

ARTICLE OPEN



Comprehensive characterization of per- and polyfluoroalkyl substances in wastewater by liquid chromatography-mass spectrometry and screening algorithms

Caiming Tang¹, Yutao Liang¹, Kai Wang¹, Jianbo Liao¹, Yanhong Zeng², Xiaojun Luo², Xianzhi Peng², Bixian Mai², Qingguo Huang³ and Hui Lin¹

Per- and polyfluoroalkyl substances (PFASs) constitute a large category of synthetic environmental pollutants, many of which remain unknown and warrant comprehensive investigation. This study comprehensively characterized PFASs in fluorinated-industrial wastewater by nontarget, quasi-target and target analyses using liquid chromatography-high-resolution mass spectrometry and data-processing algorithms. The algorithms based on characteristic in-source neutral losses and isotopologue distributions were applied to screening and identifying PFASs, while semiquantitative and quantitative analyses were utilized to determine their concentrations and distributions in the wastewater. In total, 175 formulae of PFASs, including traditional, little-known and unknown species were identified and further ascertained in terms of concentrations and distributions. The total concentrations of PFASs in the wastewater were 5.3–33.4 $\mu\text{g mL}^{-1}$, indicating serious pollution of PFASs. This study not only provides an efficient approach for screening and identification of unknown PFASs, but also presents a practicable and simple way to comprehensively depict environmental pollution signatures of PFASs.

npj Clean Water (2023)6:6; <https://doi.org/10.1038/s41545-023-00220-6>

INTRODUCTION

Per- and polyfluoroalkyl substances (PFASs) are a large class of hazardous pollutants widespread in the environment, and have been raising concerns in the last few decades^{1,2}. PFASs such as perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) have been proved to be persistent, bioaccumulative and toxic^{3,4}, and have posed serious environmental threats worldwide^{5,6}. Hitherto, over 10000 PFASs have been documented in various lists^{7–9}, but the chemical identities of many are still unclear and merely a small fraction of them have been found in the environment³. Besides known PFAS pollutants, numerous unknown PFASs are potentially present in environmental and biological matrices, such as river water, surface water, and marine mammals^{10–12}. For the global concern over PFAS pollution, it is necessary to screen, identify and comprehensively characterize PFASs in the environment, including regular, little-known, and unknown ones.

In the recent decade, nontarget and quasi-target analyses have been applied to screening and identification of environmental pollutants such as current-use chemicals¹³ and halogenated organic pollutants^{14,15}, which can mitigate the dilemma caused by the lack of reference standards^{16,17}. By virtue of recent advances in chromatography and high-resolution mass spectrometry (HRMS), numerous nontarget and suspect screening analysis methods for environmental pollutants have been rapidly developed^{18–22}. The cutting-edge analytical techniques applied in nontarget and suspect screening analyses mainly include liquid chromatography coupled with quadrupole-Orbitrap HRMS (LC-Q-Orbitrap-HRMS)^{23–27}, gas chromatography coupled with Q-Orbitrap-HRMS (GC-Q-Orbitrap-HRMS)^{28,29}, LC coupled with quadrupole time-of-flight MS (LC-QTOF-MS)^{30–32}, GC-QTOF-MS^{33,34},

Fourier transform ion cyclotron resonance HRMS (FT-ICR-HRMS)^{35,36}, etc. Nontarget and suspect analyses have also been applied in the screening and identification of PFASs, and more than 1000 previously unrecognized PFASs have been identified in various matrices recently³⁷, e.g., commercial products^{1,38,39}, environmental matrices^{40–42} and biological samples^{2,43,44}. These analytical methods mainly applied LC-Q-Orbitrap-HRMS^{37,45–47} and LC-QTOF-MS^{39,40,48}, with the detection mode of full scan^{37,49}, data-dependent acquisition (DDA)⁴⁹ and data-independent acquisition (DIA)^{9,39,46}. Characteristic fragment ions, e.g., C_2F_5^- , C_3F_5^- , SO_3F^- , PO_3^- , HSO_4^- and NSO_2^- were utilized to trace quasi-molecular ions, thus identifying PFASs^{37,45}. Particularly, in the study of Liu et al.⁴⁵, an innovative nontarget analysis strategy using in-source fragmentation flagging was developed and successfully applied, demonstrating that diagnostic fragment ions generated during in-source fragmentation can be employed to flag and identify PFASs. It thus can be extrapolated that other in-source fragmentation features of PFASs such as neutral losses, which have been observed previously⁴⁵, could facilitate nontarget identification of these substances. In addition, in-source neutral losses may be more compound-specific than neutral losses occurring in collision-induced dissociation (CID), since fragmentation is generally more difficult to happen in electrospray ionization (ESI) source than in CID cell. However, so far, no study has applied in-source neutral losses to trace and screen quasi-molecular ions of PFASs for identifying these compounds.

The LC-HRMS techniques used in nontarget analysis such as LC-Q-Orbitrap-HRMS and LC-QTOF-MS are full-information scanning techniques, which usually generate large-volume dataset for each sample, resulting in laborious and cumbersome data-processing workload^{50–53}. As a result, in nontarget analysis using these

¹Research Center for Eco-Environmental Engineering, Dongguan University of Technology, 523808 Dongguan, China. ²State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, 510640 Guangzhou, China. ³College of Agricultural and Environmental Sciences, Department of Crop and Soil Sciences, University of Georgia, Griffin, GA 30223, USA. ✉email: CaimingTang@dgut.edu.cn; linhui@dgut.edu.cn

techniques, screening and identifying compounds of concern is analogous to “finding needles in a haystack”⁵⁴, and thus requires aid of computer programs in data-processing procedures, e.g., data mining and analysis^{24,25,52,55}. However, few scripting approaches have been developed and used for nontarget analysis of PFASs in environmental matrices using LC-HRMS so far⁴⁰. As in-source neutral losses of PFASs can occur in ESI-MS and lead to constant mass discrepancies between precursor and fragment ions, this property could be readily utilized in algorithms for searching and identifying PFAS features in LC-HRMS data. In addition, specific carbon and sulfur isotopologue distributions can also be used in algorithms for screening and identifying PFASs, as some previous studies have already used isotopic patterns for aiding in nontarget identification of these compounds^{39,40,45}.

Benefiting from the rapid development in nontarget analysis, comprehensive characterization (i.e., finding and profiling all relevant compounds) of chemical components of certain groups of compounds has been realized recently^{56–58}, which can effectively depict pollution signatures of the pollutants from an overall perspective. To date, most of the reported studies relevant to comprehensive characterization of environmental pollutants focused on limited groups of compounds, such as hydrocarbons⁵⁹, halogenated dioxins⁶⁰, and atmospheric brown carbon components⁶¹. However, at present, studies focusing on comprehensive characterization of PFAS pollutants in environmental matrices or other media are still scarce⁶².

Hence, in this work, we systematically conducted nontarget, quasi-target and target analyses of PFASs in fluorinated-industrial wastewater, and further comprehensively characterized their chemical components and distributions based on results of identification, quantification and semiquantification. The nontarget analysis, an essential part of this work, was implemented by LC-Q-Orbitrap-HRMS with the aid of in-house algorithms involving characteristic in-source neutral losses and specific carbon and sulfur isotopologue distributions of PFASs. The identification results derived from the nontarget and quasi-target analyses were confirmed by DDA and DIA mass spectra. A large number of PFAS formulae were identified and assigned with tentative or exact chemical structures, and a much larger number of congeners including isomers were found. The comprehensive characteristics of the components and distributions of PFASs in the wastewater were explored. This study presents an integral method for accurate and high-efficient identification of both known and unknown PFASs in complex environmental water samples, and the comprehensive analysis outcome provides an overview of the PFASs signature in the fluorinated-industrial wastewater.

RESULTS AND DISCUSSION

Method performances

Quantitative target analysis. In this study, quantitative target analysis was performed by LC-MS/MS in MRM mode using the native standards of 21 PFASs (Supplementary Table 1). The accuracies of all the standards in all quality control (QC) samples were in range of 85.4–114.9% with relative standard deviations (RSDs) of 0.9–14.7% (Supplementary Table 1), indicating satisfactory accuracy and precision of the target analysis method. The experiments of recovery and matrix effect were carried out with 8 extraction internal standards (EISs). As documented in Supplementary Table 2, the absolute recoveries were 43.5–139.9% with RSDs of 1.0–17.9%, and the matrix effects were 56.8–120.0% with RSDs of 0.5–21.1%. The matrix effects in TE (the total effluent wastewater) were not necessarily better than those in other samples, even though the wastewater of TE had been treated by the subsidiary wastewater treatment plant before sampling. The limits of quantification (LOQs) of the individual targeted PFASs

were 0.1 ng mL⁻¹, exhibiting sufficient sensitivity for the quantitative analysis. In the wastewater samples, 8 PFASs were quantified (Supplementary Fig. 1), with the concentrations from 0.2 ng mL⁻¹ to 23.8 µg mL⁻¹ (Supplementary Table 3). The quantified PFASs in Post-RO (the wastewater after reverse osmosis treatment) generally showed higher concentrations than in Pre-RO (the wastewater before reverse osmosis treatment), which was well consistent with the fact that the reverse osmosis treatment was actually a concentration process for solutes.

