of the autochthonous DOM derived from microbial (dominated by *cyanobacteria*, **Figure 1**) cell lysis during chlorination on the X_n-DBPs formation were examined by comparing the X_n-DBPs formation in the chlorinated UEL water samples with prefiltration (UEL+ClOPF) with that for the chlorinated UEL water samples without prefiltration (UEL+ClOF). The total number of all X_n-DBPs species in the four chlorinated UEL water samples without prefiltration was more than that for the four chlorinated UEL water samples with prefiltration (3810 versus 2815, **Figure S13**). Furthermore, the total number of unique X_n-DBPs species mainly derived from microbial activities in the chlorinated UEL water samples without prefiltration was slightly larger than that for the chlorinated UEL water samples with prefiltration (861 versus 806, **Figure S13**). The number of unique X_n-DBPs species for treatments with prefiltration was comparable to those without prefiltration (1468 versus 1524, **Figure S14**). From the intersection of only chlorinated UEL water samples with and without prefiltration, the shared X_n-DBPs species composed the subset with the largest number (567, **Figure S14B**). These results suggest chlorination of UEL water without prefiltration yielded more X_n-DBPs species mainly derived from microbial cell lysis than that with prefiltration. Notably, all regulated X_n-DBPs compounds (66 regulated X_n-DBPs) reported by previous studies [12-14] were not identified in this study, highlighting the significance of comprehensive characterization of unregulated and unknown X_n-DBPs. In addition, Compared with 3871 and 2807 X_n-DBPs in the chlorinated algal-cultured samples with and without pre-ozonation, respectively[18], 1524 and 1468 unique X_n-DBPs (excluding isotopic formulae) identified in the chlorinated UEL water samples with and without prefiltration, respectively (**Figure S14**), suggesting that the potential high risk in rivers and reservoirs with slight eutrophication.

Totally, 16 newly generated X_n-DBPs species (including isotopic formulae) at the nominal mass of 263 exemplified the clear differences between chlorinated UEL

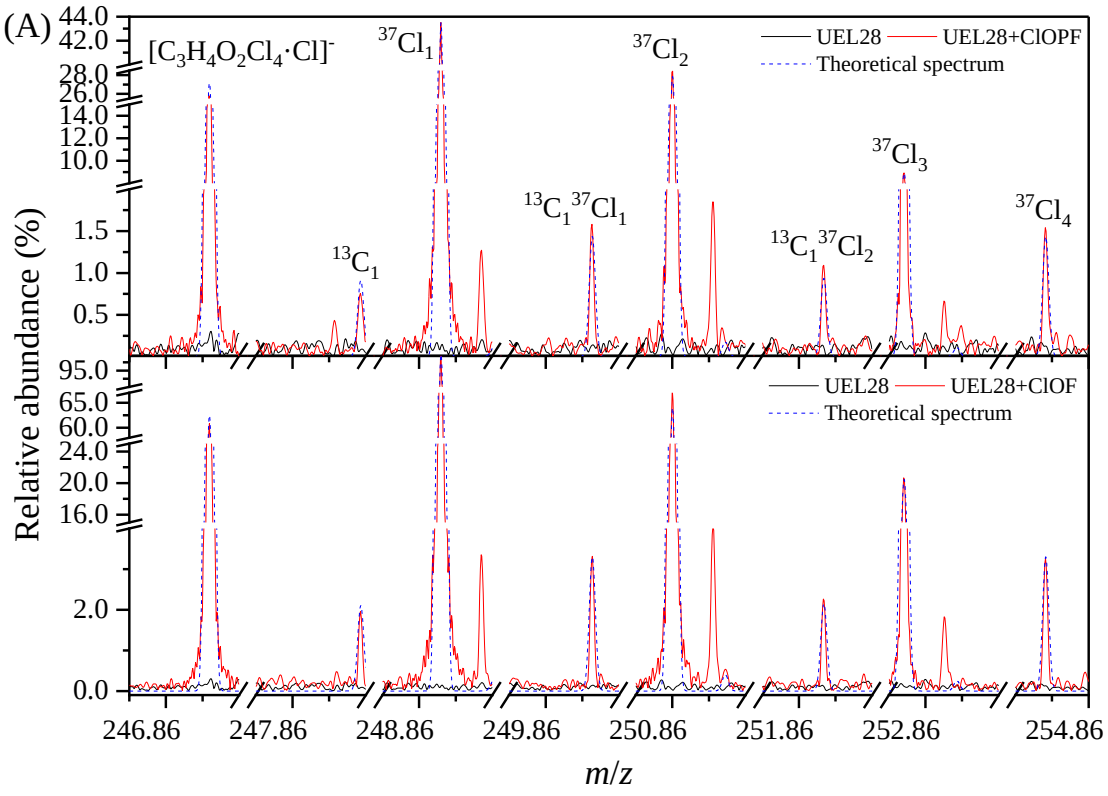
water samples with and without prefiltration (**Figure S15**). A series of X_n-DBPs species containing N and multiple Cl atoms were identified in the chlorinated UEL28 water without prefiltration but absent in the chlorinated UEL28 water with prefiltration, SRNOM1, SRNOM2, chlorinated SRNOM1, and chlorinated SRNOM2 (**Figure S16**). For example, intensities of $C_2H_1O_1N_2Cl_4$ and $C_3H_2O_3N_1S_1Cl_5$ peaks in the chlorinated UEL28 water without prefiltration were approximately 18- and 6-fold of those in the raw and chlorinated UEL28 water with prefiltration, respectively (**Figure S16**). These X_n-DBPs species containing heteroatoms (N and S) and multiple Cl atoms generally have higher ecological risks than those for nitrogen-free X_n-DBPs species with fewer Cl atoms [36]. These observations could be attributed to the additional release of AOM from disinfected algal cells in the UEL water chlorinated without prefiltration compared with that chlorinated after filtration, which further reacts with reactive chlorine to form more chlorinated DBPs [20, 37]. Moreover, the X_n-DBPs species only in the chlorinated UEL water sample without prefiltration had significantly lower ($p < 0.05$) O/C but higher ($p < 0.05$) N/C and (H+X)/C values than those for the X_n-DBPs species both in chlorinated UEL water samples with and without prefiltration (0.48 ± 0.20 versus 0.50 ± 0.15, 0.04 ± 0.08 versus 0.02 ± 0.06, and 1.46 ± 0.37 versus 1.42 ± 0.32, respectively). These results indicate the prevalence of highly saturated and N-containing X_n-DBPs species in the chlorinated UEL water without prefiltration, most likely derived from microbial activities.

