Water Research 268 (2025) 122636

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Characterization of natural and anthropogenic dissolved organic matter in the yangtze river basin using FT-ICR MS

Cuiping Ning [a,b], Shuai Sun [b,c], Yuan Gao [b], Huaijun Xie [d], Lidong Wu [e], Haijun Zhang [b], Jiping Chen [b], Ningbo Geng [b,*]

[a] School of Chemical Engineering, University of Science and Technology Liaoning, Anshan 114051, China
[b] Chinese Academy of Sciences, Dalian Institute of Chemical Physics, Dalian 116023, China
[c] Key Laboratory of Pesticide Environmental Assessment and Pollution Control, Ministry of Ecology and Environment of the People's Republic of China, Nanjing Institute of Environmental Science, Nanjing, 210042, China
[d] Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), School of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian, 116024, China
[e] Key Laboratory of Control of Quality and Safety for Aquatic Products, Chinese Academy of Fishery Sciences, Beijing 100141, China

ARTICLE INFO

Keywords:
Dissolved organic matter
FT-ICR MS
Yangtze River Basin
In-house database
Non-targeted analysis

ABSTRACT

Dissolved organic matter (DOM) is a complex mixture that plays a crucial role in global carbon cycling and climate dynamics. Understanding the chemical composition of DOM is crucial for studying its biogeochemical behavior. However, identifying individual DOM molecules is challenging. Here, using ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry and an in-house database, we developed a framework to investigate DOM characteristics in natural water. Through the developed approach, we successfully identified thousands of individual DOM molecules in the water bodies of the Yangtze River Basin. For the first time, the proportions of natural and anthropogenic organics within DOM were revealed. In total, 9006 unambiguous molecular formulas were assigned to DOM in the Yangtze River Basin. The proportions of CHNO and CHOS compounds increased from upstream to downstream regions. Moreover, 1099 DOM compounds were tentatively identified, with 85 % being endogenous organics and 15 % being exogenous organics. Notably, lipids and pharmaceuticals and personal care products were the most frequently detected endogenous and exogenous compounds. The spatial variation of the identified DOM indicated anthropogenic discharges considerably increased both the number and abundance of DOM in the downstream Yangtze River Basin. This study highlighted the importance of anthropogenic impacts on DOM in water.

1. Introduction

Dissolved organic matter (DOM) is ubiquitous in various natural aqueous environments and plays a profound role in global carbon cycling and climate change (Cole et al., 2007; Kellerman et al., 2014; Lipczynska-Kochany., 2018). It originates from both natural sources (e.g., plant and animal remains, excrement from algae activities) and anthropogenic sources (e.g., soil inputs, sewage discharges) (Bao et al., 2015; Imai et al., 2001). DOM consists of thousands of organic compounds, including humic substances, lipids, organic acids and their derivatives, sugars, proteins, and exogenous pollutants (Leenheer and Croue., 2003; Qi et al., 2022a). In water, DOM can form active intermediates (Canonica and Freiburghaus., 2001), which considerably influence the environmental fate and bioavailability of pollutants, overall water quality, and ecosystem health (Guerard et al., 2009; Yao et al., 2024; Zeng et al., 2013).

Previous studies have explored the macroscopic properties of DOM through ultraviolet–visible spectroscopy, excitation–emission matrix fluorescence spectroscopy, and stable carbon isotopic ($\delta^{13}C$) analysis (He et al., 2022; Hertkorn et al., 2016; Liao et al., 2022; Weishaar et al., 2003). These studies have identified a variety of aromatic organics, protein-like fluorescent components, and humus-like components, with sources including higher plants, microbial photochemical oxidation, and combustion-derived organic matter. Despite these advancements, comprehensively analyzing DOM at the molecular level remains a challenge owing to its complex composition (Qi et al., 2022b; Sandron

---

* Corresponding author.
E-mail address: gengningbo@dicp.ac.cn (N. Geng).

https://doi.org/10.1016/j.watres.2024.122636
Received 3 July 2024; Received in revised form 23 September 2024; Accepted 14 October 2024

0043-1354/

et al., 2015), which makes it nearly impossible to isolate individual compounds. Although the utilization of chromatography-mass spectrometry technology has expanded the potential for DOM compositional analysis, only low-molecular-weight organic acids such as lactic acid, acetic acid, and formic acid have been successfully separated and analyzed (Kvitvang et al., 2011; Weishaar et al., 2003).

The use of Ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) provides excellent analytical capabilities to explore the molecular composition of DOM (Chen et al., 2022; He et al., 2016; Kellerman et al., 2014; Li et al., 2018; Ning et al., 2019). Several studies have used FT-ICR MS to characterize DOM in oceans, lakes, rivers, and other natural water bodies. Consequently, thousands of DOM molecular formulas have been identified (Kellerman et al., 2014; Kida et al., 2021; Kida et al., 2023; Seidel et al., 2022). FT-ICR MS analysis has revealed that the Yangtze River mainstream contains notable amounts of highly aromatic and unsaturated oxygen-containing DOM during dry and cold seasons (Li et al., 2023). Moreover, the abundance and composition of DOM in the Yangtze River Basin are greatly affected by hydrologic and meteorological conditions and anthropogenic activities (Li et al., 2023).

FT-ICR MS shows great promise in analyzing DOM, but thus far, only the molecular formulas and primary functional groups of DOM have been determined through FT-ICR MS analysis. The molecular characterization of DOM represents a breakthrough in the study of its environmental behavior in natural water. Hu *et al.* developed the "energy-diversity-trait" framework via high-throughput sequencing, FT-ICR MS analysis, and ecological networks model, and the DOM–microbe associations were further clarified (Hu et al., 2022). FT-ICR MS shows great promise in identifying DOM, but thus far, only the molecular formulas and primary functional groups of DOM have been determined through FT-ICR MS analysis. Identifying thousands of molecules remains highly challenging owing to the intricate complexity and heterogeneity of DOM in natural water bodies. Furthermore, there is currently a lack of studies exploring the specific components of both natural and anthropogenic DOM. These components can impact the sources, molecular composition, and bioavailability of DOM and potentially alter the ecosystem functions and carbon cycling.

The Yangtze River, the largest river in Asia, holds economic, ecological, and scientific importance (Yang et al., 2007). As it flows from its upstream to downstream areas, natural climate varies, and both population density and economic development increased significantly. These variations can impact the composition and transformation of DOM. Moreover, the expected change in the proportion of stable and labile organics also leads to different dynamics of DOM from upstream to downstream of the Yangtze River Basin. The present study established a new approach based on non-targeted FT-ICR MS analysis and an in-house database to implement DOM screening and molecular identification. The molecular composition, spatial distribution, and sources of DOM along the Yangtze River Basin were investigated. This study is the first to reveal the proportions of natural and anthropogenic DOM along the Yangtze River Basin. These findings provide new insight into DOM molecular characterization and further elucidate its role in aquatic ecosystems.