Nontarget analysis. The nontarget analysis of PFASs were conducted by LC-HRMS with the aid of the developed data-processing algorithms. The performance of the nontarget analysis method was evaluated by standard solutions and the native PFASs and labeled PFAS standards in the wastewater samples. As shown in Supplementary Table 4, in the calibration sample at 100 ng mL⁻¹, all the native standards could be filtered and identified with the nontarget analysis method. In addition, some ¹³C₁- and ¹³C₂-substituted PFASs, the molecular carbon isotopologues of the native standards, were found. Besides, some ¹³C₃- and ¹³C₄-labeled standards were screened and identified. These results demonstrate that the nontarget analysis method developed in this study was reliable and efficient.

In addition to standard solutions, the nontarget analysis method was also validated with the wastewater samples via the PFASs identified by the target analysis method using LC-MS/MS, along with their ¹³C₁- and ¹³C₂-substituted molecular isotopologues, and the spiked internal standards containing ¹³C₃ and ¹³C₄. As listed in Supplementary Table 5, 6 of the 8 PFASs detected by LC-MS/MS were screened and identified by the nontarget analysis method, along with some of their ¹³C₁- and ¹³C₂-substituted molecular isotopologues. In addition, the ¹³C₃- to ¹³C₄-labeled internal standards were screened and identified. These results further confirm the reliability of the nontarget analysis method.

The features of PFASs in the wastewater samples

Overview of total PFASs. In total, 175 PFAS formulae were identified in the wastewater (Fig. 1), among which 119, 79, and 8 formulae were identified by means of nontarget analysis, quasi-target analysis and target analysis, respectively (Fig. 1a). Twenty-five formulae were found by both nontarget and quasi-target analyses, and 6 were identified by both nontarget and target analyses. This result indicates that nontarget analysis could find most PFASs, and quasi-target analysis could be a complementary approach for identifying PFASs that were missed in nontarget analysis. In TE, Pre-RO and Post-RO, 107, 151 and 159 formulae of PFASs were found, respectively, of which 101 could be found in all the samples, and 44 were identified in both Pre-RO and Post-RO (Fig. 1b), suggesting good reproducibility and reliability of the identification approaches. In all the PFAS formulae, 120 were PFCAs and 55 were PFASs, accounting for 69% and 31% of the total, respectively, showing that PFCAs were the predominant and more diverse PFASs (Fig. 1c). As illustrated in Fig. 1d, the total concentrations of all the PFASs found in TE, Pre-RO and Post-RO were as high as 5.3, 15.9 and 33.4 µg mL⁻¹, respectively, indicating serious PFAS contamination in all the wastewater samples. The magnitude orders of both PFAS species and total concentrations were as follows: TE < Pre-RO < Post-RO (Fig. 1b, d). It was reasonable that Post-RO had more PFAS species and higher concentrations than Pre-RO because the RO process concentrated the PFASs and higher concentrations enhanced detectable rates. TE had less PFAS species and lower concentration than other samples, implying that the subsidiary wastewater treatment plant was effective to remove partial PFASs from the wastewater.

Formula characteristics. The identified PFAS formulae were mainly in the mass range of 100–600 u (Supplementary Fig. 2),

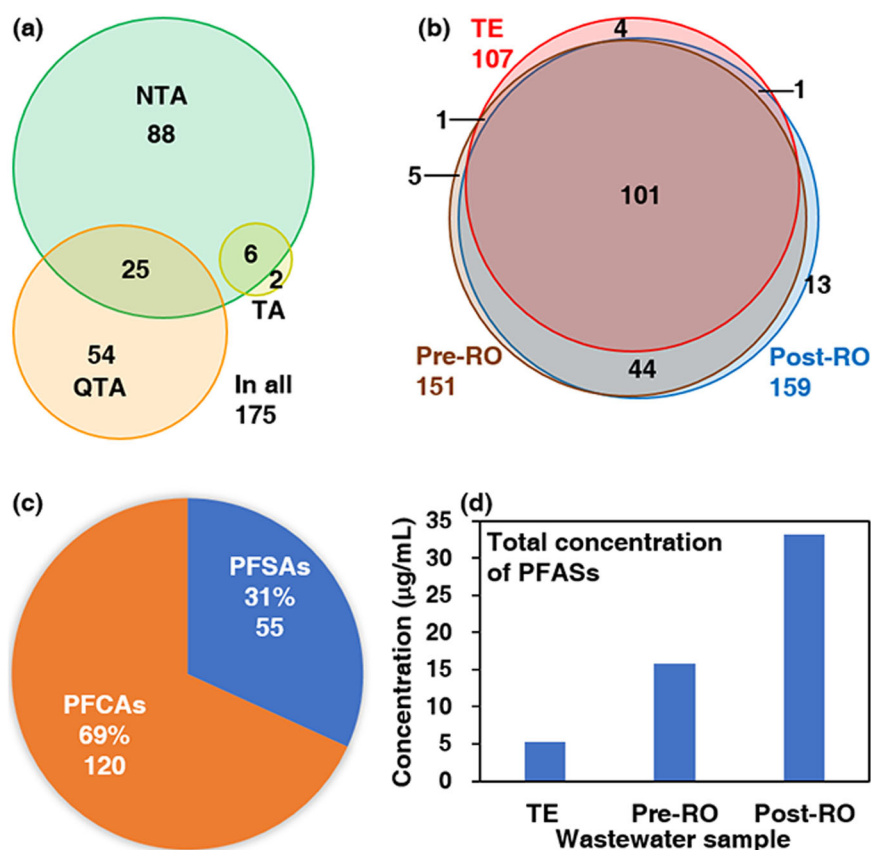


Fig. 1 Outline of the PFASs identified in the wastewater. **a** Formula numbers found by nontarget analysis (NTA), quasi-target analysis (QTA) and target analysis (TA); 88, 54 and 2 formulae were exclusively found by NTA, QTA and TA, respectively; 25 formulae were identified by both NTA and QTA; 6 formulae were found by both NTA and TA. **b** Formula numbers found in the three samples, i.e., total effluent (TE), before reverse osmosis treatment (Pre-RO), and after reverse osmosis treatment (Post-RO); 101 formulae were found in all the wastewater samples; 44 formulae were found in both Pre-RO and Post-RO; 1 formula was found in both TE and Pre-RO; 1 formula was found in both TE and Post-RO; 4, 5 and 13 formulae were exclusively found in TE, Pre-RO and Post-RO, respectively. **c** Formula numbers and proportions of the identified per- and polyfluorocarboxylic acids (PFCAs) and per- and polyfluoroalkanesulfonic acids (PFASs); 120 formulae were PFCAs, and 55 were PFASs. **d** Total concentrations of all the identified PFASs in individual wastewater samples.

indicating that the PFASs in the wastewater were mainly low to medium molecular weight compounds (≤ 600 u). The mass ranges of 150–400, 150–450, and 150–450 u had the most formula numbers in TE, Pre-RO and Post-RO, respectively. The formula numbers were generally normally distributed with the molecular weights in 100–650 u. The formula number distributions in Pre-RO and Post-RO were fairly consistent (Supplementary Fig. 2), whereas that in TE was different from others. This observation confirms that Pre-RO and Post-RO were co-genetic, and implies that the TE wastewater contained the Pre-RO/Post-RO wastewater and other wastewater with different PFAS compositions.

The graphs of CF_2 -normalized Kendrick masses against adjusted Kendrick mass defects (AKMD) of all detected ions applied for PFAS screening (Fig. 2a) and all the identified PFASs (Fig. 2b) were plotted. As illustrated in Fig. 2a, the masses of all measured ions were mainly in the range of 100–1000 u, and the AKMD were distributed in the range of 0.4–1.5. It is noteworthy that a great number of measured ions fell within the AKMD range of 0.9–1.1, possibly encompassing a large number of PFASs. Since numerous ions were within the PFAS-characteristic AKMD range (0.9–1.1), the screening and identification of PFASs were thus challenging, and the identification could not simply rely on formula assignment for quasi-molecular ions. As shown in Fig. 2b, all the normalized Kendrick masses (normalized with CF_2) of the identified PFASs were in the range from 100 to 760 u, and the AKMD fell within the range of 0.94–1.06. In addition, most of the identified PFASs were within a relatively narrow AKMD range from 0.98 to 1.04. These

results suggest that the identification outcomes were reasonable and reliable in light of the mass defect feature.

