

# 基于特征碘离子的非靶向筛查鉴定河北省 地下水中碘化消毒副产物

钟蔚 刘思琪 董艳冉 许楠<sup>†</sup>

深圳市重金属污染控制与资源化重点实验室, 北京大学深圳研究生院环境与能源学院, 深圳 518055;

<sup>†</sup> 通信作者, E-mail: xunan@pkusz.edu.cn

**摘要** 联合高分辨质谱和碘化消毒副产物(I-DBPs)的特征质谱性质, 建立一套完整的非靶向分析方法。以河北省17口监测井为研究对象, 筛查地下水样品中I-DBPs的种类、数量及分布。共筛选出含有同分异构体的I-DBPs疑似离子2408种, 其中不同质荷比的数量为839, 远高于以往研究报道的I-DBPs数量。疑似I-DBPs离子中, 对碘离子响应强度排前十位的离子进行结构鉴定, 其中响应强度排前两位的两种离子被鉴定为酚类I-DBPs, 其发育毒性比对应的脂肪族I-DBPs高数十至数百倍。进一步的分析结果表明, 酚类I-DBPs是样品中主要的I-DBPs类型之一。最后, 通过主成分分析, 讨论2408种疑似I-DBPs的分布特征。根据主成分分数, 筛选出明显偏离其他采样点的3个采样点。根据装载因子, 通过高斯混合模型对疑似I-DBPs进行聚类, 得到4类离子, 其中3类离子分别属于3个离群采样点的特征污染物。这些结果表明, I-DBPs的分布在河北省内的不同地区具有明显的差异性和复杂性。

**关键词** 碘化消毒副产物(I-DBPs); 非靶向筛查; 高分辨质谱; 地下水; 主成分分析

## Identification of Iodinated Disinfection Byproducts in Groundwater from Hebei Province Using Iodide-Based Nontarget Screening

ZHONG Wei, LIU Siqi, DONG Yanran, XU Nan<sup>†</sup>

Key Laboratory for Heavy Metal Pollution Control and Reutilization, School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen 518055; <sup>†</sup> Corresponding author, E-mail: xunan@pkusz.edu.cn

**Abstract** A complete set of nontargeted analysis method was established by combining the high-resolution MS and the characteristic mass spectrometry properties of iodinated disinfection byproducts (I-DBPs). Groundwater samples were collected from 17 sites in Hebei Province, and an integrated nontarget screening workflow was developed and implemented to identify the category, quantity and distribution of I-DBPs in the groundwater samples. As a result, a total of 2408 suspect I-DBPs ions including isomers were screened, of which 839 unique mass-to-charge ratios were identified, much higher than the numbers reported by previous studies. Among the suspect I-DBPs ions, the structures were carefully analyzed for precursors which were detected with top 10 the most abundant iodide ions in their MS/MS spectra, and the top 2 suspect ions were identified as phenolic I-DBPs, whose developmental toxicity are tens to hundreds of times higher than the corresponding aliphatic DBPs. Further analysis confirmed that the phenolic I-DBPs were one of the main categories of all the I-DBPs detected in samples. Finally, the distribution characteristics of 2408 suspect ions were analyzed based on principal component analysis (PCA). According to the PCA, 3 sites significantly different from other sampling sites were screened as outlier sites. The suspect ions were clustered into 4 classes using Gaussian mixture model according to their loading scores, and ions from 3 of 4 classes were identified as the diagnostic pollutants for 3 outlier sites respectively. The results highlighted the differences and complexity of I-DBPs in groundwater of Hebei Province.

**Key words** iodinated disinfection byproducts (I-DBPs); nontarget screening; high-resolution mass spectrometry; groundwater; principal component analysis