2. Material and methods

2.1. Sample collection

Water samples were collected along the Yangtze River Basin in September 2022. Fourteen samples were obtained, and the sampling was conducted across the upstream, midstream, and downstream areas. The geographic locations and climatic conditions of the sampling sites are detailed in **Table S1** in Supporting Information (SI). Water samples were collected from the top layer using brown glass bottles pre-rinsed with sample water three times. Each 500 mL bottle was filled, capped quickly, and wrapped with aluminum foil and sealing tape. All samples were stored in a refrigerator at -20 °C until analysis.

2.2. DOM extraction

All samples were filtered through 0.45 μm cellulose filters to remove particulate matter, and the filtrate was adjusted to pH 2 with hydrochloric acid. DOM extraction was performed using hydrophilic-lipophilic balance (HLB) solid-phase extraction (SPE) cartridges (Oasis HLB, Waters, U.S.) that were pre-activated with 3 mL of methanol followed by 3 mL of ultrapure water. The filtrate was loaded onto the SPE cartridge, washed with 3 mL of ultrapure water, and then eluted twice with 3 mL of methanol. The eluent was concentrated using a gentle stream of nitrogen at 40 °C until it was nearly dry, and then it was diluted to 1 mL with methanol-water mixture for further analysis.

2.3. FT-ICR MS analysis

Ultrahigh-resolution FT-ICR MS (Solari X 15.0 T, Bruker, Germany) coupled with an electrospray ionization (ESI) source was used for DOM determination in negative ion mode. Samples were injected directly into the ion source at a flow rate of 2.0 μL/min (Hu et al., 2022; Kellerman et al. 2014). The ESI capillary voltage, dry gas flow rate, and dry temperature were maintained at 3500 V, 4 L/min, and 200 °C, respectively. The ion transfer time of flight was set to 0.8 ms. Data acquisition covered a mass range of 150–1000 Da, with 100 mass spectra averaged per sample. The extraction and instrumental analysis methods for the blank sample were the same as those used for the Yangtze River Basin water samples. The obtained mass spectrum was initially calibrated using a sodium formate solution for external calibration (Xu et al., 2020), and subsequently internally recalibrated using an in-house reference mass list (**Table S2)** in the quadratic mode.

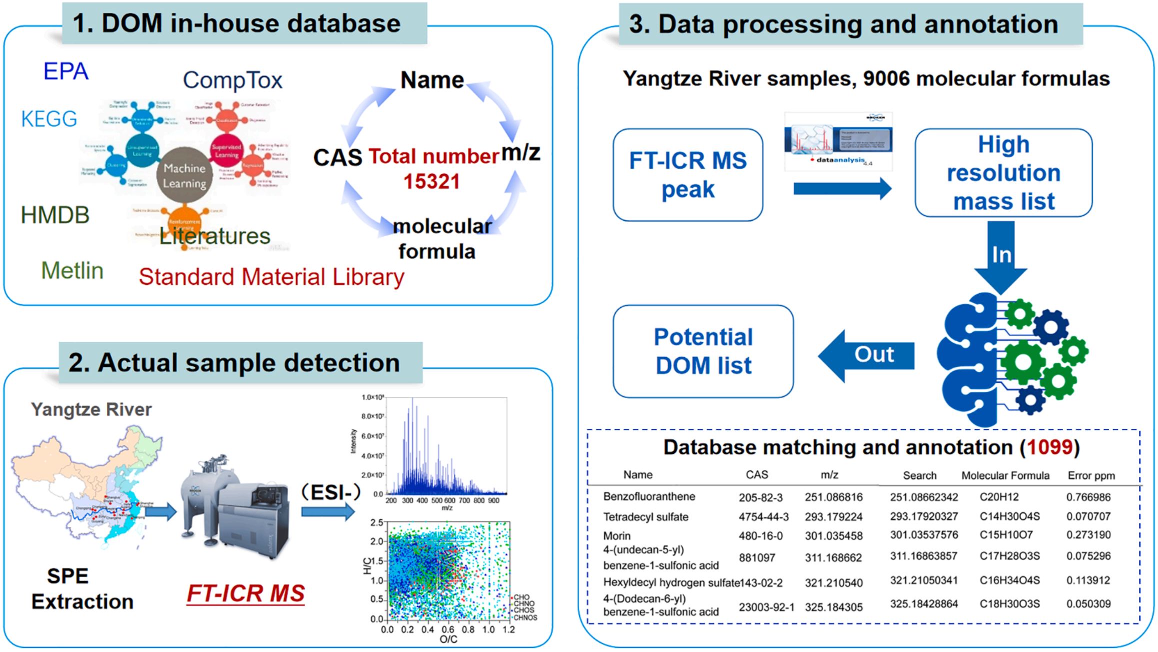
2.4. Molecular formula assignment

Mass spectrum peaks with a signal-to-noise ratio of 5 or higher were identified using a data processing system DataAnalysis (5.0, Bruker, Germany). DOM mainly contained C, H, O, N, and S elements, and the number of each element in the molecular formula was set to ^{12}C (1–50), ^{1}H (1–100), ^{16}O (1–25), ^{14}N (0–4), and ^{32}S (0–2). Additional constraints included $O/C \leq 1.2$ and $H/C \leq 2.5$ for formula calculation. The permissible mass error for the assigned molecular formula was within ±1 ppm. Finally, the identified molecular formulas were exported in Excel format. For the given chemical formula $C_cH_hN_nO_oS_s$, the unsaturation or ring structure of DOM was described by the double bond equivalent (DBE) value (Lv et al., 2016). Additionally, the aromaticity index (AI) was frequently used to evaluate the aromatic nature of DOM (Koch and Dittmar., 2006). Furthermore, Kendrick mass defect (KMD) analysis was employed to identify homologous series that contained similar repeat units such as CH_2 and COO (Kim et al., 2003). The details of DBE, AI, and KMD calculations are provided in SI.

3. Results and discussion

3.1. Workflow for DOM screening and identification

The present study developed a new workflow for DOM screening and identification (Fig. 1), which consists of three main steps: (1) the establishment of a DOM in-house database for natural water; (2) the construction of a high-resolution mass spectrometry database for DOM using actual samples; (3) data processing and annotation. The establishment of the in-house database enables high-throughput screening of DOM by integrating extensive databases of both endogenous and exogenous organic compounds. Exogenous organics were obtained from the United States Environmental Protection Agency (U.S. EPA) CompTox Chemicals Dashboard, relevant literature, and standard material libraries for water-related organic compounds. Similarly, a list of



Name	CAS	m/z	Search	Molecular Formula	Error ppm
Benzofluoranthene	205-82-3	251.086816	251.08662342	C20H12	0.766986
Tetradecyl sulfate	4754-44-3	293.179224	293.17920327	C14H30O4S	0.070707
Morin	480-16-0	301.035458	301.03537576	C15H10O7	0.273190
4-(undecan-5-yl) benzene-1-sulfonic acid	881097	311.168662	311.16863857	C17H28O3S	0.075296
Hexyldecyl hydrogen sulfate	143-02-2	321.210540	321.21050341	C16H34O4S	0.113912
4-(Dodecan-6-yl) benzene-1-sulfonic acid	23003-92-1	325.184305	325.18428864	C18H30O3S	0.050309

Fig. 1. Workflow for screening DOM based on non-targeted FT-ICR MS analysis coupled with in-house database guidance.