In addition to the AKMD plots, the van Krevelen diagrams of O/C vs. F/C (Fig. 2c) and O/C vs. (H + F)/C (Fig. 2d) of the identified PFASs were plotted. The van Krevelen diagrams explicitly show characteristic patterns of the PFASs, indicating different groups of PFAS species. The dots in the diagram of O/C vs. (H + F)/C (Fig. 2d) are less than those in the O/C vs. F/C diagram (Fig. 2c), indicating many identified PFASs were hydrogen-substituted species (H-PFASs). As shown in Fig. 2c, d, the compound groups of PFCAs and PFASs could be clearly recognized through the characteristic patterns of O/C vs. F/C, and O/C vs. (H + F)/C. These observations also demonstrate the high accuracy and reliability of the nontarget and quasi-target analyses in this work.

PFAS distributions vs. carbon, hydrogen, and oxygen numbers. The identified PFASs possess the carbon numbers of 2–16 (Supplementary Fig. 3). The predominant PFASs contain 2–9 carbon atoms, among which the three most abundant are those containing 3, 4, 6, and 8 carbon atoms (Supplementary Fig. 3a). C_8 -PFASs were the most abundant in all the samples, presenting the concentrations of 2.6, 12.8, and 24.4 $\mu\text{g mL}^{-1}$ in TE, Pre-RO and Post-RO, respectively. This observation is ascribable to the high concentrations of n-PFOA in the wastewater (1.7–23.8 $\mu\text{g mL}^{-1}$), accounting for 31.7–71.2% of the concentrations of Σ PFASs (i.e., all the PFASs found in individual wastewater samples), which suggests that the chemical industry park might mainly use n-

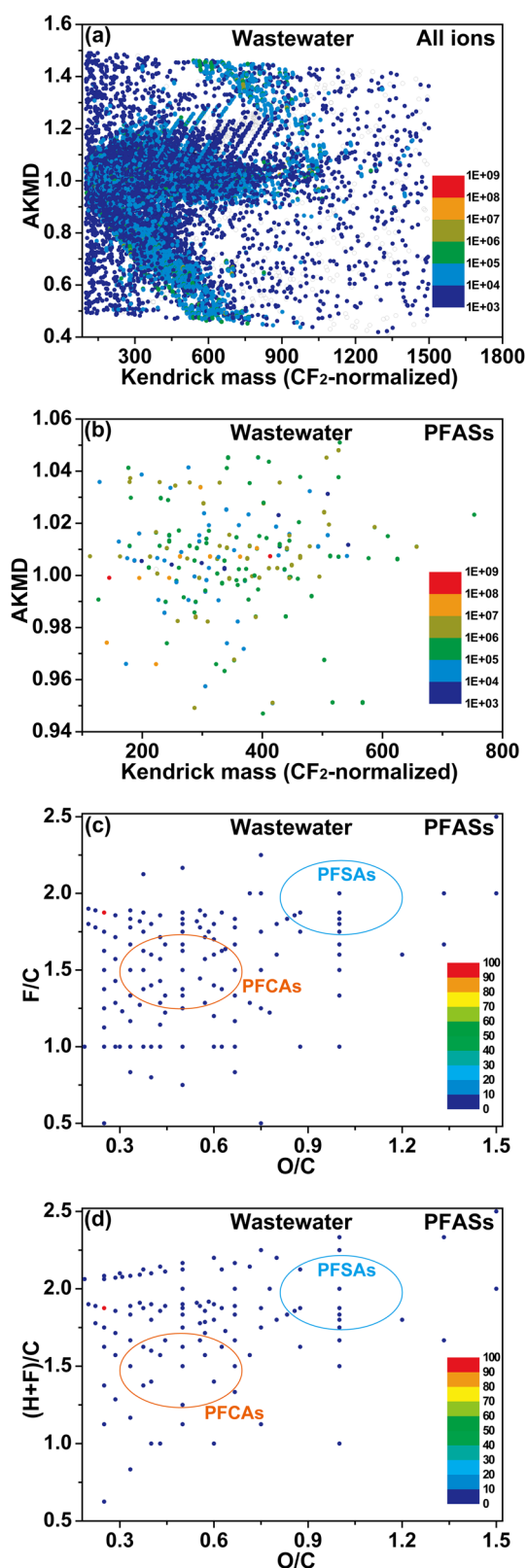


Fig. 2 Plots of CF₂-normalized Kendrick mass vs. adjusted Kendrick mass defect (AKMD) of all detected ions and the identified PFASs, and van Krevelen diagrams of the PFASs in the wastewater samples. **a** CF₂-normalized Kendrick mass vs. AKMD of all detected ions. **b** CF₂-normalized Kendrick mass vs. AKMD of the identified PFASs. **c** O/C vs. F/C of the identified PFASs. **d** O/C vs. (H+F)/C of the identified PFASs. The calculation procedures for the CF₂-normalized Kendrick masses and AKMD were referred to the study of Liu et al.⁴⁵, and are also provided in the Supplementary Information (Supplementary Equations 1–4).

mainly contain 4–9 carbon atoms, and the three PFAS groups with the most formulae are those having 6–8 carbon atoms (Supplementary Fig. 3b), with the formula numbers of 58–83, accounting for 52.2–54.2% of the total PFAS formulae. C₈-containing PFASs had the most formulae, with the numbers of 25, 33, and 32 in TE, Pre-RO and Post-RO, respectively. The PFASs with 2 and 10–16 carbon atoms presented significantly less formula numbers in contrast to others. The congener number distributions against carbon numbers were fairly consistent with the formula distributions (Supplementary Fig. 3b, c). The C₄- to C₁₀-PFASs possessed the most congeners, and C₈-PFASs had the largest congener numbers (Supplementary Fig. 3b, c), whereas the PFASs containing 2, 3, and 11–16 carbon atoms possessed significantly less congeners.

The identified PFASs had hydrogen numbers from 0 to 17 (Supplementary Fig. 4). As can be seen in Supplementary Fig. 4a, the predominant PFASs were those containing 0, 1, and 3–5 hydrogen atoms. The perfluorinated species (hydrogen number = 0) had the highest concentrations among all the PFASs, with the concentrations of 4.0, 13.7, and 26.7 $\mu\text{g mL}^{-1}$ in TE, Pre-RO and Post-RO, respectively, accounting for 76.2–86.2% of the total concentrations of PFASs. The PFASs with the hydrogen numbers of 2, 6, and 9–17 were significantly less abundant than others. Particularly in TE, merely the PFASs containing 0–11 hydrogen atoms were found. Interestingly, the PFASs containing 8, 10, 12, 14, and 16 hydrogen atoms were not found, which may relate to pathways of PFAS transformation. The PFAS abundance distributions in Pre-RO and Post-RO were relatively consistent, but different from that in TE. The formula numbers of the identified PFASs decreased as the hydrogen number increased from 0 to 17 (Supplementary Fig. 4b). The non-hydrogen-substituted PFASs encompassed the most formula numbers, possessing 53–63 species in the wastewater, accounting for 39.1–49.5% of the total formulae. It is worth noting that the formula number distributions against hydrogen numbers of the PFASs were fairly consistent in the three wastewater samples. The PFASs with 6–17 hydrogen numbers had significantly less formulae than the PFASs with other hydrogen numbers. Similar to the formula numbers, the congener numbers of PFASs generally decreased as the hydrogen numbers increased from 0 to 17 (Supplementary Fig. 4c). The PFASs containing 0–2 hydrogen atoms possessed the most congeners, in total accounting for 68.2–80.5% of the Σ PFASs.

The oxygen numbers of the identified PFASs were 2–8 (Supplementary Fig. 5). As shown in Supplementary Fig. 5a, the PFASs with two oxygen atoms were the predominant species, in other words, PFCAs containing merely two oxygen atoms were the predominant, with the concentrations of 4.7–31.2 $\mu\text{g mL}^{-1}$, accounting for 88.7–94.8% of the total concentrations of PFASs in the wastewater. The PFASs with oxygen numbers of 4 and 3 were the second and the third highest abundant, with similar concentration ranges of 184.2–370.4 and 144.3–316.0 ng mL^{-1} , respectively, accounting for 0.9–3.9% of the total concentrations of PFASs. In general, the concentrations decreased with the increasing oxygen numbers, and PFASs with 5–8 oxygen atoms had significantly lower relative abundances than those with 2–4 oxygen atoms, in all accounting for 0.1–0.7% of the total PFAS

PFOA-containing fluoride materials. On the other hand, the PFASs with relatively larger carbon numbers (10–16) showed significantly lower concentrations (0.8–20.4 ng mL^{-1}) in comparison with those with less carbon atoms. In addition, the PFASs in TE merely possess 2–10 carbon atoms. The formulae of the identified PFASs