通过杀死病原体,水处理中的消毒过程大大降低了各类介水传染病的发病率,添加含氯消毒剂(如氯气和氯胺)的消毒过程已成为现代水处理工艺中的重要环节。然而,消毒过程和消毒后的余氯均可以与被处理水体中溶解性有机质(dissolved organic matter, DOM)反应生成种类众多的消毒副产物(disinfection byproducts, DBPs)<sup>[1]</sup>。有研究表明, DBPs 具有多种毒性(如生殖、发育和基因毒性)以及致癌性<sup>[2-5]</sup>。当水体中存在卤素离子(氯、溴和碘)时, DBPs 倾向于进一步生成更加复杂的卤化 DBPs。其中碘化消毒副产物(iodinated DBPs, I-DBPs)的细胞和基因毒性显著高于对应的氯化物和溴化物 DBPs<sup>[6]</sup>, I-DBPs 中的碘元素可来自无机含碘化合物(如碘化物和碘酸盐)和各类碘化有机化合物<sup>[7]</sup>, 消毒过程中氧化生成的次碘酸也可进一步与 DOM 反应生成 I-DBPs<sup>[8]</sup>。I-DBPs 在饮用水(特别是自来水)中已有检出。另外,由于家庭加碘食盐的使用,可能会生成更多种类和数量的 I-DBPs<sup>[9]</sup>。DOM 组成复杂,生成的 I-DBPs 种类繁多,结构复杂,分析难度较高,现有的研究方法很难对样品中 I-DBPs 进行全面的分析。

基于高分辨质谱的非靶向筛查是近年来分析环境样品中未知污染物的主要方法<sup>[10-15]</sup>。高分辨质谱通常会生成海量的数据,其中包括大量的背景干扰离子<sup>[16]</sup>。从海量数据中定向地筛选出疑似目标物离子,是非靶向筛查鉴定未知污染物的第一步。碘离子可作为特征碎片离子用于疑似 I-DBPs 的筛选。以往研究中,低分辨质谱条件下的母离子扫描(precursor ion scan, PIS)已成功用于筛查 I-DBPs<sup>[9,17]</sup>。筛选疑似目标物后,后续鉴定过程包括分子式计算和结构解析两步。分子式计算的主要依据为质谱检测的离子质荷比。一般情况下,质荷比精确度越高,在误差范围内的可能元素组成个数越少,越有可能确定分子式<sup>[18]</sup>。高分辨质谱的一大优势即为质荷比检测的高精确度(一般小于  $5 \times 10^{-6}$ )<sup>[19]</sup>。同时,一些辅助手段如特征同位素峰以及碎片离子信息可帮助缩小元素组成范围。

由于 I-DBPs 的生成与消毒过程十分相关,以往的 I-DBPs 研究通常关注污水处理过程<sup>[6,8-9,20]</sup>,对天然水体关注较少。另一方面,受限于分析仪器和分

析方法,通常只能分析有限数量的 I-DBPs,覆盖度不够。本文通过 Orbitrap 高分辨质谱和自建的非靶向筛查方法,较为全面地分析地下水中 I-DBPs 的种类、数量和分布,以期为地下水 I-DBPs 污染防控以及优先控制污染物的筛选提供数据支撑。

## 1 材料与方法

### 1.1 样品信息

充分考虑风向、地下水补给模式、土壤性状及类型等因素,选定河北省 17 口监测井为采样点。地下水样品采集于 2017 年 5 月,每个采样点采集两个平行样,同时设置一个空白样。采样容器为 4.5 L 的聚丙烯采样瓶,采样后在避光和 4°C 条件下保存,运回实验室后 24 小时内处理完毕。采样点分布如图 1 所示。

### 1.2 样品前处理

本研究利用固相萃取法富集样品中的有机污染物。在进行固相萃取之前,样品经多层玻璃纤维滤膜(GF/C 1.2 μm 和 GF/F 0.7 μm, 1.2 μm 滤膜在上面;英国 Whatman 公司)过滤。Oasis HLB 固相萃取柱