The number and its proportion of $N_{n>0}$-DBPs in disinfected UEL water samples were significantly larger ($p < 0.05$) than those for disinfected SRNOM (8 - 206 versus 0 - 4 and 1.60% - 29.70% versus 0% - 0.4%, **Figure S11**). Moreover, as shown in **Figure S14C**, the number of unique $N_{n>0}$-DBPs in disinfected UEL water samples with and without prefiltration (357 and 230, respectively) was also overwhelmingly larger than that for disinfected SRNOM samples (4). The X_n-DBPs formulae identified in disinfected SRNOM samples were clustered between $0.5 < (H+X)/C < 1.4$ and $0.4 < O/C < 0.9$ in the van Krevelen diagram, with the dominant components

of lignin-like and tannin-like compounds [25] (**Figures 2** and **S9**). However, the molecular composition of X_n-DBPs species identified in ≥ 5 disinfected UEL water samples and ≤ 1 background sample were more diverse (mainly 1.25 < (H+X)/C < 2.0 and 0.2 < O/C < 0.6), with more assigned formulae (*i.e.*, aminosugar-like and saturated compounds). Although most X_n-DBPs species in both disinfected UEL water and disinfected SRNOM samples were unsaturated compounds, the X_n-DBPs species identified in ≥ 5 disinfected UEL water samples and ≤ 1 background sample contained more reduced compounds and scattered in regions II (unsaturated and reduced) and III (saturated and reduced) within plots of (DBE-0.5O)/C against NOSC (**Figure S13**). Compared with disinfected SRNOM, 5 - 74 X_n-DBPs species identified in ≥ 5 disinfected UEL water samples and ≤ 1 background sample were ascribed to saturated and reduced compounds (**Figure S13**). This observation indicated that X_n-DBPs precursors in disinfected UEL water samples were more saturated and reduced than in disinfected SRNOM. Moreover, the disinfected SRNOM samples contain significantly higher *m*/*z*, DBE, NOSC, AI_{mod}, Cl/C, and lower (H+X)/C values than disinfected UEL water samples ($p < 0.05$), suggesting higher saturation and lower aromaticity of X_n-DBPs species in disinfected UEL water samples.

There were 789 unique high bioavailable X_n-DBPs species ((H+X)/C ≥ 1.5, excluding isotopic formulae) identified in the UEL water with the number- and intensity-averaged formula of $C_{13.17}H_{21.61}O_{5.20}N_{0.58}S_{0.03}Cl_{1.70}$ and $C_{13.84}H_{22.77}O_{5.36}N_{0.33}S_{0.03}Cl_{1.57}$, but absent in the chlorinated SRNOM solution [38, 39]. For example, the X_n-DBPs species with three carbons and (H+X)/C ≥ 1.5 identified in this study had low predicted biotransformation half-life (0.02 - 1.67 days) (**Table S5**), suggesting the high bioavailability of those X_n-DBPs species [39]. Compared with X_n-DBPs species exclusively in SRNOM, the X_n-DBPs only present in the UEL water were generally saturated and N-rich, suggesting the obviously different molecular properties of autochthonous precursors and allochthonous precursors for X_n-DBPs formation [20]. These results indicated that microbial growth would promote the

formation of bioavailable X_n-DBPs species during the disinfection process. The molecular formulae of these saturated X_n-DBPs species exclusively in the disinfected UEL water were typically assigned to $C_nH_{2n}O_{3\text{-}5}Cl_2$ and $C_nH_{2n+1}O_{2\text{-}4}Cl_1$ series (n is an integer, $8 \leq n \leq 19$) (**Figure S17**). The CH_2 homologous series of these typical X_n-DBPs were related to alkyl chain elongation and chemical reactions, including methylation/demethylation reactions [40]. Moreover, the Cl_5-DBPs ($[C_3H_4O_2Cl_4 \cdot Cl]^-$, m/z = 246.865941) and Cl_6-DBPs ($[C_3H_3O_2Cl_5 \cdot Cl]^-$, m/z = 280.826969) ions were identified as the Cl-adduct ions (**Figure 3**) because HCl was not added to the neutral formulae ($C_3H_4O_2Cl_4$ and $C_3H_3O_2Cl_5$, DBE=0) by addition reaction; otherwise, the neutral chemical formulae were chemical impossible. There were 119 $Cl_{n\geq2}$-DBPs identified in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample, further supporting that saturated X_n-DBPs species containing multiple chlorine atoms in the chlorinated UEL water was mainly derived from microbial activities.