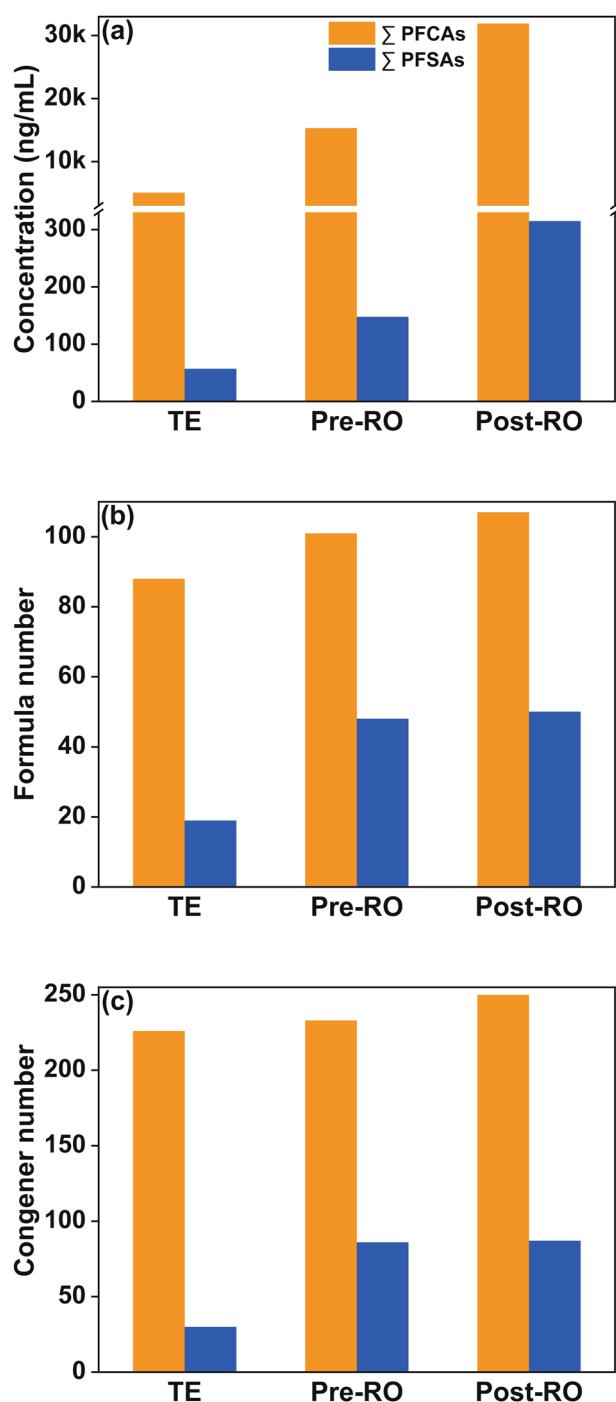


Fig. 3 Distributions of the Σ PFCAs and Σ PFASs identified in the wastewater samples. **a** Distribution of concentrations. **b** Distribution of formula numbers. **c** Distribution of congener numbers.

concentrations. With respect to formula numbers, PFASs with oxygen number of 2–4 possessed the most formulae, making up 83.6–88.8% of the total (Supplementary Fig. 5b). The formula numbers generally decreased with the increase of oxygen numbers. The congener number distributions with oxygen numbers were roughly similar to those of the formula number distributions, showing a gradual decrease with oxygen numbers from 2 to 8 (Supplementary Fig. 5c).

Distribution signatures of PFCAs and PFASs. As shown in Fig. 3a, the predominant PFASs were PFCAs, of which the concentrations were overwhelmingly higher than those of PFASs, accounting for 95.5–96.4% of the Σ PFASs in the wastewater. On the other hand, the relative abundance distributions of PFCAs and PFASs were consistent, showing the same concentration order of TE < Pre-RO < Post-RO. With respect to formula numbers, the identified PFCAs had much more formulae than PFASs (Fig. 3b), making up 66.9–82.2% of the total. The concentration orders of PFCAs and PFASs in the three samples were consistent, presenting the same order of TE < Pre-RO < Post-RO. The congener distributions between PFCAs and PFASs were similar to the formula distributions, i.e., the PFCAs had much more congeners than the PFASs, accounting for 73.0–83.3% of the total (Fig. 3c). The congener numbers of PFCAs and PFASs in the three samples shared the same order of TE < Pre-RO < Post-RO. The above results indicate that the PFASs in the wastewater were mainly PFCAs in terms of abundances, formula numbers and congener numbers, suggesting that the fluorinated materials used in the chemical industry park mainly contained PFCAs.

Isomer distributions of representative PFASs. A large number of PFAS congeners including isomers of individual formulae were found in the wastewater. Generally, several isomers were observed for individual formulae, particularly for those with relatively large carbon numbers (Supplementary Tables 6–8). For instance, as illustrated in Supplementary Fig. 6, PFHxA and PFOA presented four and five isomers in the wastewater, respectively, which were merely different in carbon skeletons, i.e., normal (n-) vs. branched carbon chains. n-PFHxA and n-PFOA showed longer retention times than their respective branched isomers on the C_{18} chromatographic column (Supplementary Fig. 6). In this work, we measured the isomeric concentrations of typical PFASs by quantitative and semiquantitative analyses (Supplementary Table 8), and the isomeric concentration distributions of some representative perfluorinated species were illustrated in Fig. 4 and Supplementary Fig. 7. In general, six representative n-PFCAs were much more abundant than their respective branched isomers. Specially, the concentrations of n-PFHxA and n-PFOA were 172.7–717.1 ng mL^{-1} (Figs. 4a) and 1.7–23.8 $\mu\text{g mL}^{-1}$ (Fig. 4b), respectively, accounting for 90.4–92.2% and 65.9–75.0% of the total concentrations of PFHxA and PFOA (Fig. 4). Interestingly, for PFHxA and PFOA, the relative abundances of their isomers gradually increased with the increasing retention times (Supplementary Fig. 6), and the abundance distributions in the three samples were similar, implying the same sources of PFHxA and PFOA in the samples (Fig. 4 and Supplementary Fig. 7). In addition, except PFPeA, all the representative PFCAs showed similar isomeric abundance distributions in the three samples, suggesting that many PFASs in the samples might be from same sources (Fig. 4 and Supplementary Fig. 7). This observation indicates that the isomeric abundance distributions of PFASs may be capable of source identification for these compounds.

Structural elucidation for PFASs

By means of the data-processing algorithms developed in this study, most F-containing carboxylic acids and sulfonic acids could be screened and identified. As a result, the identified PFASs possess at least one carboxyl/sulfonic acid group. Since the in-source neutral loss CF_2O of PFCAs and characteristic ion FSO_3^- of PFASs were taken into account in the identification process, the carbon linking to the carboxyl should bond to no less than two fluorine atoms, and that neighboring the sulfonic acid group should link with \geq one fluorine atom. Using reference standards, we unambiguously identified the structures of eight PFCAs (Supplementary Table 3). In addition, some PFASs processing sole structure

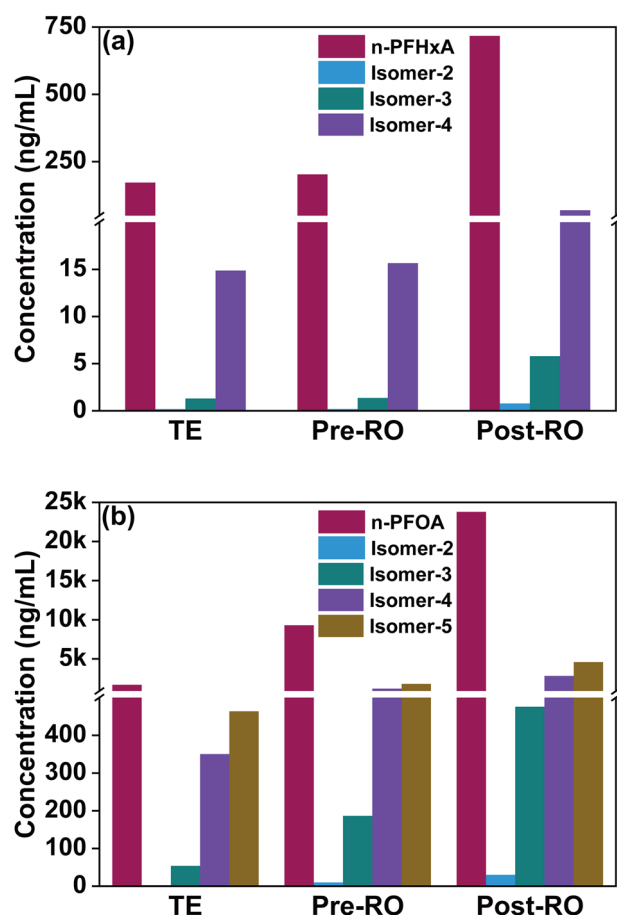


Fig. 4 Isomeric concentration distributions of two typical PFASs detected in the wastewater samples. **a** Perfluorohexanoic acid (PFHxA). **b** Perfluorooctanoic acid (PFOA).