endogenous organics was compiled from metabolite databases such as the Human Metabolome Database (HMDB), the Metabolite Link (Metlin), and the Kyoto Encyclopedia of Genes and Genomes (KEGG). Compounds relevant to microorganisms and human metabolites, including organic acids and their derivatives, nucleotides and polysaccharides were included as endogenous organics to establish our in-house database. After the removal of duplicates, the final in-house database comprises over 15,000 organic compounds, with 38.2 % being exogenous organics and 61.8 % being endogenous organics. As shown in SI, the in-house database contains basic information of organic compounds, such as names, chemical abstracts service registry numbers (CASRN), and molecular formulas (**Fig. S1**). The extracted DOM from water samples was analyzed via FT-ICR MS. The MS features generated in the FT-ICR MS raw data were first deconvoluted, extracted, and assigned to the accurate mass and molecular composition of the DOM using the DataAnalysis data processing system. DOM screening was conducted through a comparison of its accurate mass data obtained from FT-ICR MS with entries in the in-house database, using a mass tolerance of ±1 ppm. The identified DOM was then reported with its name, m/z value, CASRN, molecular formula, and molecular mass error.

Previous studies have explored DOM in the Yangtze River Basin using FT-ICR MS; however, the data obtained only provided a molecular list of DOM (Li et al., 2023; Pang et al., 2021). The workflow developed in the present study addresses the qualitative challenges posed by the large number of DOM components and the lack of available reference standards, enabling the molecular identification of DOM. Water samples from the Yangtze River Basin were analyzed using this workflow. A total of 9006 unambiguous molecular formulas were obtained using DataAnalysis software, and the formulas were matched with a customized in-house database using the self-built software (MRMSaid, version 2023SR0404677). First, the in-house database was imported into MRMSaid in CSV format. Through the “Degradation product screening module”, the molecular formula was split to obtain the accurate molecular masses. In addition, the precise masses of 14 types of adduct ions (such as $[M-H]^{-}$ and $[M+Cl]^{-}$) in negative ion mode was acquired. Then, these values were matched with the measured data (9006 unambiguous molecular formulas), allowing a mass error of ±1 ppm. All matched DOM molecular formulas were filtered according to their specific masses, and any redundant DOM molecular formulas was removed. Ultimately, 1099 suspected DOM compounds were tentatively identified through the established workflow.

3.2. *General characterization of DOM molecular composition*

Non-targeted analysis of DOM was conducted on water samples from the Yangtze River Basin. **Fig. S2** shows the reconstructed mass spectra of DOM obtained through FT-ICR MS. A total of 9006 molecular formulas were assigned to DOM in the water samples collected from various sites along the Yangtze River Basin. The number of distinct molecular formulas of DOM ranged from 1291 to 2242 across all sampled sites (Fig. 2**a**). The assigned molecular weights (MWs) of DOM ranged from 200 to 800 Da, with average MW values ranging from 370 Da to 444 Da (Fig. 2**b**). These values correspond to the mass distribution observed in DOM from salt lakes and non-pristine lakes (Xu et al., 2020; Zhou et al., 2022). The chemical diversity of DOM exhibited both commonalities and uniqueness. Among the assigned formulas, 461 molecular formulas were consistently detected in over 50 % of the samples with the characteristic of H/C>1 and O/C<0.6, whereas 8545 unique molecular formulas were found in less than 50 % of the samples (Fig. 2**c**). The nitrogen- and sulfur- containing compounds were more abundant in unique compounds than in common compounds. This high proportion of heteroatoms was mainly due to the different sources of the samples, which led to the heterogeneity of DOM.

The assigned molecular formulas were categorized into five subgroups according to their elemental composition: CHO, CHNO, CHOS, CHNOS, and remainder compounds. Among these, CHO compounds (51.0 %–75.6 %) were most abundant in the water bodies of the Yangtze River Basin (Fig. 2**d**), followed by S-containing compounds (CHOS and CHNOS) (13.4 %–43.2 %). In contrast, the relative abundance of CHNO compounds (5.8 %–16.9 %) in DOM from the Yangtze River Basin was relatively low (Fig. 2**d**). Overall, the relative abundance of these DOM subgroups in the Yangtze River Basin differed considerably from that observed in DOM from salt lakes (CHNO > CHO > CHNOS > CHOS), highlighting distinct characteristics of DOM in different aquatic environments (Xu et al., 2020).

The molecular formula $C_cH_hO_oN_nS_s$ was classified into eight types of compounds according to the H/C ratio, O/C ratio, number of nitrogen atoms, and AI value. These types include condensed aromatics (CAS),

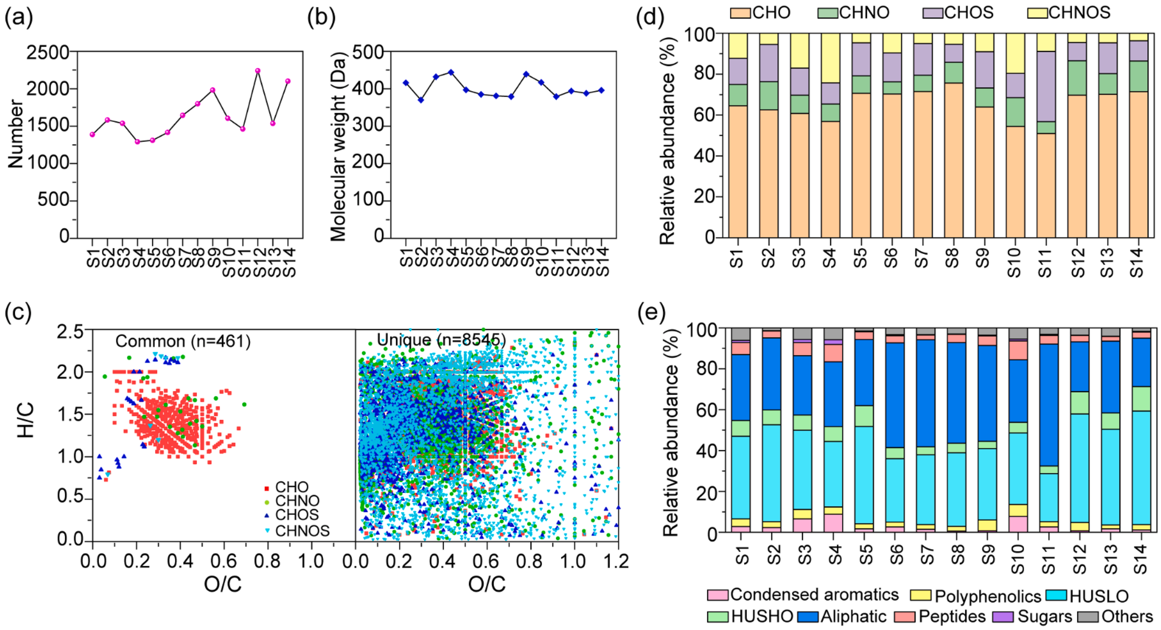


Fig. 2. (a) Number of assigned molecular formulas. (b) Average molecular weight of DOM across all samples. (c) Van Krevelen diagrams showing common compounds (detected in >50 % of the samples) and unique compounds (present in <50 % of the samples). Molecular characterization of (d) five subgroups (CHO, CHNO, CHOS, CHNOS, and remainder) and (e) eight classes (condensed aromatics, polyphenolics, HUSLO, HUSHO, aliphatic, peptides, and sugars) in DOM from all samples. (Samples S1–S6, S7–S10, and S11–S14 were collected from upstream, midstream, and downstream locations of the Yangtze River Basin, respectively).

highly unsaturated structures with low oxygen (HUSLO), highly unsaturated structures with high oxygen (HUSHO), polyphenolics, aliphatics, peptides, and sugars (He et al., 2016). Detailed classification criteria can be found in **Table S3**. HUSLO and aliphatic compounds were the predominant species across all samples, constituting 23.6 %–55.5 % (average: 39.7 %) and 23.7 %–59.7 % (average: 38.1 %) of the total DOM in the Yangtze River Basin (Fig. 2**e**). HUSLO compounds are mainly derived from lignin metabolism, while aliphatic compounds typically originate from phytoplankton fatty acids (Behnke et al., 2021).