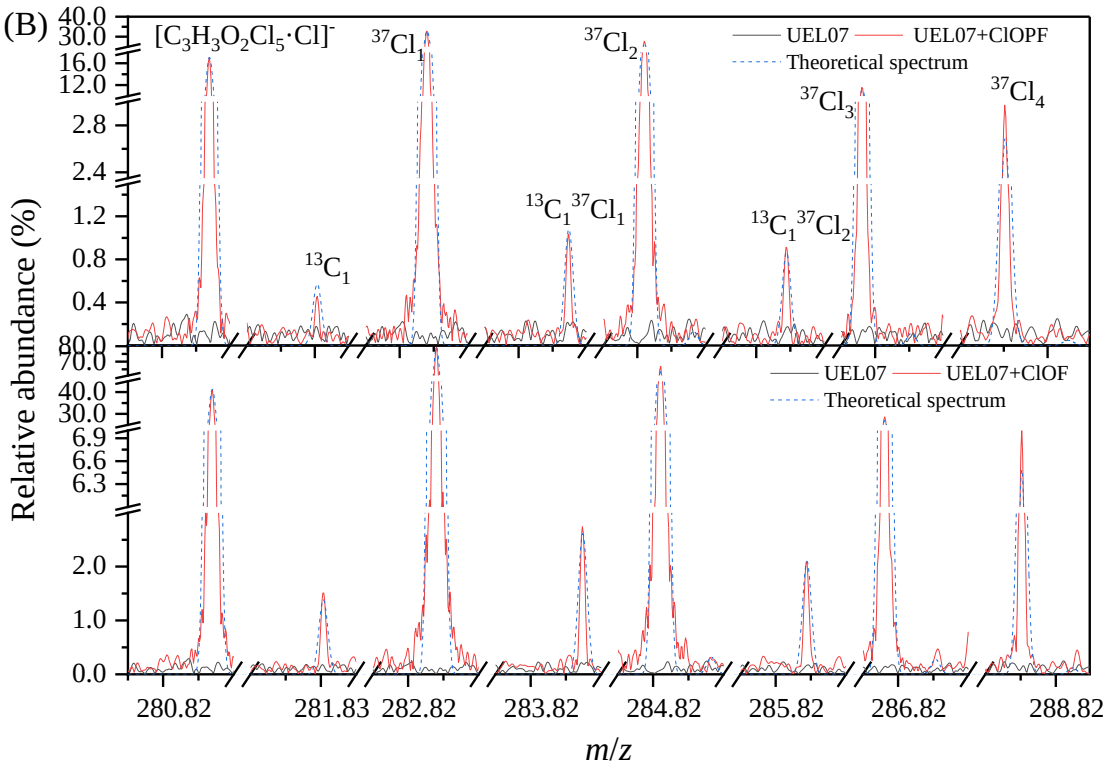


Figure 3. Enlarged representative FT-ICR MS spectra of the negatively charged Cl-adduct ions. The measured spectra and theoretical spectra of (A) $[C_3H_4O_2Cl_4 \cdot Cl]^-$ (for samples UEL28, UEL28+ClOPF, and UEL28+ClOF) and (B) $[C_3H_3O_2Cl_5 \cdot Cl]^-$ (for samples UEL07, UEL07+ClOPF, and UEL07+ClOF).

The possible chemical structures of all X_n-DBPs species with 3 carbons identified in this study were retrieved from the PubChem database due to their high bioavailability ((H+X)/C ≥ 1.5) [39] (**Table S5**). The DBE value of a typical saturated X_n-DBPs compound, $C_3H_4O_2Cl_4$ (**Figure 3A**), was zero, suggesting its absence of double bond and the impossibility of forming the false true ion $[C_3H_5O_2Cl_5]^-$ (DBE = -1) ion by adding H+Cl atoms in the ESI(-)-FT-ICR MS spectrum. Similarly, the peak clusters at *m*/*z* =280.826932, 282.823980, 284.821030, 286.818079, and 288.815169 (mass error and intensity deviation were -0.13 ± 0.01 ppm and 5.2% ± 4.7%, respectively, **Figure 3B**) were annotated to the Cl-adduct of $C_3H_3O_2Cl_5$ (DBE = 0).

All 12 C_3-bearing X_n-DBPs were of a highly saturated nature (0 ≤ DBE ≤ 3) with their predicted toxicities or the number-average toxicities of all possible isomers tabulated in **Table S5**. As exemplified by the predicted toxicity of 96-h LC_{50} for green

algae, two-thirds of C_3-bearing X_n-DBPs isomers in chlorinated UEL water samples could be identified as harmful compounds (**Table S5**). In addition to one toxic isomer [2.1 mg/L for 1,1,3,3-tetrachloro-2-hydroperoxypropan-2-ol ($C_3H_4O_3Cl_4$)], three other isomers [0.7 mg/L, 0.4 mg/L, and 0.2 mg/L for 2,3,3-trichloropropanamide ($C_3H_4O_1N_1Cl_3$), 2,2-dichloro-N-(chloromethyl)acetamide ($C_3H_4O_1N_1Cl_3$), and 2,2,2-trichloro-N-cyanoacetamide ($C_3H_1O_1N_2Cl_3$)] were very toxic. The $C_3H_1O_1N_2Cl_3$ had the highest toxicity for green algae (96-h LC_{50} = 0.2 mg/L, **Figure S18** and **Table S5**), related to its N-rich nature [36, 41]. The LC_{50} and EC_{50} values of C_3-bearing X_n-DBPs species detected only in the disinfected UEL water samples were significantly (p < 0.05) lower than those for C_3-bearing X_n-DBPs species in all disinfected samples (**Figure S18**). Moreover, three C_3-bearing X_n-DBPs detected both in chlorinated UEL water and SRNOM samples were classified as harmless (LC_{50}/EC_{50} values of $C_3H_3O_3Cl_3$, $C_3H_2O_1Cl_4$, and $C_3H_1O_1Cl_5$ were 849.3 -3520.0 mg/L, 274.5 - 1086.6 mg/L, and 128.0 - 365.6 mg/L, respectively). The predicted toxicity of C_3-bearing N_0-DBPs species was generally positively related to their Cl atom number [36]. For example, the averaged predicted toxicity of 96-h LC_{50} for fish was in the order of $C_3H_4O_2Cl_4$ (least toxic, 195.5 mg/L) < $C_3H_3O_1Cl_5$ (moderate toxic, 74.0 mg/L) < $C_3H_2O_1Cl_6$ (most toxic, 28.4 mg/L). In contrast, the toxic order of 48-h LC_{50} for daphnid was $C_3H_3O_3Cl_3$ (least toxic, 849.3 mg/L) < $C_3H_2O_1Cl_4$ (moderate toxic, 557.4 mg/L) < $C_3H_1O_1Cl_5$ (most toxic, 200.0 mg/L, **Figure S18**). Although the toxicity of a limited number of highly saturated DBPs was evaluated in this study, these results have highlighted the different molecular diversity and toxicity of microbe-derived and terrestrially X_n-DBPs species. Additionally, two highly saturated C_3-bearing X_n-DBPs [$C_3H_3O_3N_1S_1Cl_4$ (identified in ≥ 5 disinfected UEL water samples and ≤ 1 background sample) and $C_3H_2O_3N_1S_1Cl_5$] were detected only in disinfected UEL water samples. However, no information about these two highly saturated C_3-bearing X_n-DBPs is retrievable in the PubChem database, highlighting the necessity of investigating unknown X_n-DBPs species derived from microbial activities.