were also unambiguously elucidated with structures, such as chlorodifluoroacetic acid ($C_2O_2ClF_2$), trifluoroacetic acid ($C_2O_2F_3$), pentafluoroethanesulfonic acid ($C_2O_3F_5S$), difluoro-hydroxy-acetyl fluoride ($C_2O_4F_3S$), 3,3,3-trifluoropropanoic acid ($C_3H_2O_2F_3$), fluoropropanate ($C_3HO_2F_4$), 3-chlorotetrafluoropropionic acid ($C_3O_2ClF_4$), perfluoropropionic acid ($C_3O_2F_5$), difluoro(trifluoromethoxy)acetic acid ($C_3O_3F_5$) and carbonic acid mono-(difluoro-trifluoromethoxymethyl) ester ($C_3O_4F_5$) (Supplementary Table 9). A number of identified PFASs could be proposed with probable structures by searching the formulae in ChemSpider and PubChem. The existence of these PFASs in the databases indicates that they might be found and/or synthesized previously. Many formulae of the screened PFASs could not be matched with rational structures or no relevant structure was present in the databases, which were proposed with tentative structures in light of their formulae, categories (PFCAs or PFASs) and mass spectra. These PFASs are likely unknown pollutants, even though their exact structures could not be elucidated. Some N-containing PFAS formulae could not be assigned with putative structures, due to high complexity and large numbers of potential structures. Nonetheless, the categories of these PFASs could be ascertained, as their carboxyl/sulfonic acid group and fluorine numbers could be specified (Supplementary Table 9).

Rarely reported and representative unknown PFASs

In this study, the quantitatively analyzed eight PFCAs and some perfluorinated short-chain PFASs (C_2 and C_3) were common species reported previously. These short-chain PFASs included trifluoroacetic acid ($C_2O_2F_3$), pentafluoroethanesulfonic acid

($C_2O_3F_5S$) and perfluoropropionic acid ($C_3O_2F_5$). However, most of the identified PFASs also contain other elements (e.g., H, Cl and N) beyond traditional perfluorinated species, or with extra oxygen atoms. These PFASs are potentially little-known or unknown fluorinated pollutants in the environment. However, it is challenging to determine their exact structures and difficult to ascertain whether they have been reported previously. In this work, if some PFAS formulae cannot be matched with rational structures via online chemical databases such as ChemSpider and PubChem or even no relevant information is available in these databases (Supplementary Table 9), they can be regarded as potentially unknown PFASs irrespective of undetermined exact chemical structures. With this recognition standard, 84 potentially unknown PFASs were identified, such as difluoro-fluorocarbonylmethanesulfonic acid ($C_2O_4F_3S$) and 2,2,3-trifluoropropanoic acid ($C_3H_2O_2F_3$), and their structures have also been tentatively proposed (Supplementary Table 9). These PFASs may be of high research significance and warrant further in-depth investigation.

Most importantly, three iodinated PFASs (I-PFASs), i.e. $C_6H_6OF_6I-SO_3H$ (including two isomers) and $C_6H_4F_8I-SO_3H$ (Supplementary Tables 7,9) were discovered in this work. To the best of our knowledge, I-PFASs have neither been reported previously, nor can be found in any available database. As shown in Supplementary Table 7, the three I-PFASs were found not only in Pre-RO, but also in Post-RO, with concentrations of 1.2–20.9 $ng\ mL^{-1}$. The total concentrations of the three I-PFASs in Pre-RO and Post-RO were 11.9 and 23.7 $ng\ mL^{-1}$, respectively, accounting for 8.2% and 7.7% of the total concentrations of all the identified PFASs in the respective samples. In particular, the isomers of $C_6H_6OF_6I-SO_3H$ were two of the major PFASs in the samples (Supplementary Table 7). These I-PFASs may find their way into the environment ultimately like some ubiquitous PFAS pollutants. As a consequence, I-PFASs may be a group of non-negligible unknown PFASs in the environment, needing further research and serious concerns.

Implications and prospect

This study presents an integrated method for comprehensive characterization of PFAS pollutants in fluorinated-industrial wastewater by combined utilization of nontarget, quasi-target and target analyses using LC-Q-Orbitrap-HRMS and LC-MS/MS. Data-processing algorithms in light of characteristic in-source neutral losses and isotopologue distributions were applied to the screening and identification of PFAS pollutants. Semiquantitative analysis based on the quantification results was employed to determine the concentrations and distributions of PFASs in the wastewater. Comprehensive characterization of PFASs in the wastewater was successfully conducted. A large number of PFAS pollutants (175 formulae and >350 congeners), including traditional, rarely known and unknown species were found and further ascertained in terms of concentrations and distributions. The total concentrations of PFASs in the samples were from 5.3 to 33.4 $\mu g\ mL^{-1}$, indicating heavy PFAS pollution in the wastewater. A number of potentially unknown PFASs were identified (84 formulae), which merit further in-depth research. The present study not only provides a highly efficient screening and identification approach for PFASs, but also presents a practicable way to comprehensively delineate pollution signatures of PFASs in the environment. The developed method involving instrumental analysis and data-processing algorithms can be extended to comprehensive analysis of PFAS pollutants in other matrices for revealing the full picture of PFASs in the environment. The identification, semiquantification and quantification outcomes provide crucial understandings about the environmental pollution caused by PFASs, especially in the aspect of rarely known and unknown PFASs. It will be worthwhile to study the rarely known and newly discovered PFASs in the aspects

of exact structure elucidation, accurate quantification, environmental occurrence and health risk assessment.

METHODS

Chemicals and materials

Three standard solutions, i.e., native perfluorinated compound (PFC) stock solution (PFAC-MXC, 21 PFASs), mass-labeled PFC EISs solution (MPFAC-C-ES, 13 ^{13}C -labeled PFASs), and mass-labeled PFC injection internal standards (IIS) solution (MPFAC-C-IS, 4 ^{13}C -labeled PFASs) were purchased from Wellington Laboratories Inc. (Ontario, Canada). Details of these standard solutions are provided in Supplementary Table 10. Methanol (MeOH) and acetonitrile (ACN) were of chromatographic grade and purchased from Merck Corp. (Darmstadt, Germany). Nylon filters (0.45 μm) were bought from Jinteng Experiment Equipment Company (Tianjin, China). Chromatographic grade ammonium acetate (NH_4Ac) was purchased from Thermo-Fisher Scientific Co., Ltd. (Hampton NH, USA). Ultrapure water (electrical resistivity: 18.2 $\text{M}\Omega\text{cm}$) was produced by a Millipore water purification apparatus (Millipore Corporation, Bellerica, MA, USA).

Calibration and quality control (QC) working solutions were prepared by serial dilution of the PFAC-MXC solution using MeOH/ H_2O (1:1, v/v), with the concentration ranges of 1–1000 ng mL^{-1} and 1–800 ng mL^{-1} , respectively. Working solutions of EIS and IIS were prepared by diluting the MPFAC-C-ES and the MPFAC-C-IS solutions to 500 ng mL^{-1} with MeOH/ H_2O (1:1, v/v), respectively. Calibration and QC samples were prepared by 10-fold dilution of the corresponding working solutions with MeOH/ H_2O (1:4, v/v), followed by addition of 10 μL of the EIS working solution and 10 μL of the IIS working solution, and the final volume of each sample was 1 mL. The concentrations of the calibration samples were 0.1, 0.5, 1, 5, 10, 50, and 100 ng mL^{-1} for all the analytes, and those of the QC samples were 0.1 (LLOQ), 0.25 (LQC), 5 (MQC), and 80 ng mL^{-1} (HQC). These solutions were stored in a freezer at $-20\text{ }^\circ\text{C}$ before use.

Sample information and pretreatment

In June 2021, three wastewater samples were collected from a chemical industry park in North China, where PFASs were used as materials in industrial production. Among the three samples, two were collected before and after a reverse osmosis (RO) treatment process, and labeled as Pre-RO and Post-RO, respectively. The third sample was collected from the total effluent (TE) of a subsidiary wastewater treatment plant of the chemical industry park, and labeled as TE. Specifically, the wastewater treatment plant involved an activated carbon adsorption process. The samples were placed in polypropylene bottles and immediately transported to our laboratory for pretreatment. A volume of 800 μL of each sample was transferred to a 2-mL glass vial, and spiked with 10 μL of the EIS working solution and vortex-mixed for 2 min. Then, 180 μL MeOH was added in the sample mixture and vortex-mixed for 2 min, followed by addition of 10 μL of the IIS working solution. Thereafter, the sample mixture was filtered with a Nylon filter, and then sealed for instrumental analysis. Specifically, for quantitative target analysis of PFASs with high concentrations in the wastewater samples using LC-MS/MS, the raw samples were subjected to 100-fold dilution with ultrapure water before the pretreatment procedures described above.