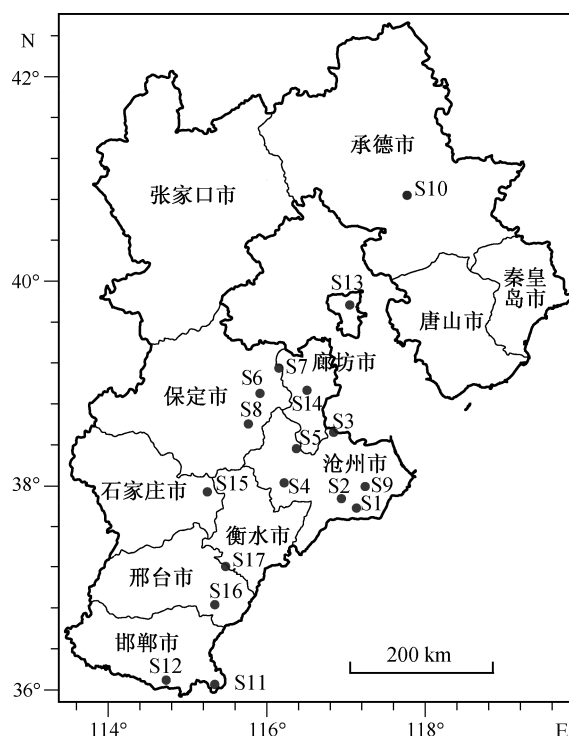


图 1 17 口采样井分布

Fig. 1 Distribution of 17 sampling sites

(500 mg, 6 mL; 美国 Waters 公司)先用 6 mL 甲醇(色谱纯; 德国 Merck 公司)和 6 mL 超纯水(电阻率 18.2 mΩ/cm, 美国 Millipore 公司)活化。样品经过萃取柱后, 使用 5 mL 超纯水淋洗, 去除样品中的亲水杂质, 并在真空下抽干 2 小时。然后使用 10 mL 甲醇洗脱样品, 洗脱液用 15 mL 聚丙烯离心管收集, 在氮吹仪中吹至近干, 再用甲醇定容至 1 mL。最后, 溶液通过 0.22 μm 针头过滤器(中国 Anpel 公司)过滤, 移入聚丙烯进样小瓶, 放入 -20°C 冰箱保存。

### 1.3 超高效液相色谱串联高分辨质谱分析

本研究分析平台为 Dionex UltiMate 3000 超高效液相色谱串联 Q-Exactive Focus 高分辨质谱系统(美国 Thermo Fisher 公司)。液相色谱选用 ZORBAX Extend plus C18 色谱柱(2.1×100 mm, 1.8 μm; 美国 Agilent 公司), 并连接一个 ZORBAX Extend C18 保护柱(2.1×100 mm, 1.8 μm; 美国 Agilent 公司)。

色谱系统流动相 A 是 2 mM 醋酸铵(中国百灵威公司)水溶液, 并用氨水调节 pH 至 10.25。流动相 B 由 5 mM N-甲基哌啶(中国百灵威公司)的甲醇/乙腈(体积比为 1:1, 色谱纯; 德国 Merck 公司)溶液组成。流速为 0.25 mL/min, 柱温为 40°C。初始流动相比例为 A:B=90%:10%, 持续 2 分钟后, 溶液 B 的比例在 6 分钟内提高到 95%, 并持续 7 分钟, 之后溶液 B 的比例在 0.2 分钟内降低到 10%, 并持续 3.8 分钟, 直到结束。

质谱的离子源采用负模式的电喷雾电离(electrospray ionization, ESI), 质谱数据采集模式为 full scan MS/data-dependent MS/MS (dd-ms<sup>2</sup>)。其中二级质谱母离子选择范围为当前时刻前三强离子, 母离子分步加热碰撞能(15, 30 和 45 eV)碎裂, 最终二级质谱为 3 种碰撞能情形下离子碎裂结果的加权平均植。一级质谱分辨率为 70000(质荷比=200 时), 扫描范围为 150~1000; 二级质谱分辨率为 17500(质荷比=200 时), 扫描范围为 60~750。

### 1.4 非靶向筛查方法

如图 2 所示, 本研究针对 I-DBPs 质谱特征, 建立一套完整的非靶向筛查流程。按顺序, 流程可分为数据预处理、I-DBPs 疑似离子筛查和 I-DBPs 鉴定三部分。

数据预处理的第一步为数据格式转换。仪器分析的原始数据为 raw 格式文件, 为了满足后续分析需求, 要将其转换为 mzXML 格式文件。本研究使用 Proteowizard 软件<sup>[21]</sup>进行格式转换, 同时将原始文