3.4 Metabolic pathways responses for X_n-DBPs formation

The metabolic pathways collated in the KEGG metabolome database revealed the microbial metabolic pathway for X_n-DBPs precursors determined according to the reaction stoichiometry of the substitution and addition reactions [26]. In total, 240 and 981 precursors were identified exclusively in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample (**Table S6**), and both in ≥ 5 chlorinated UEL water samples and chlorinated SRNOM, respectively. There were 392 and 1,003 KEGG compounds matched for these precursors, respectively (**Table S7**). Moreover, 20 and 66 KEGG compounds in metabolic pathway maps were identified for the precursors exclusively in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample, and both in ≥ 5 chlorinated UEL water and chlorinated SRNOM samples, respectively (**Figures 4** and **Table S7**). There were 10 and 33 unique structures retrieved from the KEGG database for these KEGG compounds, respectively. The metabolites involved in the metabolic pathways of these precursors were then scattered in the KEGG metabolic pathway map. As depicted in **Figure 4**, "lipid metabolism" is an important metabolic pathway for these precursors exclusively in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample (7 KEGG compounds). In comparison, precursors both in ≥ 5 chlorinated UEL water and chlorinated SRNOM samples (1 KEGG compound) were less involved in this metabolic pathway (**Figure 4** and **Table S7**). Moreover, no KEGG compounds in metabolic pathway maps were identified for the precursors exclusively in ≥ 5 chlorinated UEL water samples and ≥ 2 background samples. The unique and considerable contribution of lipid metabolism to the precursors exclusively identified in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample suggests that lipid metabolism is an important metabolic pathway of microbe-derived precursors because of the absence of related KEGG compound identified in chlorinated SRNOM and ≥ 2 background samples. Thus, the saturated X_n-DBPs species is expected to be generated from highly saturated autochthonous precursors yielded by lipid metabolism. Moreover, amino acid metabolism (6 KEGG compound)

is also involved in the synthesis of precursors mainly derived from microbial activities, followed by the metabolism of cofactors and vitamins (4 KEGG compound), nucleotide metabolism (2 KEGG compound), and carbohydrate metabolism (1 KEGG compound).

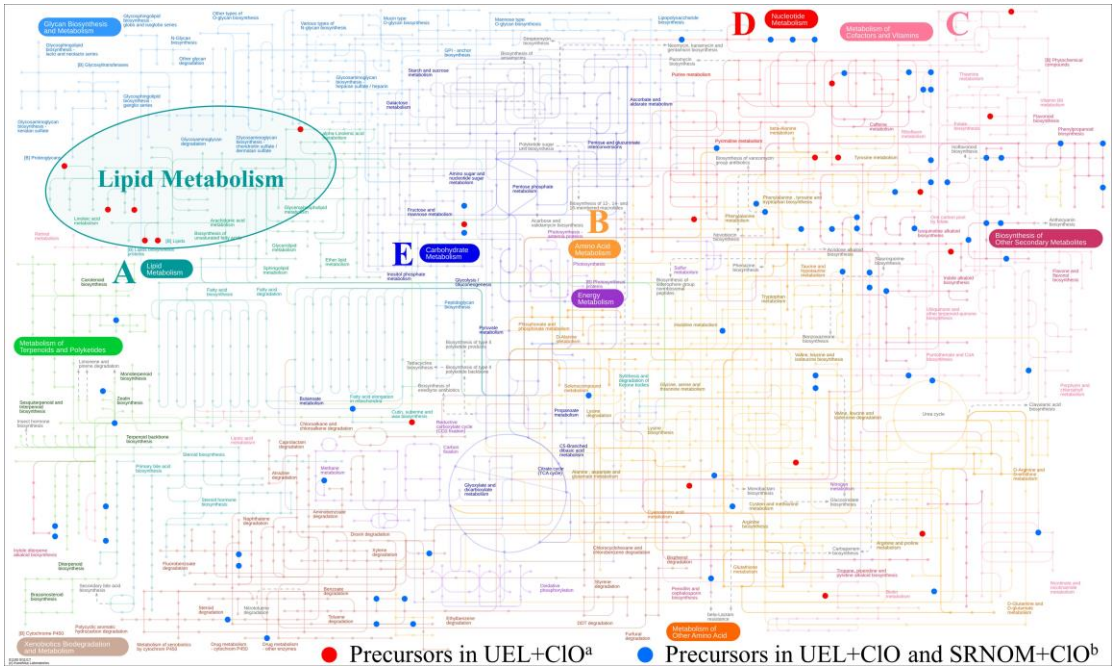


Figure 4. KEGG pathways map of the three categorized X_n-DBPs precursors matched with KEGG compounds using Interactive Pathway Explorer v3.0. Colour code of nodes: Precursors in UEL+ClO[a]: precursors identified exclusively in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample: red; Precursors in UEL+ClO and SRNOM+ClO[b]: precursors identified both in ≥ 5 chlorinated UEL water and chlorinated SRNOM samples: blue. The order of the letter codes represents the order of the most likely involved metabolism for precursors identified only in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample: Lipid metabolism (A); Amino acid metabolism (B); Metabolism of cofactors and vitamins (C); Nucleotide metabolism (D); and Carbohydrate metabolism (E). The dominant type of metabolism highlighted with green circles is lipid metabolism. Note that dots of different colours may overlap.