Instrumental analysis

Nontarget and quasi-target analyses. The nontarget and quasi-target analyses were conducted by an LC-Q-Orbitrap-HRMS system comprised of a Dionex ultra performance liquid chromatograph and a Q-Extractive Plus mass spectrometer equipped with a heated-electrospray ionization (HESI) source (Thermo-Fisher

Scientific, Waltham, USA). Chromatographic separation was carried out on an Acquity UPLC[®] BEH C_{18} column (2.1 \times 100 mm, 1.7 μm , Waters, Milford, MA, USA), with a protective column with the same packing material (1 mm). The mobile phase A was ultrapure water containing 2 mM NH_4Ac , and the mobile phase B was ACN. The gradient elution program was set as follows: from 0 to 0.2 min, mobile phase B kept at 20%; 0.2–8 min, B linearly increased to 80%; 8–10 min, B linearly increased to 95%; 10–12 min, B maintained at 95%; 12–12.1 min, B decreased to 20%; 12.1–15 min, B kept at 20%. The flow rate was 250 $\mu\text{L min}^{-1}$, and the column was kept at room temperature. The injection volume was 5 μL . The HESI source was operated in negative mode. Other ionization source parameters are listed as follows, spray voltage: -3200 V , sheath gas flow rate: 45 arb, auxiliary gas flow rate: 10 arb; capillary temperature: 320 $^\circ\text{C}$, auxiliary temperature: 350 $^\circ\text{C}$. Full scan, DDA, and DIA modes were applied, with the mass resolutions of 140,000, 70,000, and 70,000, respectively (at 200 u). The scan ranges of full scan mode were set at m/z 50–750 and 100–1500, and that of DIA mode was m/z 70–1000.

Target analysis. The target analysis was performed by an LC-MS/MS system consisting of a Waters Acquity H-Class UPLC and a Xevo TQD triple quadrupole tandem mass spectrometer equipped with an ESI source (Waters Corporation). The LC column used and the LC working conditions were identical to those in the LC-Q-Orbitrap-HRMS analysis as described above. The ESI source was operated in negative mode. The ESI parameters are provided as the following, capillary voltage: -2500 V , source temperature: 150 $^\circ\text{C}$, desolvation temperature: 500 $^\circ\text{C}$, desolvation gas flow rate: 800 L h^{-1} . Multiple reaction monitoring (MRM) mode was applied for data acquisition. The MRM parameters including ion transition channels, cone voltages and collision energies are documented in Supplementary Table 11.

Data-processing

Nontarget analysis. The data volume for each sample generated by LC-Q-Orbitrap-HRMS was fairly large ($\sim 120\text{ MB}$), and the data processing was thereby crucial and challenging. The chromatogram of each sample (0.5–15 min) was equally divided into 29 segments with a steplength of 0.5 min, and the full scan mass spectral data of individual segments were exported from the software Xcalibur 4.1 (Thermo-Fisher) to Excel files in form of plain text.

PFCAs. The mass spectral data were first screened with a Matlab script for total per- and polyfluorocarboxylic acids (PFCAs). The algorithm was based on the premise that each PFCA should generate three ions, namely, $[\text{M}-\text{H}]^-$, $[\text{M}-\text{CO}_2-\text{H}]^-$ and $[\text{M}-\text{CF}_2\text{O}-\text{H}]^-$, with the mass discrepancy between the first ions of $22.00189 \pm 0.001\text{ u}$ and that between the last two ions of $43.98983 \pm 0.001\text{ u}$. In addition, based on the mass spectra of the native PFCA standards, the abundance ratio between the first two ions and that between the last two were set within the ranges of 0.5–100 and 0.1–40, respectively. Therefore, all the potential PFCAs containing one $-\text{COOH}$ and at least one $-\text{CF}_2$ were filtered. Then, the exact masses of the quasi-molecular ions ($[\text{M}-\text{H}]^-$) were determined. The exact mass of each quasi-molecular ion was then checked with the formula calculation program implemented in Xcalibur for formula assignment, and the general formula was preset as $\text{C}_{2-30}\text{H}_{0-60}\text{O}_{2-10}\text{N}_{0-2}\text{P}_{0-2}\text{F}_{2-60}$. The mass error should be $\leq 3\text{ ppm}$, and the ring and double bond equivalence (RDBE) values should be with a decimal part of 0.5. In addition, formulae which had incommensurate RDBE values so that could not constitute reasonable molecules were regarded as fragment ions rather than quasi-molecular ions. The quasi-molecular ion of each PFCA candidate, along with the in-source fragment ions, namely, $[\text{M}-\text{CO}_2-\text{H}]^-$ and $[\text{M}-\text{CF}_2\text{O}-\text{H}]^-$ were checked in the full scan

chromatogram, and the extracted ion chromatograms (EICs) of these ions were obtained. Then, the chromatographic peaks (including possible isomeric peaks) of PFCA candidates were determined. Afterwards, DDA (if detected) and DIA mass spectra of the PFCA candidates were extracted, and their chemical structures were tentatively elucidated via analysis of characteristics of the mass spectra and fragmental pathways of the ions. For complex PFCA with uncommon structures difficult to elucidate, their molecular formulae were searched in online chemical databases, e.g., ChemSpider and PubChem to match potential chemical structures, thus facilitating the structure identification.

PFASs. The mass spectral data were first filtered by a Matlab script for overall potential per- and polyfluoroalkanesulfonic acids (PFASs), which based on the algorithm for searching monosulfur-containing organic compounds via their specific carbon and sulfur isotopologue distributions. For each potential PFSA, at least 3 isotopologues of the quasi-molecular ion should be detected, i.e., $[M-H]^-$, $[^{13}C_1-M-H]^-$ and $[^{34}S_1-M-H]^-$, and the mass discrepancies between $[M-H]^-$ and $[^{13}C_1-M-H]^-$, and that between $[M-H]^-$ and $[^{34}S_1-M-H]^-$ should be within 1.00335 ± 0.001 and 1.99580 ± 0.001 u, respectively. In addition, the abundance ratio of $[^{13}C_1-M-H]^-$ to $[M-H]^-$, and that of $[^{34}S_1-M-H]^-$ to $[M-H]^-$ should be within 0.02–0.3 and 0.024–0.064, respectively. Moreover, in each chromatogram segment, the in-source fragment ions SO_3^- (m/z 79.95736), HSO_4^- (m/z 96.96010) and/or FSO_3^- (m/z 98.95576) should be found. After filtering, quasi-molecular ions of potential PFSA candidates were identified, which were then sent to the formula matching program in Xcalibur for formula assignment and checking. The general formula for the assignment was preset as: $C_{2-30}H_{0-60}O_{3-10}N_{0-2}P_{0-2}SF_{2-60}$. Like PFCA, the mass error tolerance for screening PFASs was ≤ 3 ppm, and the RDBE values should contain a decimal part of 0.5. The three quasi-molecular ion isotopologues of each potential PFSA were further checked in the full scan chromatogram, and their EICs were checked to ascertain the chromatographic peak or isomeric peaks of the PFSA. Thereafter, the DDA (if detected) and DIA mass spectra of each PFSA candidate were extracted and checked for the existence of the diagnostic product ions SO_3^- and FSO_3^- , for confirming the PFSA formula. In addition, according to the MS/MS spectra obtained by DDA/DIA, chemical structures of the PFASs were tentatively elucidated.

The workflow of the data-processing procedures is outlined in Supplementary Fig. 8. All the screening scripts were written with Matlab®R2020a (Mathworks, Inc., Natick, MA, USA), and the data processing was performed by Matlab.

Quasi-target analysis. According to the structures and mass spectra of the 21 native reference standards of PFASs, we in-silico devised a large number of homologues of the reference standards by substitution with and/or introduction of H, O, Cl and Br atoms, wherein the introduced oxygen atom(s) are involved in ether, alcoholic hydroxyl and carbonyl. The exact masses of quasi-molecular ions of the devised PFASs were then checked in the full scan chromatograms with the mass error ≤ 3 ppm. If a suspect chromatographic peak of a devised PFAS was found in the chromatogram, the reasonability of retention time was assessed by comparison with that of corresponding reference standards in consideration of the introduced group with regard to polarity variations. If the retention time was rational, then the full scan, DDA (if available) and DIA mass spectra of the peak were extracted and checked for the presence of diagnostic fragment ions of the PFAS. If the mass spectra were consistent with the devised PFAS, then the compound was identified with a tentative structure.

Quantitative target analysis and semiquantitative analysis. In this work, 13 PFCA and 8 PFASs were quantitatively analyzed. The

quantification was performed with an internal standard method. Linear calibration was applied with the weight factor of $1/x$. As for the identified PFASs without reference standards, semiquantification was conducted by comparing their MS signal intensities of quasi-molecular ions with those of their structurally similar PFASs that were quantitatively analyzed, e.g., hydrogen-substituted PFOA (H-PFOA) vs. PFOA.