The total numbers of molecular formulas identified in the upstream, midstream, and downstream regions of the Yangtze River Basin were 4829, 4227, and 3897, respectively (Fig. 3**a**). The percentage of CHO compounds showed little variation along the Yangtze River Basin (**Fig. S3**). Most of these CHO compounds in the unique molecular formulas had high oxygen numbers (**Fig. S4**), suggesting the potential existence of diverse oxygen-containing functional groups (eg hydroxyl, carboxyl, peroxy, ester groups) (Leyva et al., 2020; Wu et al., 2019; Zhao et al., 2013). However, CHNO and CHOS compounds increased, while

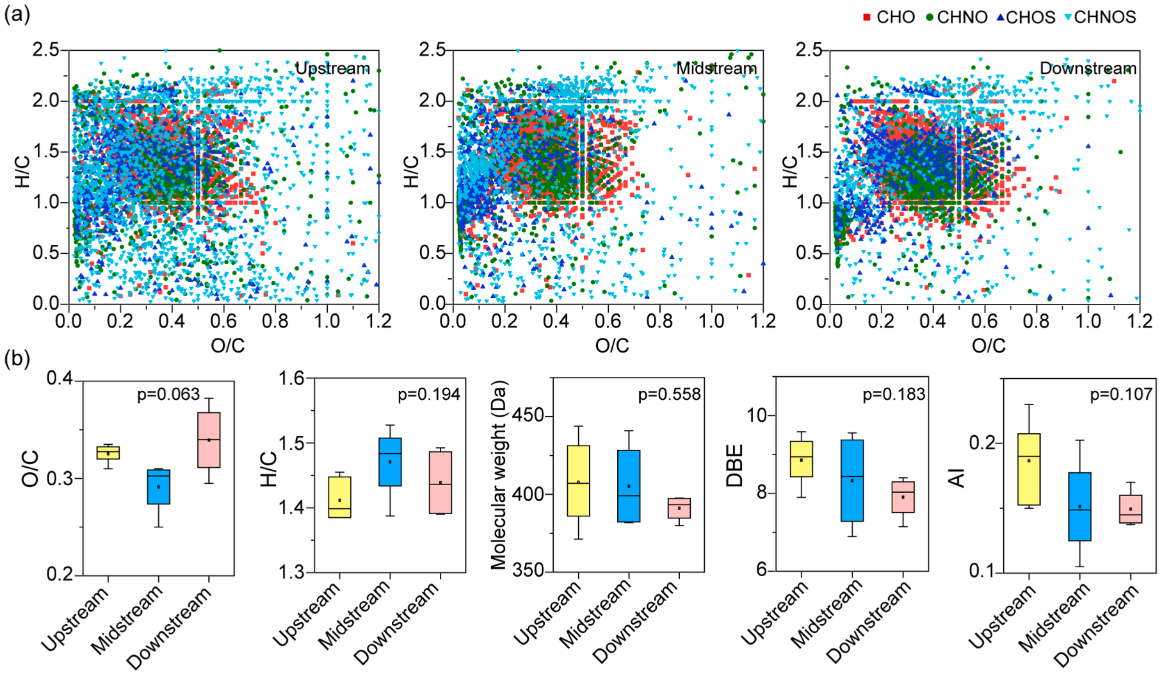


Fig. 3. (a) Van Krevelen diagrams and (b) box plots showing the weighted average O/C ratio, H/C ratio, molecular weight, DBE value, and AI value of DOM in the upstream, midstream, and downstream Yangtze River Basin. The boxes represent the interquartile range (IQR, 25 %–75 %), while the whiskers extend to 1.5 times the IQR. The black squares represent mean values.

CHNOS compounds decreased from upstream to downstream in the Yangtze River Basin (**Fig. S3**). CHNO compounds predominantly occupied the upper left part of the KMD plot (O = 4–12, KMD (COO) = -0.40 to 0.00) (**Fig. S5**), indicating their relatively unsaturated nature with a higher proportion of oxidized carbon (Xu et al., 2020). High abundance of CHNO compounds, such as $C_{17}H_{25}NO_4$, $C_{17}H_{25}NO_5$, $C_{21}H_{27}NO_6$, $C_{16}H_{27}NO_7$, and $C_{18}H_{27}NO_7$, with corresponding DBE values of 6, 6, 9, 4, and 6, respectively, were detected in the downstream Yangtze River Basin (**Fig. S6**). These CHNO compounds exhibit structures resembling lignins/carboxyl-rich alicyclic molecules and are considered recalcitrant molecules originating mainly from soil and higher plants (Hertkorn et al., 2006; Lambert et al., 2016; Wagner et al., 2019). The photodegradation and biodegradation of DOM were significantly affected by high turbidity and fluctuations of dissolved oxygen during periods of high discharge in the Yangtze River Basin, leading to the preferential preservation of recalcitrant components (ward et al., 2013). S-containing compounds with 15–19 carbon atoms, such as $C_{19}H_{32}O_3S$, $C_{18}H_{30}O_3S$, $C_{17}H_{28}O_3S$, $C_{16}H_{26}O_3S$, and $C_{15}H_{20}O_5S$, exhibiting DBE values of 4, 4, 4, 4, and 6, respectively, showed the highest relative abundance in the downstream Yangtze River Basin (**Fig. S7**). These CHOS compounds likely originated from linear alkylbenzene sulfonate (LAS), one of the most widely used surfactants in China (Gong et al., 2022; Gonsior et al., 2011). The observed high abundance of LAS downstream can be directly linked to substantial discharges resulting