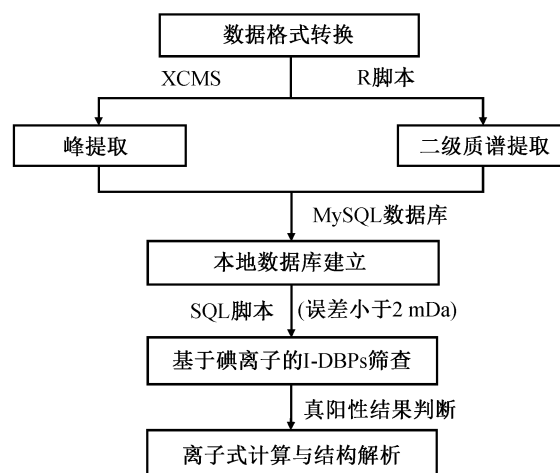


图 2 针对 I-DBPs 建立的非靶向筛查流程  
Fig. 2 Workflow of nontarget screening for I-DBPs

件质谱格式从 profile 转换为 centroid, 从而减小输出文件的大小, 提高后续数据处理效率。格式转换后进行峰提取, 使用 R(3.5.2 版本, R Development Core Team)环境下的 XCMS(3.4.2 版本)<sup>[22]</sup>作为峰提取工具。峰提取主要依赖 XCMS 提供的 centwave 函数, 主要参数设置为 peakwidth=c(5,20), noise=5000, ppm=5, mzdifff=0.003。上述参数是根据实验数据迭代处理优化而设置的, 其中 peakwidth 接近本实验中大部分色谱峰的峰宽, noise 和 ppm 基于质谱的检测能力设置, 主要用于筛除干扰数据, 而 mzdifff 的设置值(0.003)与仪器分辨率接近, 经数据处理验证可以有效地避免错误的峰合并。输出文本数据, 进一步通过 CAMERA(1.38.1 版本)<sup>[23]</sup>进行峰合并, 用来检测来源于同一分子的不同色谱峰。CAMERA 在默认参数条件下运行。最终的峰提取结果输出为 CSV 文件。峰提取后, 数据文件通过 PeakDensity 算法进行峰匹配, 提取特征(Features)。另外, 本研究基于 XCMS 提供的 chromPeakSpectra 函数, 编写 R 脚本来提取二级质谱, 并输出为 CSV 文件。经过以上步骤后, 进一步通过 MySQL(Community Server 8.0.11)将峰表和二级质谱转换为本地关系数据库。得到数据库文件后, 通过 SQL 脚本即可实现筛查需求。与纯 CSV 文件或其他软件相比, 通过 SQL 检索数据库速度更快、灵活性更高。

数据预处理后, 进行 I-DBPs 疑似离子的筛查。在上述数据库基础上, 通过 SQL 脚本实现基于碘离子的非靶向筛查, 即通过筛选二级质谱存在碘离子

碎片(质荷比=126.9044)的离子,定向地获取 I-DBPs。筛查过程中,规定碎片离子质荷比误差应小于 2 mDa。另外,由于碘离子显著的质量亏损(negative mass defect)<sup>[24]</sup>,在常见的有机物组成元素限制下,质荷比为 126.9044 的离子在高分辨质谱的误差范围内(一般小于  $5 \times 10^{-6}$ ),碘离子是唯一可能的选项,因此可以准确地筛查 I-DBPs。

在鉴定疑似 I-DBPs 前需要进行数据检查来过滤背景干扰<sup>[25]</sup>,提高分析准确性。过滤标准包括平行样间离子响应相对标准偏差小于 30%,样品平均响应大于空白样品对应离子响应的 10 倍。鉴定过程中,分子式通过 ChemCalc<sup>[26]</sup>在线工具计算,设定的元素组成范围为 C=1~20, H=0~50, O=0~10, N=0~10, P=0~5, S=0~5, I=0~5。另外,特征同位素峰<sup>[27]</sup>也用于辅助判断分子式。结构解析依赖二级质谱,碎片离子元素组成计算方法如上所述。

非靶向筛查的结果可信度一般通过赋予置信等级来判断,本文使用 Schymanski 等<sup>[28]</sup>提出的五等级判断标准,该标准在非靶向筛查中使用最为广泛。

## 2 结果与讨论

### 2.1 疑似 I-DBPs 离子筛查结果

本研究 17 个采样点的 34 个样品中检测到峰的平均数量为 9304 个,峰匹配后检测到的特征总数为 49909 个,高于以往文献报道的峰或特征数量<sup>[29-30]</sup>。通过基于碘离子搜索的非靶向筛查,在 17 个采样点的 34 个样品中共检测到含碘离子的母离子 7732 个,扣除质荷比重复的母离子后,剩余 1200 个离子。这些离子中可能包含背景离子或仪器数据采集误差导致的假阳性结果以及由于同位素分布、不同加和物形式或中性损失形成的同源峰,因此实际 I-DBPs