Quality assurance and quality control. Quality assurance and quality control were conducted to ensure accuracy and reliability. The calibration samples and QC samples were analyzed by both LC-Q-Orbitrap-MS and LC-MS/MS. The performances of the nontarget analysis were evaluated with the reference standards in the calibration and the QC samples, the ^{13}C -labeled standards spiked in the wastewater samples, and the PFASs found in the wastewater samples by the target analysis using LC-MS/MS. The accuracy and detection rates of these PFASs and reference standards were utilized to validate the reliability of the nontarget analysis. The quantitative target analysis of the 21 native PFASs was validated with the QC samples in terms of accuracy and precision at low, middle, and high concentration levels. In addition, the recoveries and matrix effects of the PFCA in the LC-MS/MS analysis were evaluated by procedures used in our previous study⁶³. Procedure blanks, and control blanks (spiked with internal standards only) were prepared and analyzed, and in these samples no analyte should be detected with signal intensity $\geq 20\%$ of that in the LLOQ samples. The carryover effects should be $\leq 20\%$ of the signal intensities of the analytes in the LLOQ samples.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

CODE AVAILABILITY

The code used to process the experimental data is available upon request to the corresponding authors.

Received: 14 May 2022; Accepted: 23 January 2023;
Published online: 09 February 2023

REFERENCES

- Barzen-Hanson, K. A. et al. Discovery of 40 classes of per- and polyfluoroalkyl substances in historical aqueous film-forming foams (AFFFs) and AFFF-impacted groundwater. *Environ. Sci. Technol.* **51**, 2047–2057 (2017).
- Spain, K. M. et al. Fluorine mass balance and suspect screening in marine mammals from the northern hemisphere. *Environ. Sci. Technol.* **54**, 4046–4058 (2020).
- Wang, Z., DeWitt, J. C., Higgins, C. P. & Cousins, I. T. A never-ending story of per- and polyfluoroalkyl substances (PFASs)? *Environ. Sci. Technol.* **51**, 2508–2518 (2017).
- Rahman, M. F., Peldszus, S. & Anderson, W. B. Behaviour and fate of perfluoroalkyl and polyfluoroalkyl substances (PFASs) in drinking water treatment: A review. *Water Res* **50**, 318–340 (2014).
- Ludwicki, J. K. et al. Hazard quotient profiles used as a risk assessment tool for PFOS and PFOA serum levels in three distinctive European populations. *Environ. Int.* **74**, 112–118 (2015).
- Lin, H. et al. Defect engineering on a Ti_4O_7 electrode by Ce^{3+} doping for the efficient electrooxidation of perfluorooctanesulfonate. *Environ. Sci. Technol.* **55**, 2597–2607 (2021).
- USEPA, United States Environmental Protection Agency (USEPA). CompTox chemistry Dashboard PFAS Master list of PFAS Substances (2020). <https://comptox.epa.gov/dashboard/chemical-lists/PFASSTRUCT> (Accessed 05 May 2022).
- Wang, Z. et al. A new OECD definition for per- and polyfluoroalkyl substances. *Environ. Sci. Technol.* **55**, 15575–15578 (2021).
- Liu, Y., D'Agostino, L. A., Qu, G., Jiang, G. & Martin, J. W. High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of

- poly-and per-fluoroalkyl substances (PFASs) in environmental and human samples. *TrAC-Trend Anal. Chem.* **121**, 115420 (2019).
10. Heydebreck, F., Tang, J., Xie, Z. & Ebinghaus, R. Alternative and legacy per-fluoroalkyl substances: differences between European and Chinese river/estuary systems. *Environ. Sci. Technol.* **49**, 8386–8395 (2015).
 11. Strynar, M. et al. Identification of novel perfluoroalkyl ether carboxylic acids (PFECAs) and sulfonic acids (PFESAs) in natural waters using accurate mass time-of-flight mass spectrometry (TOFMS). *Environ. Sci. Technol.* **49**, 11622–11630 (2015).
 12. Gebbink, W. A. et al. Observation of emerging per- and polyfluoroalkyl substances (PFASs) in Greenland marine mammals. *Chemosphere* **144**, 2384–2391 (2016).
 13. Samanipour, S., Martin, J. W., Lamoree, M. H., Reid, M. J. & Thomas, K. V. Letter to the editor: Optimism for nontarget analysis in environmental chemistry. *Environ. Sci. Technol.* **53**, 5529–5530 (2019).
 14. Tang, C. & Tan, J. Quasi-targeted analysis of halogenated organic pollutants in fly ash, soil, ambient air and flue gas using gas chromatography-high resolution mass spectrometry with isotopologue distribution comparison and predicted retention time alignment. *J. Chromatogr. A* **1555**, 74–88 (2018).
 15. Tang, C. et al. Quantitative and semiquantitative analyses of hexa-mix-chlorinated/brominated benzenes in fly ash, soil and air using gas chromatography-high resolution mass spectrometry assisted with isotopologue distribution computation. *Environ. Pollut.* **255**, 1131629 (2019).
 16. Moschet, C., Piazzoli, A., Singer, H. & Hollender, J. Alleviating the reference standard dilemma using a systematic exact mass suspect screening approach with liquid chromatography-high resolution mass spectrometry. *Anal. Chem.* **85**, 10312–10320 (2013).
 17. Schymanski, E. L. et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
 18. Goto, A. et al. Nontarget and target screening of organohalogen compounds in mussels and sediment from Hiroshima Bay, Japan: occurrence of novel bioaccumulative substances. *Environ. Sci. Technol.* **54**, 5480–5488 (2020).
 19. Tue, N. M. et al. Nontarget screening of organohalogen compounds in the liver of wild birds from Osaka, Japan: specific accumulation of highly chlorinated POP homologues in raptors. *Environ. Sci. Technol.* **55**, 8691–8699 (2021).
 20. Peng, H. et al. Untargeted identification of organo-bromine compounds in lake sediments by ultrahigh-resolution mass spectrometry with the data-independent precursor isolation and characteristic fragment method. *Anal. Chem.* **87**, 10237–10246 (2015).
 21. Wang, L., Kang, Q., Jia, Y., Li, X. & Hu, J. Identification of three novel chloroalkyl organophosphate triesters in house dust using halogenation-guided nontarget screening combined with suspect screening. *Environ. Sci. Technol.* **55**, 2482–2490 (2021).
 22. Phillips, K. A. et al. Suspect screening analysis of chemicals in consumer products. *Environ. Sci. Technol.* **52**, 3125–3135 (2018).
 23. Lu, Y. et al. Combination of high resolution mass spectrometry and a halogen extraction code to identify chlorinated disinfection byproducts formed from aromatic amino acids. *Water Res.* **190**, 116710 (2021).
 24. Lu, Y. et al. Nontargeted identification of chlorinated disinfection byproducts formed from natural organic matter using orbitrap mass spectrometry and a halogen extraction code. *J. Hazard. Mater.* **416**, 126198 (2021).
 25. Cariou, R. et al. Nontargeted LC/ESI-HRMS detection of polyhalogenated compounds in marine mammals stranded on French Atlantic coasts. *ACS EST Water* **1**, 309–318 (2021).
 26. Liu, X., Lin, Y., Ruan, T. & Jiang, G. Identification of N-nitrosamines and nitrogenous heterocyclic byproducts during chloramination of aromatic secondary amine precursors. *Environ. Sci. Technol.* **54**, 12949–12958 (2020).
 27. Postigo, C. et al. Unraveling the chemodiversity of halogenated disinfection byproducts formed during drinking water treatment using target and non-target screening tools. *J. Hazard. Mater.* **401**, 123681 (2021).
 28. Yang, L. et al. Gas chromatography-orbitrap mass spectrometry screening of organic chemicals in fly ash samples from industrial sources and implications for understanding the formation mechanisms of unintentional persistent organic pollutants. *Sci. Total Environ.* **664**, 107–115 (2019).
 29. Dubocq, F., Bæringsdóttir, B. B., Wang, T. & Kärrman, A. Comparison of extraction and clean-up methods for comprehensive screening of organic micropollutants in fish using gas chromatography coupled to high-resolution mass spectrometry. *Chemosphere* **286**, 131743 (2022).
 30. Li, J. et al. Identification of unknown disinfection byproducts in drinking water produced from Taihu Lake source water. *J. Environ. Sci.* **113**, 1–11 (2022).
 31. Ul'yanovskii, N. V., Kosyakov, D. S., Varsegov, I. S., Popov, M. S. & Lebedev, A. T. Identification of novel disinfection byproducts in pool water: chlorination of the algacide benzalkonium chloride. *Chemosphere* **239**, 124801 (2020).
 32. Kosyakov, D. S., Ul'yanovskii, N. V., Popov, M. S., Latkin, T. B. & Lebedev, A. T. Characterization of disinfection by-products in Arkhangelsk tap water by liquid chromatography/high-resolution mass spectrometry. *J. Anal. Chem.* **73**, 1260–1268 (2018).
 33. Baygi, S. F., Fernando, S., Hopke, P. K., Holsen, T. M. & Crimmins, B. S. Automated isotopic profile deconvolution for high resolution mass spectrometric data (APGC-QToF) from biological matrices. *Anal. Chem.* **91**, 15509–15517 (2019).
 34. Hernández, F., Portolés, T., Pitarch, E. & López, F. J. Target and nontarget screening of organic micropollutants in water by solid-phase microextraction combined with gas chromatography/high-resolution time-of-flight mass spectrometry. *Anal. Chem.* **79**, 9494–9504 (2007).
 35. Fu, Q. L., Fujii, M. & Kwon, E. Development and application of a high-precision algorithm for nontarget identification of organohalogenes based on ultrahigh-resolution mass spectrometry. *Anal. Chem.* **92**, 13989–13996 (2020).
 36. Luek, J. L. et al. Halogenated organic compounds identified in hydraulic fracturing wastewaters using ultrahigh resolution mass spectrometry. *Environ. Sci. Technol.* **51**, 5377–5385 (2017).
 37. Barrett, H. et al. Suspect and nontarget screening revealed class-specific temporal trends (2000–2017) of poly- and perfluoroalkyl substances in St. Lawrence Beluga whales. *Environ. Sci. Technol.* **55**, 1659–1671 (2021).
 38. D'Agostino, L. A. & Mabury, S. A. Identification of novel fluorinated surfactants in aqueous film forming foams and commercial surfactant concentrates. *Environ. Sci. Technol.* **48**, 121–129 (2014).
 39. Dubocq, F., Wang, T., Yeung, L., Sjöberg, V. & Kärrman, A. Characterization of the chemical contents of fluorinated and fluorine-free firefighting foams using a novel workflow combining nontarget screening and total fluorine analysis. *Environ. Sci. Technol.* **54**, 245–254 (2020).
 40. Wang, Y. et al. Suspect and nontarget screening of per- and polyfluoroalkyl substances in wastewater from a fluorochemical manufacturing park. *Environ. Sci. Technol.* **52**, 11007–11016 (2018).
 41. Yu, N. et al. Non-target and suspect screening of per- and polyfluoroalkyl substances in airborne particulate matter in China. *Environ. Sci. Technol.* **52**, 8205–8214 (2018).
 42. Getzinger, G. J., Higgins, C. P. & Ferguson, P. L. Structure database and in silico spectral library for comprehensive suspect screening of per- and polyfluoroalkyl substances (PFASs) in environmental media by high-resolution mass spectrometry. *Anal. Chem.* **93**, 2820–2827 (2021).
 43. Li, Y. et al. Transplacental transfer of per-and polyfluoroalkyl substances identified in paired maternal and cord sera using suspect and nontarget screening. *Environ. Sci. Technol.* **54**, 3407–3416 (2020).
 44. Wang, Q. et al. Target, nontarget, and suspect screening and temporal trends of per- and polyfluoroalkyl substances in marine mammals from the South China Sea. *Environ. Sci. Technol.* **55**, 1045–1056 (2021).
 45. Liu, Y., Pereira, A. D. S. & Martin, J. W. Discovery of C5–C17 poly and per-fluoroalkyl substances in water by in-line SPE-HPLC Orbitrap with in-source fragmentation flagging. *Anal. Chem.* **87**, 4260–4268 (2015).
 46. Hensema, T. J., Berendsen, B. J. & van Leeuwen, S. P. Non-targeted identification of per-and polyfluoroalkyl substances at trace level in surface water using fragment ion flagging. *Chemosphere* **265**, 128599 (2021).
 47. Liu, Y., Qian, M., Ma, X., Zhu, L. & Martin, J. W. Nontarget mass spectrometry reveals new perfluoroalkyl substances in fish from the Yangtze River and Tangxun Lake, China. *Environ. Sci. Technol.* **52**, 5830–5840 (2018).
 48. McCord, J. & Strynar, M. Identification of per- and polyfluoroalkyl substances in the Cape Fear River by high resolution mass spectrometry and nontargeted screening. *Environ. Sci. Technol.* **53**, 4717–4727 (2019).
 49. Wang, X. et al. Non-target and suspect screening of per- and polyfluoroalkyl substances in Chinese municipal wastewater treatment plants. *Water Res.* **183**, 115989 (2020).
 50. Gupta, S., Aga, D., Pruden, A., Zhang, L. & Vikesland, P. Data analytics for environmental science and engineering research. *Environ. Sci. Technol.* **55**, 10895–10907 (2021).
 51. Samanipour, S., Reid, M. J., Baek, K. & Thomas, K. V. Combining a deconvolution and a universal library search algorithm for the nontarget analysis of data-independent acquisition mode liquid chromatography–high-resolution mass spectrometry results. *Environ. Sci. Technol.* **52**, 4694–4701 (2018).
 52. Pena-Abaurrea, M. et al. Identification of potential novel bioaccumulative and persistent chemicals in sediments from Ontario (Canada) using scripting approaches with GCxGC-TOF MS analysis. *Environ. Sci. Technol.* **48**, 9591–9599 (2014).
 53. Cariou, R., Omer, E., Léon, A., Dervilly-Pinel, G. & Le Bizec, B. Screening halogenated environmental contaminants in biota based on isotopic pattern and mass defect provided by high resolution mass spectrometry profiling. *Anal. Chim. Acta* **936**, 130–138 (2016).
 54. Zahn, D. & Frömel, T. Finding a needle in a haystack—analyte-driven tools and techniques for information extraction and prioritization of chemicals from