The KEGG compounds identified in the lipid metabolic pathway were mainly responsible for lipid metabolism, fatty acid metabolism, linoleic acid metabolism, and alpha-linolenic acid metabolism (**Figure 4**). Over 40 $Cl_{1\text{-}5}$-DBPs species were identified from fatty acids during the disinfection of *M. aeruginosa* cells [20]. For example, a series of Cl_n-DBPs species was formed from the reaction between active chlorine and fatty acids, including linoleic acid ($C_{18}H_{32}O_2$), γ-linolenic acid ($C_{18}H_{30}O_2$), stearidonic acid ($C_{18}H_{28}O_2$), and 17-hydroxylinolenic acid ($C_{18}H_{30}O_3$). Indeed, KEGG-matched precursors identified only in ≥ 5 disinfected UEL water samples and ≤ 1 background sample were featured in a high degree of C-C bond saturation (-1 ≤ DBE-0.5O ≤ 4, **Table S8**). These precursors could be assigned to fatty acid molecules containing several hydroxyl and/or carboxyl functional groups and double bonds because negative and positive values of DBE-0.5O signify the presence of saturated function groups (*e.g.*, hydroxyl, methoxy, or ether groups) and unsaturated functional groups (*e.g.*, double bonds and rings) in a molecule [42]. In this study, 14 unique formulae of KEGG-matched precursors were related to 32 saturated $Cl_{1\text{-}5}$-DBPs species (**Tables S8** and **S9**). There were six saturated Cl_1-DBPs ($C_{10}H_{13}O_5N_2Cl_1$, $C_{10}H_{15}O_5N_2Cl_1$, $C_{16}H_{29}O_4Cl_1$, $C_{16}H_{31}O_5Cl_1$, $C_{18}H_{31}O_5Cl_1$, and $C_{18}H_{33}O_5Cl_1$) exclusively identified in ≥ 5 chlorinated UEL water samples and ≤ 1 background sample (**Table S9**). The KEGG-matched precursors exclusively in ≥ 5 disinfected UEL water samples exhibited distinct compositional space in the van Krevelen diagram, which differed from that observed both in ≥ 5 chlorinated UEL water and chlorinated SRNOM samples. In addition to more N-containing compounds, most KEGG-matched precursors detected exclusively in ≥ 5 disinfected UEL water samples and ≤ 1 background sample was saturated, aminosugar-like, and reduced compounds (**Figure S19**). The microbial oxidation (such as aerobic respiration) of highly oxidized (higher NOSC) DOM is more thermodynamically favorable and tends to be removed or consumed [43, 44]. The NOSC and ΔG^o_{Cox} values of X_n-DBPs precursors detected exclusively in ≥ 5 disinfected UEL water samples and ≤ 1

603 background sample were significantly lower and higher ($p < 0.05$) than those for the
604 precursors identified both in ≥ 5 chlorinated UEL water and chlorinated SRNOM
605 samples (**Figure S20**), suggesting that autochthonous X_n-DBPs precursors exclusively
606 identified in the UEL water were less thermodynamically favourable and easily
607 consumed, thus yielding higher ecological and human risks than allochthonous X_n-
608 DBPs precursors.

609 ## 4. Environmental Implications

610 Contrasting to the SRNOM, a representative of terrestrially derived aquatic
611 NOM [24], the UEL water has more autochthonous and bioavailable X_n-DBPs
612 precursors. In addition to the investigated Lake A, eutrophication and algal blooms
613 were often reported in other drinking water sources, such as Biwa Lake in Japan [45],
614 Okeechobee Lake in the USA [46] and Taihu Lake, Chaohu Lake, and East Lake in
615 China [2, 47]. Although slight eutrophication exhibited minor effects on the molecular
616 composition of aquatic DOM, results of this study have demonstrated that substantial
617 microbial growth yields more autochthonous DOM and promotes the formation of
618 more highly saturated and bioavailable X_n-DBPs species in Lake A upon disinfection.
619 Furthermore, these X_n-DBPs species with multiple N and Cl atoms (*e.g.*, unknown
620 $C_3H_3O_3N_1S_1Cl_4$ and $C_3H_3O_3N_1S_1Cl_5$) mainly derived from autochthonous DOM and
621 generally more toxic than allochthonous-derived X_n-DBPs species, highlighting the
622 significance of appropriate treatments for eutrophic source water before disinfection.
623 In this study, compared with few known regulated X_n-DBPs, the structure and toxicity
624 of the approximately 1,500 unique unknown X_n-DBPs (including 272 microbe-
625 derived X_n-DBPs) identified in the chlorinated UEL water samples with or without
626 prefiltration warrants further investigation. Additionally, the similar molecular
627 properties (*i.e.*, saturated and reduced fractions) of DOM compounds in the
628 UEL28+GAC sample to microbial-derived X_n-DBPs precursors (**Figures S21**, **S22**
629 and **Table S11**) suggest that autochthonous DOM in GAC-treated eutrophic lake
630 water will contribute to the formation of microbial-derived DBPs. Elucidating the

chemical composition of DOM contributes primarily to tracking and restoring the health of fresh waters [48]. In this study, for the first time, the relationship between microbial metabolism and X_n-DBPs species was established by linking X_n-DBPs precursors with the KEGG compounds, which enables the in-depth tracking of the sources of autochthonous X_n-DBPs precursors in terms of metabolic pathways. In addition to the unique contribution of lipid metabolism, other metabolisms (*e.g.*, amino acid metabolism) are also crucial for the precursors identified only in the chlorinated natural water with algae, suggesting that these microbial biometabolites are most likely involved in forming microbe-derived X_n-DBPs species during disinfection.

In eutrophic lakes, various autochthonous biometabolites, including carbohydrates, lipids, and proteins, are present throughout the year [5]. The molecular composition of DOM in eutrophic lakes (*e.g.*, Taihu Lake) is dominated by microbial metabolism in a temporal-dependent manner [5]. For example, the abundance of cyanobacteria gradually increased from August to the highest level in September, followed by a gradual decrease in October [31]. Thus, microbial metabolism (*e.g.*, lipid metabolism) is expected to be more significant during the growth phase in the summer. Algae cells yield the highest DOM per cell during the late exponential or death phase [11]. In this study, the algae were generally in the lag or early exponential phase. Therefore, it is expected that the release of autochthonous DOM and the formation of microbe-derived X_n-DBPs during disinfection will be more significant when algae in the UEL water is in the late exponential or death phase. Moreover, the inactivation of pathogenic bacteria (*e.g.*, *Escherichia coli*) using bactericides in water will release autochthonous substances, such as proteins, from the pathogenic cells [49]. These results suggest that disinfection may cause the release of different autochthonous substances from different microbes, which consequently contributes to the formation of different X_n-DBPs species. Although the inactivation of pathogenic bacteria and the detection of X_n-DBPs species have been extensively reported, the connection between

659 microbial metabolism and microbe-derived X_n-DBPs species is generally overlooked.
660 In addition to the important role of the precursors mainly derived from microbial
661 activities in the formation of highly saturated X_n-DBPs species with high ecological
662 and health risks upon the disinfection, this study has highlighted the significance of
663 linking microbial metabolism with X_n-DBPs species to shed new insights in treatment
664 and governance of eutrophic source water.