应小于此数字。17 个采样点检测到的疑似 I-DBPs 离子数量及相对丰度(用仪器响应表示)如图 3 所示。与峰数量相比,采样点间 I-DBPs 疑似离子数量差别更大,每个采样点平均检测到 227 个 I-DBPs 疑似离子,其中采样点 10 检测到的疑似离子最多,达到 429 个(取平行样的最大值),采样点 8 仅检测到 39 个疑似离子。由图 3 可知,疑似离子检测数量越多,其响应总和越高,但图中越靠近右上角(即离子数量多、响应高的采样点),采样点分布越稀疏,显示出不同采样点间 I-DBPs 种类和数量的差异性,具体分布情况将在 2.3 节讨论。

在进行进一步鉴定前,需要对数据进行清洗,过滤掉背景干扰离子。在初步观察数据时,我们发现大量液相色谱中常见的干扰离子被鉴定为 I-DBPs 疑似离子,如质荷比为 263.1393(离子式  $C_{14}H_{21}O_5^-$ )、263.1288(离子式  $C_{15}H_{19}O_4^-$ )和 277.1445(离子式  $C_{16}H_{21}O_4^-$ )的离子。这些离子及其碎片离子不可能含有碘离子,只是因为对应的二级质谱中检测到碘离子碎片而被筛选为疑似离子,出现这种现象主要是由于对应的干扰离子保留时间附近存在真实的含碘化合物,这些含碘化合物碎裂后生成的碘离子即是这些干扰离子中碘离子的来源。一般的含碘化合物在 ESI 条件下非常容易脱离母离子,因此碘离子响应强度较高,而这些干扰离子的二级质谱中碘离子响应强度普遍很低。因此,我们首先设定疑似 I-DBPs 离子中,只保留二级质谱中碘离子响应强度至少达到母离子响应 20%的离子。此外,由于 I-DBPs 大多带有较高的质量亏损,本文设置分段的质量亏损过滤器,对于质荷比<350 的离子,要求质量亏损必须为负,其余离子要求质量亏损<0.15。

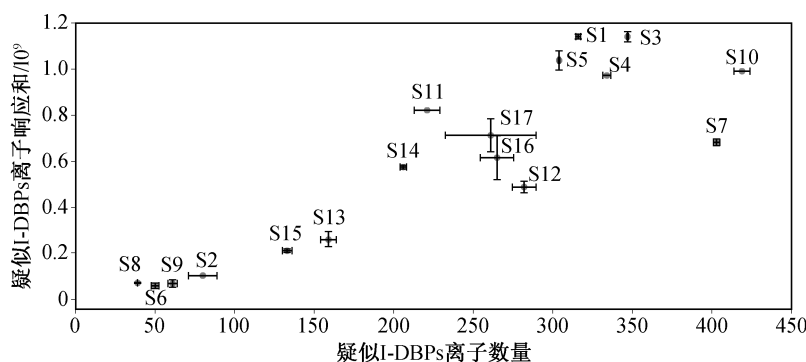


图 3 17 个采样点的疑似 I-DBPs 数量以及相对丰度(误差棒基于平行样计算)

Fig. 3 Number and relative abundance of suspect I-DBPs in 17 sampling sites (error bars were calculated based on replicates)

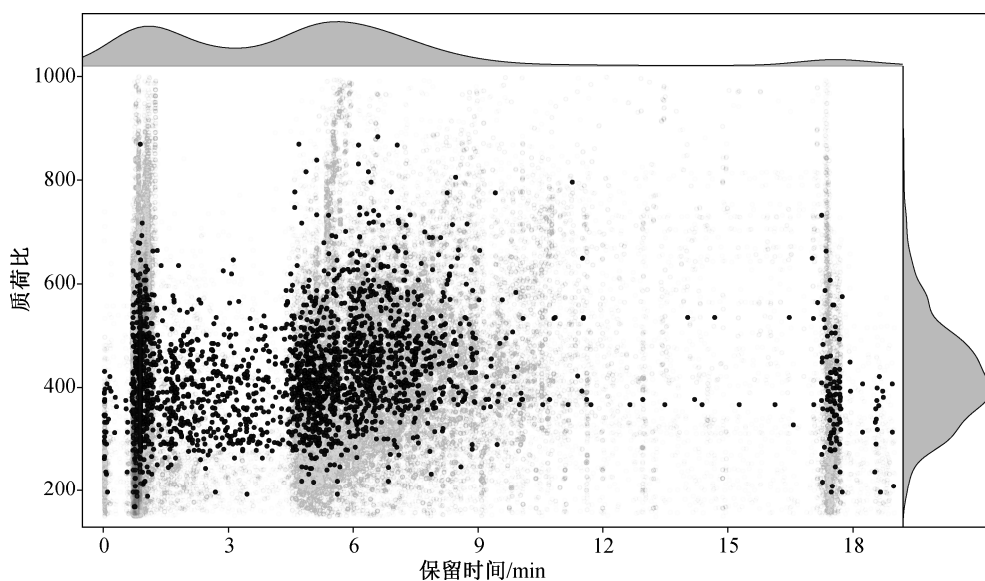
经过以上两步过滤后, 1200个疑似离子最终剩余 839个离子用于后续分析, 分布如图 4所示。839个离子对应的实际特征总数为 2408个, 这是因为一些离子的同分异构体被色谱分开, 从而被识别为不同的特征。由图 4可看出, 大部分 I-DBPs 疑似离子的质荷比集中在 400左右, 保留时间则明显存在两个聚集区, 分别为 1分钟和 6分钟附近, 根据本研究的实验条件, 可推测保留时间的两个聚集区分别对应极性强的(1分钟)和极性一般的(6分钟)的两类物质。