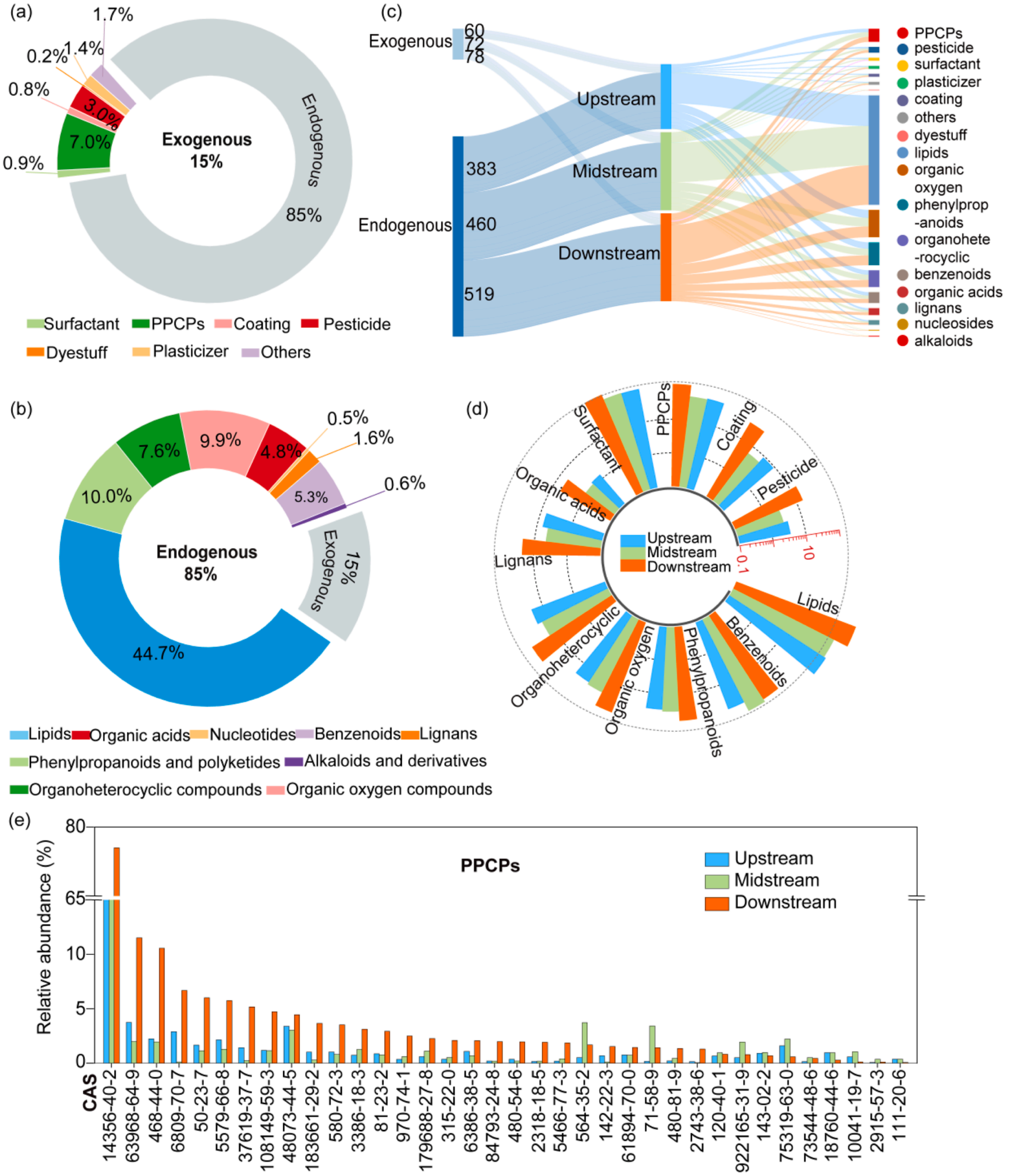


Fig. 4. (a) and (b) Composition of identified DOM in the Yangtze River Basin based on non-targeted analysis coupled with in-house database guidance. The two pie charts depict the percentage of identified DOM by number. (c) Number of identified exogenous and endogenous organics in the upstream, midstream, and downstream Yangtze River Basin. (d) Total relative abundance of commonly identified exogenous and endogenous organics in the upstream, midstream, and downstream Yangtze River Basin. (e) Relative abundance of the tentatively identified PPCPs in the Yangtze River Basin.

from high population density. Some of the CHNOS compounds, such as $C_{20}H_{24}N_4O_6S$, $C_{19}H_{26}N_4O_5S$, $C_{17}H_{10}N_2O_8S$, $C_{16}H_8N_2O_{11}S$, and $C_{32}H_{65}NO_6S$, exhibited higher relative abundance in the upstream Yangtze River Basin (**Fig. S8**). Studies have indicated that the heteroatomic formulas of DOM in the Yangtze River Basin possibly originated from terrestrial organic matter transported during periods of high discharge (Pang et al., 2021). In addition, atmospheric deposition was also an important source of CHNOS compounds in the upstream Yangtze River Basin. Previous study has reported that the wet precipitation in the upstream Yangtze River Basin contained abundant CHNOS compounds (Chen et al., 2022). It can be seen that the chemodiversity of DOM in the Yangtze River Basin was caused by both natural and anthropogenic sources.

Overall, DOM in water samples collected from the upstream Yangtze River Basin exhibited a higher O/C ratio, lower H/C ratio, and higher molecular weight, while the Kruskal-Wallis test showed no significant results (Fig. 3**b**). In contrast, DOM in the downstream Yangtze River Basin showed higher O/C and H/C ratios and lower molecular weight. Microbial degradation processes may lead to the consumption of DOM with lower oxygen content and the production of DOM with higher oxygen content, thereby contributing to the increased O/C ratio of DOM in the downstream regions (Minor et al., 2012). Generally, the DBE value and AI value of DOM decreased gradually from the upstream to downstream Yangtze River Basin (Fig. 3**b and Fig. S9**). This decrease was not statistically significant. DOM in the upstream Yangtze River Basin exhibited high unsaturation and aromaticity, indicating the abundance of terrestrial substances such as HUSLO, HUSHO, polyphenolic compounds, and CAS (Li et al., 2023). The decreased abundance of these compounds in the downstream Yangtze River Basin is attributable to a stronger photooxidation process, which selectively removes oxidized aromatic compounds (Song et al., 2019).

3.3. Identification of suspected DOM

A total of 1099 suspected DOM compounds were tentatively identified in the water bodies of the Yangtze River Basin based on non-targeted FT-ICR MS analysis coupled with in-house database guidance. According to Schymanski's confidence level criteria, the 1099 compounds were classified as "Unequivocal molecular formula" (level 4) (Schymanski et al., 2014). Fig. 4**a** and **b** illustrate the wide variety of molecules identified, with 85 % of them being endogenous organics and 15 % being exogenous organics. The majority of the DOM consisted of endogenous organics, including lipids (44.7 %), phenylpropanoids and polyketides (10.0 %), organic oxygen compounds (9.9 %), organoheterocyclic compounds (7.6 %), benzenoids (5.3 %), organic acids (4.8 %), lignans (1.6 %), alkaloids and derivatives (0.6 %), and nucleosides (0.5 %). For example, endogenous organics such as gingerol, palmitic acid, rimexolone, stearic acid, and jasmolone glucoside exhibited high relative abundance (**Table S4**). The workflow employed in this study enabled the identification of DOM in water bodies. Notably, substantial amounts of endogenous organics such as lipids, phenylpropanoids, polyketides, and organic oxygen compounds have rarely or never been individually identified in aquatic environments prior to our work.