665 ## Declaration of competing interest

666 The authors declare no competing financial interest.

667 ## Acknowledgements

668 This study was financially supported by the National Natural Science Foundation
669 of China (No. 42477006), the Japan Society for the Promotion of Science (No.
670 23KK0074), the Opening Fund of the State Key Laboratory of Environmental
671 Geochemistry (SKLEG2024215), and the Fundamental Research Funds for the
672 Central Universities, China University of Geosciences (Wuhan). The authors
673 appreciated the assistance in field sampling and FT-ICR MS measurement from
674 Qiguang Shan and Hiroyuki Momma, respectively.

675 ## Data availability

676 All FT-ICR MS spectra and formula assignment results are freely available at
677 https://doi.org/10.6084/m9.figshare.25996630.v2.

678 ## References

679 (1) Wu, S.; Fujii, M.; Yang, X.; Fu, Q.-L. 2023, Characterization of halogenated
680 organic compounds by the Fourier transform ion cyclotron resonance mass
681 spectrometry: A critical review. Water Res. 246, 120694. DOI:
682 10.1016/j.watres.2023.120694.
683 (2) Liu, S.; Hou, J.; Suo, C.; Chen, J.; Liu, X.; Fu, R.; Wu, F. 2022, Molecular-level
684 composition of dissolved organic matter in distinct trophic states in Chinese lakes:
685 Implications for eutrophic lake management and the global carbon cycle. Water Res.
686 217,118438. DOI: 10.1016/j.watres.2022.118438.
687 (3) Liu, S.; He, Z.; Tang, Z.; Liu, L.; Hou, J.; Li, T.; Zhang, Y.; Shi, Q.; Giesy, J. P.;
688 Wu, F. 2020, Linking the molecular composition of autochthonous dissolved organic
689 matter to source identification for freshwater lake ecosystems by combination of
690 optical spectroscopy and FT-ICR-MS analysis. Sci. Total Environ. 703,134764. DOI:

10.1016/j.scitotenv.2019.134764.
(4) Qi, J.; Lan, H.; Liu, R.; Miao, S.; Liu, H.; Qu, J. 2016, Prechlorination of algae-laden water: The effects of transportation time on cell integrity, algal organic matter release, and chlorinated disinfection byproduct formation. Water Res. 102, 221-228. DOI: 10.1016/j.watres.2016.06.039.
(5) Zhang, F.; Harir, M.; Moritz, F.; Zhang, J.; Witting, M.; Wu, Y.; Schmitt-Kopplin, P.; Fekete, A.; Gaspar, A.; Hertkorn, N. 2014, Molecular and structural characterization of dissolved organic matter during and post cyanobacterial bloom in Taihu by combination of NMR spectroscopy and FTICR mass spectrometry. Water Res. 57, 280-294. DOI: 10.1016/j.watres.2014.02.051.
(6) Hu, C.; Xu, H.; Shi, S.; Lan, J.; Zhou, K. e.; Zhang, J.; Song, Y.; Wang, J.; Fu, P. 2023, Sedimentary organic matter molecular composition reveals the eutrophication of the past 500 years in Lake Daihai, Inner Mongolia. Environ. Res. 111,115753. DOI: 10.1016/j.envres.2023.115753.
(7) McIntyre, A. M.; Gueguen, C. 2013, Binding interactions of algal-derived dissolved organic matter with metal ions. Chemosphere 90 (2), 620-626. DOI: 10.1016/j.chemosphere.2012.08.057.
(8) Zhao, Z.; Gonsior, M.; Schmitt-Kopplin, P.; Zhan, Y. C.; Zhang, R.; Jiao, N. Z.; Chen, F. 2019, Microbial transformation of virus-induced dissolved organic matter from picocyanobacteria: coupling of bacterial diversity and DOM chemodiversity. ISME J. 13 (10), 2551-2565. DOI: 10.1038/s41396-019-0449-1.
(9) Longnecker, K.; Kujawinski, E. B. 2011, Composition of dissolved organic matter in groundwater. Geochim. Cosmochim. Acta 75 (10), 2752-2761. DOI: 10.1016/j.gca.2011.02.020.
(10) Fu, Q.-L.; Chen, C.; Liu, Y.; Fujii, M.; Fu, P. 2023, FT-ICR MS Spectral Improvement of Dissolved Organic Matter by the Absorption Mode: A Comparison of the Electrospray Ionization in Positive-Ion and Negative-Ion Modes. Anal. Chem. 96 (1), 522-530. DOI: 10.1021/acs.analchem.3c04651.
(11) Hua, L. C.; Lai, C. H.; Wang, G. S.; Lin, T. F.; Huang, C. P. 2019, Algogenic organic matter derived DBPs: Precursor characterization, formation, and future perspectives - A review. Crit. Rev. Environ. Sci. Technol. 49 (19), 1803-1834. DOI: 10.1080/10643389.2019.1586057.
(12) Wert, E. C.; Rosario-Ortiz, F. L. 2013, Intracellular Organic Matter from Cyanobacteria as a Precursor for Carbonaceous and Nitrogenous Disinfection Byproducts. Environ. Sci. Technol. 47 (12), 6332-6340. DOI: 10.1021/es400834k.
(13) Aziz, M. T.; Granger, C. O.; Westerman, D. C.; Putnam, S. P.; Ferry, J. L.; Richardson, S. D. 2022, *Microseira wollei* and *Phormidium* algae more than doubles DBP concentrations and calculated toxicity in drinking water. Water Res. 216,118316. DOI: 10.1016/j.watres.2022.118316.
(14) Aziz, M. T.; Granger, C. O.; Ferry, J. L.; Richardson, S. D. 2023, Algae impacted drinking water: Does switching to chloramination produce safer drinking water? Sci. Total Environ. 877,162815. DOI: 10.1016/j.scitotenv.2023.162815.