## 2.2 主要 I-DBPs 鉴定

由于疑似 I-DBPs 数量极大, 难以对其中每个离子进行细致的结构鉴定, 本研究仅对其中较为重要的离子进行鉴定。由于二级质谱中碘离子碎片为 I-DBPs 的质谱特征, 我们以碘离子响应强度为标准, 筛选用于结构鉴定的离子。首先按 839个疑似 I-DBPs 离子二级质谱中碘离子响应强度从高到低排序, 然后取响应最强的前 10种不同质荷比的离子用于分析, 这 10种离子的基本信息及可能离子结构如表 1所示。可以看出, 10种主要离子中, 除质荷比为 403.0409的离子(离子式  $C_{16}H_{20}IO_4^-$ , 质量误差为 0.27 mDa)外, 均有显著的质量亏损, 可根据 1.4节的方法准确地确定其离子组成。对于质荷比为 403.0409的离子, 在设置的元素组成及误差范围内存在超过 10种可能组成, 但通过搜索同源峰, 仅观

察到一个质荷比等于  $[M+1.003]^-$  的峰(即  $^{13}C$  同源峰), 排除离子中含硫、氯和溴元素的可能。另外, 由于离子由分子脱氢生成, 排除不饱和度为整数的结果, 最后仅剩  $C_{16}H_{20}IO_4^-$  这一种结果。因此, 所有 10种 I-DBPs 疑似离子的离子式都可确认。10种离子中包含不属于 I-DBPs 类物质的三碘阴离子  $I_3^-$ 。

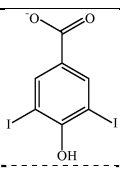
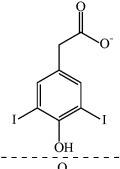
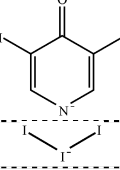
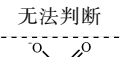
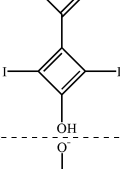
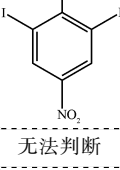
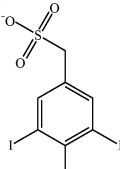
除  $I_3^-$  离子外, 其余 9种离子均可认定为 I-DBPs 物质, 可基于它们的离子组成和二级质谱谱图, 可对其离子结构进行分析。以质荷比为 402.8332的离子(离子式  $C_8H_5I_2O_3^-$ , 0.39 mDa)为例, 采样点 3是此离子响应强度最高的采样点, 其提取离子图(图 5(a))显示, 在保留时间 0.5~7分钟内存在大量分离度高的色谱峰, 表示存在多种同分异构体。图 5(a)中色谱峰较明显地聚集在 2分钟和 5分钟附近, 分别以其中具代表性的 B 和 C 峰为例(图 5(b)和(c))提取二级质谱, 可观察到不同的碎裂性质。图 5(b)和(c)均显示检测到显著的碘离子(质荷比 126.9050, 误差 0.60 mDa)和母离子(质荷比 402.8333, 误差 0.40 mDa)。图 5(c)显示同时检出 358.8437 离子, 对应母离子脱去一个  $CO_2$  基团后的质量(0.72 mDa)。另外, 根据母离子的离子式  $C_8H_5I_2O_3^-$ , 5.5的不饱和度, 最可能的来源为苯环结构, 因此可推断图 5(c)中的同分异构体应为碘酚酸结构, 其可能的结构在表 1中给出。然而, 根据现有数据, 碘原子、羟基和羧酸基团均不能确认其位置, 因此置信等级仅为 3。5分