The tentatively identified exogenous organics in the Yangtze River Basin included pharmaceuticals and personal care products (PPCPs) (7.0 %), pesticides (3.0 %), plasticizers (1.4 %), surfactants (0.9 %), coatings (0.8 %), dyestuffs (0.2 %), and others (1.7 %). PPCPs were the most frequently identified exogenous organics. PPCPs, which are emerging environmental contaminants, are ubiquitously detected in aquatic ecosystems (Bu et al., 2023) and interfere with the physiological activities of aquatic organisms owing to their strong biological activity (Osuoha et al., 2023). Additionally, considerable amounts of pesticides such as pyrethrin II, binapacryl, bentazone, coumafuryl, and acetochlor oxanilic acid were detected in the water samples from the Yangtze River Basin. Among these exogenous organics, compounds such as $C_{19}H_{32}O_3S$, $C_{15}H_{24}O_6$, $C_{18}H_{38}O_{10}$, $C_{19}H_{30}O_5$, and $C_{12}H_{26}O_4S$, which exhibited higher prevalence, were tentatively identified as tridecylbenzenesulfonic acid, tripropylene glycol diacrylate, nonatethylene glycol, piperonyl butoxide, and lauryl sulfate, respectively (**Table S4**). Non-targeted screening and identification of DOM provided an unbiased assessment of potential anthropogenic organics in the Yangtze River Basin. More attention should be focused on the environmental risks and ecological effects of PPCPs and pesticides in future research.

3.4. Spatial variation of natural and anthropogenic DOM

The identified numbers of the exogenous and endogenous organics presented an obvious increasing trend from the upstream to downstream Yangtze River Basin. Specifically, 60, 72, and 78 exogenous organics were identified in the upstream, midstream, and downstream Yangtze River Basin, respectively (Fig. 4**c**). Correspondingly, the numbers of identified endogenous organics in the upstream, midstream, and downstream Yangtze River Basin were 383, 460, and 519, respectively (Fig. 4**c**). Zhou et al. observed an increase in assigned aliphatic molecular formulas in the downstream Yangtze River Basin (Zhou et al., 2021).

Among the identified DOM, four categories of exogenous organics and seven categories of endogenous organics were detected ubiquitously in all samples. Similarly, the abundances of these commonly detected exogenous and endogenous organics were notably higher in the downstream areas compared with the upstream and midstream areas (Fig. 4**d**). Specifically, the abundances of PPCPs and surfactants increased more than 60 % and 40 % in the Yangtze River downstream compared with up and midstream areas. The identified PPCPs, such as tridecylbenzenesulfonic acid (CASRN. 14,356–40–2), artemisinin (CASRN. 63,968–64–9), gibberellin A4 (CASRN. 468–44–0), undecaethylene glycol (CASRN. 6809–70–7), and hydrocortisone (CASRN. 50–23–7), showed a notable increase in their relative abundances in the downstream Yangtze River Basin (Fig. 4**e and Table S5**). In previous studies, substantial concentrations of PPCPs have also been detected in downstream waters (Ebele et al., 2017; Veras et al., 2019), indicating the potential effect of increased anthropogenic activities such as municipal wastewater treatment plants, animal feeding operations, hospitals, and pharmaceutical manufacturers (Zhao et al., 2021).

In the present study, a large number of lipids that have rarely or never been previously identified in DOM were detected in the Yangtze River Basin. The relative abundance of lipids increased by more than 400 % in the downstream compared with up and midstream areas (Fig. 4**d**). Specifically, the relative abundances of identified lipids such as toxin T2 tetrol (CASRN. 34,114–99–3), secoeremopetasitolide A (CASRN. 178,402–70–5), deoxynivalenol (CASRN. 51,481–10–8), millefin (CASRN. 39,262–27–6), and 15-deacetylneosolaniol (CASRN. 76,348–84–0) elevated 4–6 times in the Yangtze River Basin downstream (**Fig. S10 and Table S6**). Lipids originate from autochthonous species and domestic sewage (Zhou et al., 2020; Zhou et al., 2016). The high relative abundance of lipids in the downstream Yangtze River Basin possibly originates from both natural sources and intense anthropogenic activities in the region. These natural inputs influence the metabolism of phytoplankton, plants, and microbes, resulting in the production of several refractory endogenous organics (He et al., 2022; Wen et al., 2022). The region's high degree of industrialization, dense population, and developed economy contribute to the substantial discharge of anthropogenic organic compounds into water bodies, considerably increasing the DOM content in the downstream Yangtze River Basin.

4. Conclusions

FT-ICR MS is a powerful tool for characterizing DOM at the molecular level within complex matrices, and thus, it considerably contributes to the advancement of earth and environmental sciences. Previous studies employing FT-ICR MS have identified thousands of DOM molecular formulas. However, conducting further qualitative and quantitative analysis of individual molecules remains challenging. This study innovatively developed a tailored database of DOM in water bodies by integrating extensive databases of both endogenous and exogenous organics. The database narrows the matching range of DOM in water bodies by pre-filtering out compounds irrelevant to water bodies. Furthermore, thousands of DOM molecular formulas were identified in water bodies through the matching of high-resolution DOM data with the customized database. The workflow established in this study overcomes the qualitative challenges posed by the substantial number of DOM components and the lack of available reference standards, enabling the molecular identification of DOM. For the first time, the study revealed the natural and anthropogenic distribution of identified DOM in the Yangtze River Basin. Among the 1099 tentatively identified DOM compounds in the Yangtze River Basin, 84.9 % were endogenous organics and 15.1 % were exogenous organics. Moreover, higher chemodiversity and a greater relative abundance of identified DOM were observed in the water samples from the downstream areas, especially concerning PPCPs and lipids. The results revealed that anthropogenic discharges considerably increased the number and abundance of both exogenous and endogenous organics in DOM.

CRediT authorship contribution statement

Cuiping Ning: Writing – review & editing, Methodology, Investigation, Data curation. **Shuai Sun:** Writing – review & editing. **Yuan Gao:** Resources, Investigation. **Huaijun Xie:** Resources, Investigation. **Lidong Wu:** Investigation. **Haijun Zhang:** Supervision. **Jiping Chen:** Supervision, Funding acquisition. **Ningbo Geng:** Writing – review & editing, Supervision, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 22036006, 22306083, 22306130) and Department of Education of Liaoning Province (No. JYTQN2023238).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2024.122636.

Data availability

The authors do not have permission to share data.

References

Bao, H., Wu, Y., Zhang, J., 2015. Spatial and temporal variation of dissolved organic matter in the Changjiang: fluvial transport and flux estimation. J. Geophys. Res. Biogeosciences 120 (9), 1870–1886.

Behnke, M.I., McClelland, J.W., Tank, S.E., Kellerman, A.M., Holmes, R.M., Haghipour, N., Eglinton, T.I., Raymond, P.A., Suslova, A., Zhulidov, A.V., Gurtovaya, T., Zimov, N., Zimov, S., Mutter, E.A., Amos, E., Spencer, R.G.M., 2021. Pan-Arctic riverine dissolved organic matter: synchronous molecular stability, shifting sources and subsidies. Glob. Biogeochem. Cycles 35 (4), e2020GB006871.

Bu, Q., Wang, B., Huang, J., Deng, S., Yu, G., 2023. Pharmaceuticals and personalcare products in the aquatic environment in China: a review. J. Hazard. Mater. 262, 189–211.