(15) Wang, X. X.; Liu, B. M.; Lu, M. F.; Li, Y. P.; Jiang, Y. Y.; Zhao, M. X.; Huang, Z. X.; Pan, Y.; Miao, H. F.; Ruan, W. Q. 2021, Characterization of algal organic matter as precursors for carbonaceous and nitrogenous disinfection byproducts formation: Comparison with natural organic matter. J. Environ. Manage. 282,111951. DOI: 10.1016/j.jenvman.2021.111951.
(16) Zhang, T.-Y.; Lin, Y.-L.; Xu, B.; Cheng, T.; Xia, S.-J.; Chu, W.-H.; Gao, N.-Y. 2016, Formation of organic chloramines during chlor(am)ination and UV/chlor(am)ination of algae organic matter in drinking water. Water Res. 103, 189-196. DOI: 10.1016/j.watres.2016.07.036.
(17) Chen, H.; Tsai, K.-P.; Liu, Y.; Tolić, N.; Burton, S. D.; Chu, R.; Karanfil, T.; Chow, A. T. 2021, Characterization of Dissolved Organic Matter from Wildfire-induced Microcystis aeruginosa Blooms controlled by Copper Sulfate as Disinfection Byproduct Precursors Using APPI(-) and ESI(-) FT-ICR MS. Water Res. 189, 116640. DOI: 10.1016/j.watres.2020.116640.
(18) Wu, Y. W.; Bu, L. J.; Zhu, S. M.; Chen, F.; Li, T. B.; Zhou, S. Q.; Shi, Z. 2022, Molecular transformation of algal organic matter during sequential ozonation-chlorination: Role of pre-ozonation and properties of chlorinated disinfection byproducts. Water Res. 223,119008. DOI: 10.1016/j.watres.2022.119008.
(19) Wu, Y.; Bu, L.; Duan, X.; Zhou, S.; Crittenden, J. C. 2022, Insights into the molecular compositions of CX_3R-type disinfection byproduct precursors in algal organic matter from algae-laden water. Chem. Eng. J. 446,136921. DOI: 10.1016/j.cej.2022.136921.
(20) Gonsior, M.; Powers, L. C.; Williams, E.; Place, A.; Chen, F.; Ruf, A.; Hertkorn, N.; Schmitt-Kopplin, P. 2019, The chemodiversity of algal dissolved organic matter from lysed Microcystis aeruginosa cells and its ability to form disinfection by-products during chlorination. Water Res. 155, 300-309. DOI: 10.1016/j.watres.2019.02.030.
(21) Wang, R.; Zhou, J.; Qu, G.; Wang, T.; Jia, H.; Zhu, L. 2022, FT-ICR/MS deciphers formation of unknown macromolecular disinfection byproducts from algal organic matters after plasma oxidation. Water Res. 218, 118492. DOI: 10.1016/j.watres.2022.118492.
(22) McCarthy, M. J.; Lavrentyev, P. J.; Yang, L.; Zhang, L.; Chen, Y.; Qin, B.; Gardner, W. S. 2007, Nitrogen dynamics and microbial food web structure during a summer cyanobacterial bloom in a subtropical, shallow, well-mixed, eutrophic lake (Lake Taihu, China). Hydrobiologia 581, 195-207. DOI: 10.1007/s10750-006-0496-2.
(23) Dong, H.; Aziz, M. T.; Richardson, S. D. 2023, Transformation of Algal Toxins during the Oxidation/Disinfection Processes of Drinking Water: From Structure to Toxicity. Environ. Sci. Technol. 57 (35), 12944-12957. DOI: 10.1021/acs.est.3c01912.
(24) Chin, Y. P.; McKnight, D. M.; D'Andrilli, J.; Brooks, N.; Cawley, K.; Guerard, J.; Perdue, E. M.; Stedmon, C. A.; Tratnyek, P. G.; Westerhoff, P.; et al. 2023, Identification of next-generation International Humic Substances Society reference materials for advancing the understanding of the role of natural organic matter in the

Anthropocene. Aquat. Sci. 85 (1), 32,32. DOI: 10.1007/s00027-022-00923-x.
(25) Fu, Q.-L.; Fujii, M.; Kwon, E. 2020, Development and Application of a High-Precision Algorithm for Nontarget Identification of Organohalogens Based on Ultrahigh-Resolution Mass Spectrometry. Anal. Chem. 92 (20), 13989-13996. DOI: 10.1021/acs.analchem.0c02899.
(26) Fu, Q.-L.; Fujii, M.; Watanabe, A.; Kwon, E. 2022, Formula Assignment Algorithm for Deuterium-Labeled Ultrahigh-Resolution Mass Spectrometry: Implications of the Formation Mechanism of Halogenated Disinfection Byproducts. Anal. Chem. 94 (3), 1717-1725. DOI: 10.1021/acs.analchem.1c04298.
(27) Dittmar, T.; Koch, B.; Hertkorn, N.; Kattner, G. 2008, A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol. Oceanogr. Meth. 6 (6), 230-235. DOI: 10.4319/lom.2008.6.230.
(28) Liu, W.; Li, Y.; Liu, F.; Jiang, W.; Zhang, D.; Liang, J. 2019, Visible-light-driven photocatalytic degradation of diclofenac by carbon quantum dots modified porous g-C_3N_4: Mechanisms, degradation pathway and DFT calculation. Water Res. 151, 8-19. DOI: 10.1016/j.watres.2018.11.084.
(29) Zhang, J.; Chen, Y.; Song, X.; Liu, Y.; Zhao, J.; Wang, F. 2022, Synergistic adsorption and degradation of diclofenac by zero-valent iron modified spent bleaching earth carbon: Mechanism and toxicity assessment. J. Hazard. Mater. 432,128753. DOI: 10.1016/j.jhazmat.2022.128753.
(30) U.S.EPA. *Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11*. United States Environmental Protection Agency, Washington, DC, USA., 2012. https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface#citing (accessed 2024 15 June).
(31) Arii, S.; Yamashita, R.; Tsuji, K.; Tomita, K.; Hasegawa, M.; Bober, B.; Harada, K. 2021, Differences in susceptibility of cyanobacteria species to lytic volatile organic compounds and influence on seasonal succession. Chemosphere 284,131378. DOI: 10.1016/j.chemosphere.2021.131378.
(32) Fang, J.; Ma, J.; Yang, X.; Shang, C. 2010, Formation of carbonaceous and nitrogenous disinfection by-products from the chlorination of *Microcystis aeruginosa*. Water Res. 44 (6), 1934-1940. DOI: 10.1016/j.watres.2009.11.046.
(33) Hua, L.-C.; Tsia, S. R.; Ngo, D. N. G.; Huang, C. 2021, Bromide-intrusion into *Chlorella* sp. and *Microcystis aeruginosa* growing environments: Its impacts on algal growth and the formation potential of algal-derived DBPs upon chlorination. Sci. Total Environ. 795,148772. DOI: 10.1016/j.scitotenv.2021.148772.
(34) Wang, T. T.; Yang, X.; Li, Z. H.; Chen, W. H.; Wen, X.; He, Y. B.; Ma, C.; Yang, Z. Z.; Zhang, C. 2023, MeHg production in eutrophic lakes: Focusing on the roles of algal organic matter and iron-sulfur-phosphorus dynamics. J. Hazard. Mater. 457,131682. DOI: 10.1016/j.jhazmat.2023.131682.
(35) Wu, S.; Fujii, M.; Fu, Q. 2023, Molecular characterization of coastal seawater dissolved organic matter by ultrahigh-resolution mass spectrometry: a photochemical study of the Tokyo Bay, Japan. Carbon Res. 2 (1), 46. DOI: 10.1007/s44246-023-