黑色圆点代表疑似 I-DBPs 离子, 灰色圆点代表其他离子, 上方和右方的核密度图为疑似 I-DBPs 的分布特征

图 4 疑似 I-DBPs 离子分布  
Fig. 4 Distribution of suspect I-DBPs

表 1 二级质谱中碘离子响应强度前 10 位的含碘离子对应的母离子信息  
Table 1 Information of precursors detected with top 10 iodide ion abundances in their MS/MS spectra

质荷比	离子式	质量误差/mDa	保留时间/min	可能结构	检出率/%
388.8177	$C_7H_3I_2O_3^-$	0.55	5.88, 0.72, 4.76, 1.77		94.12
402.8332	$C_8H_5I_2O_3^-$	0.39	5.97, 5.73, 5.30, 5.10, 0.90, 2.25, 8.41		88.24
345.8231	$C_5H_2I_2NO^-$	0.53	2.75, 1.63, 5.13, 0.90, 6.54		64.71
380.7141	$I_3^-$	0.70	0.83		100.00
339.9064	$C_5H_{12}I_2N^-$	0.49	0.78	无法判断	100.00
362.8022	$C_5HI_2O_3^-$	0.70	1.40, 0.89, 3.12, 2.13, 4.68		76.47
389.8129	$C_6H_2I_2NO_3^-$	0.55	5.38, 1.33, 5.92, 4.65		94.12
471.7196	$C_5HI_3NO^-$	0.43	5.34, 6.02	无法判断	35.29
438.8001	$C_7H_5I_2O_4S^-$	0.31	5.82, 4.65, 0.94		88.24
403.0409	$C_{16}H_{20}I_4^-$	0.27	5.78, 7.33, 6.43	无法判断	76.47

说明: 一些离子的多个保留时间对应不同的同分异构体的保留时间; 列出可能结构的离子置信等级为 3, 其余离子无法判断结构, 离子置信等级为 4。

钟附近的其他同分异构体的二级质谱与图 5(c)相同, 无法进一步分析。从图 5(b)中可观察到一种响应微弱的质荷比为 246.9262 的离子, 通过计算得到其离子式为  $C_7H_4IO_2^-$  (0.60 mDa), 是在母离子的基础上脱去  $COHI^-$  形成的。另外, 图 5(b)中完全检测不到  $[M-CO_2]^-$  碎片。以上两点特征暗示, 图 5(b)代表的同分异构体并不是碘酚酸结构。考虑到其显著短的保留时间, 更可能是极性更强的脂肪族化合物结构, 但具体结构尚无法确定。其他离子的结构鉴定过程与质荷比为 402.8333 的离子类似。

除质荷比为 402.8332 的离子属于碘酚类 I-DBPs 外, 碘离子响应前 10 位的离子中还有质荷比为 388.8177 (离子式  $C_7H_3I_2O_3^-$ , 0.55 mDa)、438.8001 (离子式  $C_7H_5I_2O_4S^-$ , 0.30 mDa) 和 389.8129 (离子式  $C_6H_2I_2NO_3^-$ , 0.55 mDa) 的 3 种离子也属碘酚类结构。4 种离子中除质荷比为 438.8001 的离子 (离子结构中带有磺酸基团) 外, 其余 3 种离子均在饮用水中有过报道<sup>[9,17]</sup>, 且存在多种同分异构体, 本研究结果与文献<sup>[9,17]</sup>类似。另据报道, 芳香族 I-DBPs 的发育毒性比对应的脂肪族 I-DBPs 高十甚至几百