Canonica, S., Freiburghaus, M., 2001. Electron-rich phenols for probing the photochemical reactivity of freshwaters. Environ. Sci. Technol. 35 (4), 690–695.

Chen, S., Xie, Q.R., Su, S.H., Wu, L.B., Zhong, S.J., Zhang, Z.M., Ma, M., Qi, Y.L., Hu, W., Deng, J.J., Ren, L.J., Zhu, D.Q., Guo, Q.J., Liu, C.Q., Jang, K.S., Fu, P.Q., 2022. Source and formation process impact the chemodiversity of rainwater dissolved organic matter along the Yangtze River Basin in summer. Water Res. 211, 118024.

Cole, J.J., Prairie, Y.T., Caraco, N.F., McDowell, W.H., Tranvik, L.J., Striegl, R.G., Duarte, C.M., Kortelainen, P., Downing, J.A., Middelburg, J.J., Melack, J., 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon Budget. Ecosystems 10, 171–184.

Ebele, A.J., Abou-Elwafa Abdallah, M., Harrad, S., 2017. Pharmaceuticals and personal careproducts (PPCPs) in the freshwater aquatic environment. Emerg. Contam. 3, 1–16.

Gong, C., Jiao, R., Yan, W., Yu, Q., Li, Q., Zhang, P., Li, Y., Wang, D., 2022. Enhanced chemodiversity, distinctive molecular signature and diurnal dynamics of dissolved organic matter in streams of two headwater catchments, Southeastern China. Water Res. 211, 118052.

Gonsior, M., Zwartjes, M., Cooper, W.J., Song, W., Ishida, K.P., Tseng, L.Y., Jeung, M.K., Rosso, D., Hertkorn, N., Schmitt-Kopplin, P., 2011. Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry. Water Res. 45 (9), 2943–2953.

Guerard, J.J., Miller, P.L., Trouts, T.D., Chin, Y.P., 2009. The role of fulvic acid composition in the photosensitized degradation of aquatic contaminants. Aquat. Sci. 71 (2), 160–169.

He, D., Li, P., He, C., Wang, Y., Shi, Q., 2022. Eutrophication and watershed characteristics shape changes in dissolved organic matter chemistry along two river estuarine transects. Water Res. 214, 118196.

Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., Hedges, J.I., 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochim. Cosmochim. Acta 70 (12), 2990–3010.

Hertkorn, N., Mourad Harir, M., Kaelin, M., Cawley, K.M., Schmitt-Kopplin, P., Jaffé, R., 2016. Molecular characterization of dissolved organic matter from subtropical wetlands: a comparative study through the analysis of optical properties, NMR and FTICR/MS. Biogeosciences 13, 2257–2277.

He, W., Chen, M.L., Park, J.E., Hur, J., 2016. Molecular diversity of riverine alkaline-extractable sediment organic matter and its linkages with spectral indicators and molecular size distributions. Water Res. 100, 222–231.

Hu, A., Choi, M., Tanentzap, A.J., Liu, J.F., Jang, K.S., Lennon, J.T., Liu, Y.Q., Soininen, J., Lu, X.C., Zhang, Y.L., Shen, J., Wang, J.J., 2022. Ecological networks of dissolved organic matter and microorganisms under global change. Nat. Commun. 13, 3600.

Imai, A., Fukushima, T., Matsushige, K., Kim, Y.H., 2001. Fractionation and characterization of dissolved organic matter in a shallow eutrophic lake, its inflowing rivers, and other organic matter sources. Water Res. 35 (17), 4019–4028.

Kellerman, A.M., Dittmar, T., Kothawala, D.N., Tranvik, L.J., 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nat. Commun. 5, 3804.

Kida, M., Fujitake, N., Kojima, T., Tanabe, Y., Hayashi, K., Kudoh, S., Dittmar, T., 2021. Dissolved organic matter processing in pristine antarctic streams. Environ. Sci. Technol. 55 (14), 10175–10185.

Kida, M., Merder, J., Fujitake, N., Tanabe, Y., Hayashi, K., Kudoh, S., Dittmar, T., 2023. Determinants of microbial-derived dissolved organic matter diversity in antarctic lakes. Environ. Sci. Technol. 57 (13), 5464–5473.

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.

Koch, B.P., Dittmar, I., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 20, 926–932.

Kvitvang, H.F.N., Andreassen, T., Adam, T., Villas-Bôas, S.G., Bruheim, P., 2011. Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids. Anal. Chem. 83 (7), 2705–2711.

Lambert, T., Teodoru, C.R., Nyoni, F.C., Bouillon, S., Darchambeau, F., Massicott, P., Borges, A.V., 2016. Along-stream transport and transformation of dissolved organic matter in a large tropical river. Biogeosciences 13, 2727–2741.

Leenheer, J.A., Croue, J.P., 2003. Characterizing aquatic dissolved organic matter. Environ. Sci. Technol. 37 (1), 18A–26A.

Leyva, D., Jaffe, R., Fernandez-Lima, F., 2020. Structural characterization of dissolved organic matter at the chemical formula level using TIMS-FT-ICR MS/MS. Anal. Chem. 92 (17), 11960–11966.

Liao, Z.Y., Wang, Y., Xie, K.T., Xie, N.G., Cai, X.X., Zhou, L.H., Yuan, Y., 2022. Photochemistry of dissolved organic matter in water from the Pearl river (China): Seasonal patterns and predictive modelling. Water Res. 208, 117875.

Lipczynska-Kochany, E., 2018. Effect of climate change on humic substances and associated impacts on the quality of surface water and groundwater: A review. Sci. Total Environ. 640, 1548–1565.

Li, S.D., Meng, L.Z., Zhao, C., Gu, Y., Spencer, R.G.M., Alvarez–Salgado, X.A., Kellerman, A.M., McKenna, A.M., Huang, T., Yang, H., Huang, C.C., 2023. Spatiotemporal response of dissolved organic matter diversity to natural and anthropogenic forces along the whole mainstream of the Yangtze River. Water Res. 234, 119812.

Li, X.M., Sun, G.X., Chen, S.C., Fang, Z., Yuan, H.Y., Shi, Q., Zhu, Y.G., 2018. Molecular chemodiversity of dissolved organic matter in paddy soils. Environ. Sci. Technol. 52 (3), 963–971.

Lv, J.T., Zhang, S.Z., Wang, S.S., Lou, L., Cao, D., Christie, P., 2016. Molecular-scale investigation with ESI-FT-ICR-MS on fractionation of dissolved organic matter induced by adsorption on iron oxyhydroxides. Environ. Sci. Technol. 50 (5), 2328–2336.

Minor, E.C., Steinbring, C.J., Longnecker, K., Kujawinski, E.B., 2012. Characterization of dissolved organic matter in Lake Superior and its watershed using ultrahigh resolution mass spectrometry. Org. Geochem. 43, 1–11.

Ning, C.P., Gao, Y., Zhang, H.J., Yu, H.Y., Wang, L., Geng, N.B., Cao, R., Chen, J.P., 2019. Molecular characterization of dissolved organic matters in winter atmospheric fine particulate matters ($PM_{2.5}$) from a coastal city of northeast China. Sci. Total Environ. 689, 312–321.