00083-z.
(36) Hanigan, D.; Truong, L.; Simonich, M.; Tanguay, R.; Westerhoff, P. 2017, Zebrafish embryo toxicity of 15 chlorinated, brominated, and iodinated disinfection by-products. J. Environ. Sci. 58, 302-310. DOI: 10.1016/j.jes.2017.05.008.
(37) Qi, J.; Lan, H. C.; Liu, R. P.; Miao, S. Y.; Liu, H. J.; Qu, J. H. 2016, Prechlorination of algae-laden water: The effects of transportation time on cell integrity, algal organic matter release, and chlorinated disinfection byproduct formation. Water Res. 102, 221-228. DOI: 10.1016/j.watres.2016.06.039.
(38) Gao, S.-X.; Zhang, X.; Fan, W.-Y.; Sheng, G.-P. 2021, Molecular insight into the variation of dissolved organic phosphorus in a wastewater treatment plant. Water Res. 203, 117529,117529. DOI: 10.1016/j.watres.2021.117529.
(39) D'Andrilli, J.; Cooper, W. T.; Foreman, C. M.; Marshall, A. G. 2015, An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Commun. Mass Spectrom. 29 (24), 2385-2401. DOI: 10.1002/rcm.7400.
(40) Kim, S.; Kramer, R. W.; Hatcher, P. G. 2003, Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. Anal. Chem. 75 (20), 5336-5344. DOI: 10.1021/ac034415p.
(41) Allen, J. M.; Plewa, M. J.; Wagner, E. D.; Wei, X.; Bokenkamp, K.; Hur, K.; Jia, A.; Liberatore, H. K.; Lee, C.-F. T.; Shirkhani, R.; et al. 2022, Drivers of Disinfection Byproduct Cytotoxicity in U.S. Drinking Water: Should Other DBPs Be Considered for Regulation? Environ. Sci. Technol. 56 (1), 392-402. DOI: 10.1021/acs.est.1c07998.
(42) Lechtenfeld, O. J.; Kattner, G.; Flerus, R.; McCallister, S. L.; Schmitt-Kopplin, P.; Koch, B. P. 2014, Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. Geochim. Cosmochim. Acta 126, 321-337. DOI: 10.1016/j.gca.2013.11.009.
(43) Boye, K.; Noel, V.; Tfaily, M. M.; Bone, S. E.; Williams, K. H.; Bargar, J. R.; Fendorf, S. 2017, Thermodynamically controlled preservation of organic carbon in floodplains. Nat. Geosci. 10 (6), 415-419. DOI: 10.1038/ngeo2940.
(44) Garayburu-Caruso, V. A.; Stegen, J. C.; Song, H.-S.; Renteria, L.; Wells, J.; Garcia, W.; Resch, C. T.; Goldman, A. E.; Chu, R. K.; Toyoda, J.; et al. 2020, Carbon Limitation Leads to Thermodynamic Regulation of Aerobic Metabolism. Environ. Sci. Technol. Lett. 7 (7), 517-524. DOI: 10.1021/acs.estlett.0c00258.
(45) Gurung, T. B.; Urabe, J. 1999, Temporal and vertical difference in factors limiting growth rate of heterotrophic bacteria in Lake Biwa. Microb. Ecol. 38 (2), 136-145. DOI: 10.1007/s002489900167.
(46) Havens, K. E.; Beaver, J. R.; East, T. L. 2007, Plankton biomass partitioning in a eutrophic subtropical lake: comparison with results from temperate lake ecosystems. J. Plankton Res. 29 (12), 1087-1097. DOI: 10.1093/plankt/fbm083.
(47) Liu, H.; He, B. Y.; Zhou, Y. D.; Yang, X. Q.; Zhang, X. Y.; Xiao, F.; Feng, Q.; Liang, S. W.; Zhou, X. M.; Fu, C. J. 2021, Eutrophication monitoring of lakes in Wuhan based on Sentinel-2 data. GISci. Remote Sens. 58 (5), 776-798. DOI:

10.1080/15481603.2021.1940738.
(48) Tanentzap, A. J.; Fonvielle, J. A. 2024, Chemodiversity in freshwater health. Science 383 (6690), 1412-1414. DOI: doi:10.1126/science.adg8658.
(49) Fu, Y. H.; Wang, F.; Sheng, H. J.; Xu, M.; Liang, Y.; Bian, Y. R.; Hashsham, S. A.; Jiang, X.; Tiedje, J. M. 2020, Enhanced antibacterial activity of magnetic biochar conjugated quaternary phosphonium salt. Carbon 163, 360-369. DOI: 10.1016/j.carbon.2020.03.010.