- environmental (chromatography-) HRMS non-target screening data. *Curr. Opin. Environ. Sci. Health* **18**, 70–78 (2020).
55. Léon, A. et al. HaloSeeker 1.0: A user-friendly software to highlight halogenated chemicals in nontargeted high-resolution mass spectrometry data sets. *Anal. Chem.* **91**, 3500–3507 (2019).
 56. Dubocq, F., Kärrman, A., Gustavsson, J. & Wang, T. Comprehensive chemical characterization of indoor dust by target, suspect screening and nontarget analysis using LC-HRMS and GC-HRMS. *Environ. Pollut.* **276**, 116701 (2021).
 57. Lowe, C. N. et al. Chemical characterization of recycled consumer products using suspect screening analysis. *Environ. Sci. Technol.* **55**, 11375–11387 (2021).
 58. Knorr, A. et al. Performance evaluation of a nontargeted platform using two-dimensional gas chromatography time-of-flight mass spectrometry integrating computer-assisted structure identification and automated semiquantification for the comprehensive chemical characterization of a complex matrix. *Anal. Chem.* **91**, 9129–9137 (2019).
 59. Worton, D. R. et al. Comprehensive chemical characterization of hydrocarbons in NIST standard reference material 2779 Gulf of Mexico crude oil. *Environ. Sci. Technol.* **49**, 13130–13138 (2015).
 60. Organtini, K. L. et al. Comprehensive characterization of the halogenated dibenzo-p-dioxin and dibenzofuran contents of residential fire debris using comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry. *J. Chromatogr. A* **1369**, 138–146 (2014).
 61. Lin, P., Fleming, L. T., Nizkorodov, S. A., Laskin, J. & Laskin, A. Comprehensive molecular characterization of atmospheric brown carbon by high resolution mass spectrometry with electrospray and atmospheric pressure photoionization. *Anal. Chem.* **90**, 12493–12502 (2018).
 62. Paige, J. Comprehensive characterization of per- and polyfluoroalkyl substances (PFASs) in manufacturing wastewater (A conference poster, SciMeetings, ACS Fall 2020 Virtual Meeting, USA, August 17. <https://doi.org/10.1021/scimeetings.0c06636> (2020).
 63. Tang, C., Tan, J., Wang, C. & Peng, X. Determination of perfluorooctanoic acid and perfluorooctane sulfonate in cooking oil and pig adipose tissue using reversed-phase liquid-liquid extraction followed by high performance liquid chromatography tandem mass spectrometry. *J. Chromatogr. A* **1341**, 50–56 (2014).

ACKNOWLEDGEMENTS

This study was financially supported by the National Natural Science Foundation of China (Grant No. 51878170), the Guangdong Basic and Applied Basic Research Foundation (Grant Nos. 2021B1515140019, 2022A1515140049), the National Key R&D Program of China (2019YFC1804502) and the Scientific Research Team Project of Dongguan University of Technology (No. TDQN2019010).

AUTHOR CONTRIBUTIONS

C.T. conceived the project, designed and conducted the experiment, wrote the manuscript, and organized the work. Y.L., K.W. and J.L. carried out the sample pretreatment and instrumental analysis. Y.Z., X.L., X.P., B.M., and Q.H. improved the manuscript and figures. H.L. contributed to the conceptualization, provided resources and supervised the work.

COMPETING INTERESTS

The authors declare no competing interests

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41545-023-00220-6>.

Correspondence and requests for materials should be addressed to Caiming Tang or Hui Lin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023