Osuohaa, J.O., Anyanwuc, B.O., Ejileugha, C., 2023. Pharmaceuticals and personal care products as emerging contaminants: Need for combined treatment strategy. J. Hazard. Mater. Adv. 9, 100206.

Pang, Y., Wang, K., Sun, Y.G., Zhou, Y.P., Yang, S.Y., Lid, Y.Y., He, C., Quan Shi, Q., He, D., 2021. Linking the unique molecular complexity of dissolved organic matter to flood period in the Yangtze River mainstream. Sci. Total Environ. 764, 142803.

Qi, Y.L., Fu, P.Q., Volmer, D.A., 2022a. Analysis of natural organic matter via fourier transform ion cyclotron resonance mass spectrometry: an overview of recent non-petroleum applications. Mass Spectrom. Rev. 41 (5), 647–661.

Qi, Y.L., Xie, Q.R., Wang, J.J., He, D., Bao, H.Y., Fu, Q.L., Su, S.H., Sheng, M., Li, S.K., Volmer, D.A., Wu, F.C., Jiang, G.B., Liu, C.Q., Fu, P.Q., 2022b. Deciphering dissolved organic matter by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS): from bulk to fractions and individuals. Carbon Res. 1, 3.

Sandron, S., Rojas, A., Wilson, R., Davies, N.W., Haddad, P.R., Shellie, R.A., Nesterenko, P.N., Kelleher, B.P., Paull, B., 2015. Chromatographic methods for the isolation, separation and characterisation of dissolved organic matter. Environ. Sci. Process. Impact 17 (9), 1531–1567.

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48 (4), 2097–2098.

Seidel, M., Vemulapalli, S.P.B., Mathieu, D., Dittmar, T., 2022. Marine dissolved organic matter shares thousands of molecular formulae yet differs structurally across major water masses. Environ. Sci. Technol. 56 (6), 3758–3769.

Song, K., Shang, Y., Wen, Z., Jacinthe, P.A., Liu, G., Lyu, L., Fang, C., 2019. Characterization of CDOM in saline and freshwater lakes across China using spectroscopic analysis. Water Res. 150, 403–417.

Veras, T.B., Luiz Ribeiro de Paiva, A., Duarte, M.M.M.B., Napoleão, D.C., da Silva Pereira Cabral, J.J., 2019. Analysis of the presence of anti-inflflammatories drugs in surface water: a case study in Beberibe river - PE, Brazil. Chemosphere 222, 961–969.

Ward, N.D., Keil, R.G., Medeiros, P.M., Brito, D.C., Cunha, A.C., Dittmar, T., Yager, P.L., Krusche, A.V., Richey, J.E., 2013. Degradation of terrestrially derived macromolecules in the Amazon River. Nat. Geosci. 6, 530–533.

Wagner, S., Fair, J.H., Matt, S., Hosen, J.D., Raymond, P., Saiers, J., Shanley, J.B., Dittmar, T., Stubbins, A., 2019. Molecular hysteresis: hydrologically driven changes in riverine dissolved organic matter chemistry during a storm event. J. Geophys. Res. Biogeosciences 124 (4), 759–774.

Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Farm, M.S., Fujii, R., Mopper, K., 2003. Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. Environ. Sci. Technol. 37 (20), 4702–4708.

Wen, Z.D., Shang, Y.X., Song, K.S., Liu, G., Hou, J.B., Lyu, L.L., Tao, H., Li, S.J., He, C., Shi, Q., He, D., 2022. Composition of dissolved organic matter (DOM) in lakes responds to the trophic state and phytoplankton community succession. Water Res. 224, 119073.

Wu, C.H., Yang, J., Fu, Q., Zhu, B., Ruan, T., Jiang, G.B., 2019. Molecular characterization of water-soluble organic compounds in $PM_{2.5}$ using ultrahigh resolution mass spectrometry. Sci. Total Environ. 668, 917–924.

Xu, W., Gao, Q., He, C., Shi, Q., Hou, Z.Q., Zhao, H.Z., 2020. Using ESI FT-ICR MS to characterize dissolved organic matter in salt lakes with different salinity. Environ. Sci. Technol. 54 (20), 12929–12937.

Yang, S.L., Zhang, J., Xu, X.J., 2007. Influence of the three gorges dam on downstream delivery of sediment and its environmental implications, yangtze river. Geophys. Res. Lett. 34 (10), L10401.

Yao, W.R., Qi, Y.L., Han, Y.F., Ge, J.F., Dong, Y.Y., Wang, J.W., Yi, Y.B., Volmer, D.A., Li, S.L., Fu, P.Q., 2024. Seasonal variation and dissolved organic matter influence on the distribution, transformation, and environmental risk of pharmaceuticals and personal care products in coastal zone: a case study of Tianjin, China. Water Res. 249, 120881.

Zeng, T., Arnold, W.A., 2013. Pesticide photolysis in prairie potholes: Probing photosensitized processes. Environ. Sci. Technol. 47 (13), 6735–6745.

Zhao, W.X., Yu, G., Blaney, L., Wang, B., 2021. Development of emission factors to estimate discharge of typical pharmaceuticals and personal care products from wastewater treatment plants. Sci. Total Environ. 769, 144556.

Zhao, Y., Hallar, A.G., Mazzoleni, L.R., 2013. Atmospheric organic matter in clouds: exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry. Atmos. Chem. Phys. 13, 12343–12362.

Zhou, Y.Q., Liu, M., Zhou, L., Jang, K.S., Xu, H., Shi, K., Zhu, G.W., Liu, M.L., Deng, J.M., Zhang, Y.L., Spencer, R.G.M., Kothawala, D.N., Jeppesen, E., Wu, F.C., 2020. Rainstorm events shift the molecular composition and export of dissolved organic matter in a large drinking water reservoir in China: high frequency buoys and field observations. Water Res. 187, 116471.

Zhou, Y.Q., Yao, X.L., Zhou, L., Zhao, Z.H., Wang, X.L., Jang, K.S., Tian, W., Zhang, Y.L., Podgorski, D.C., Spencer, R.G.M., Kothawala, D.N., Jeppesen, E., Wu, F.C., 2021. How hydrology and anthropogenic activity influence the molecular composition and export of dissolved organic matter: Observations along a large river continuum. Limnol. Oceanogr. 66, 1730–1742.

Zhou, Y.Q., Zhang, Y.L., Jeppesen, E., Murphy, K.R., Shi, K., Liu, M.L., Liu, X.H., Zhu, G. W., 2016. Inflow rate driven changes in the composition and dynamics of chromophoric dissolved organic matter in a large drinking water lake. Water Res. 100, 211–221.

Zhou, Y.Q., Zhou, L., Zhang, Y.L., Zhu, G.W., Qin, B.Q., Jang, K.S., Spencer, R.G.M., Kothawala, D.N., Jeppesen, E., Brookes, J.D., Wu, F.C., 2022. Unraveling the role of anthropogenic and natural drivers in shaping the molecular composition and biolability of dissolved organic matter in non-pristine lakes. Environ. Sci. Technol. 56 (7), 4655–4664.