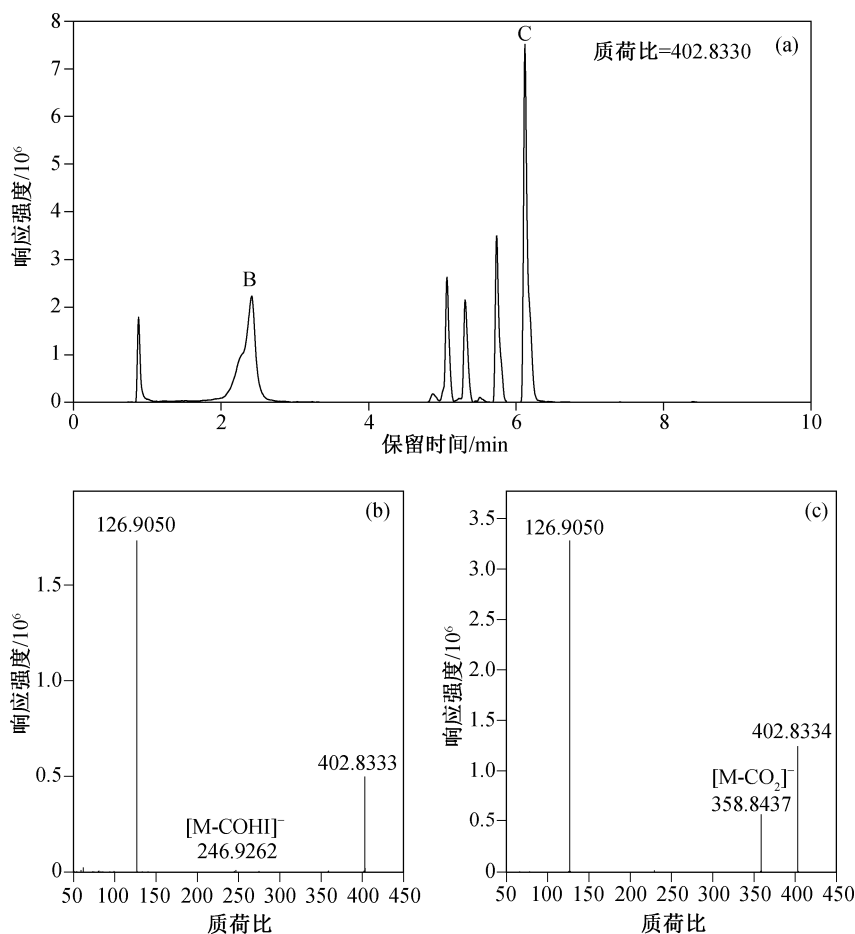


图5 采样点3中的 $C_8H_5I_2O_3^-$ 提取离子色谱图(a)和两种同分异构体的二级质谱图((b)和(c))  
Fig. 5 Extracted ion chromatograph of  $C_8H_5I_2O_3^-$  in site 3 (a) and MS/MS spectra from two different isomers ((b) and (c))

倍<sup>[31]</sup>。这里的4种芳香族I-DBPs离子检出率均大于85%，且其中质荷比为388.8177和402.8332的两种芳香族I-DBPs是检测响应强度最高的两种I-DBPs。样品中还存在更多不同离子式的芳香族I-DBPs，是某些采样点I-DBPs中的特征离子。

### 2.3 基于主成分分析的I-DBPs分布特征

本研究筛查出的疑似I-DBPs种类、数量众多，本文通过主成分分析将数据降维，可视化地展现样品间I-DBPs的分布特征。由于不同疑似I-DBPs离子的响应强度差异巨大( $10^5 \sim 10^9$ )，如果直接使用原始响应强度，降维结果将主要受高响应离子影响，掩盖其他离子的分布特征。因此，在进行主成分分析前，我们对原始数据矩阵进行归一化，使得每个疑似离子在样品间的平均值为0，标准差为1。主成分分析中主成分个数取最大值17。

提取17个采样点的前3个主成分后，绘制散点图(图6(a))。除采样点1、3和7外，其他14个采样点有明显的聚集性，在3个主成分方向没有观察到

明显的分布趋势。采样点1、3和7分别在第2、1和3主成分方向上明显地偏离其他采样点，说明这3个采样点有显著不同的污染特征。进一步提取所有主成分的方差解释率，结果如图6(b)所示。第一主成分仅能解释16%的方差，而前3个主成分总共只能解释40%的方差，充分说明样品中I-DBPs组成的复杂性，不同采样点间I-DBPs包含的种类差异较大。由于I-DBPs疑似离子数量庞大，检出率不一，不适合进行相关性及类似的分析。综上所述，前3个主成分可分别代表采样点1、3和7的分布特点。

我们提取2408个疑似离子在前3个主成分上的装载因子得分(loading scores)，结果见图6(c)。由三维散点图可观察到，2408个疑似离子存在明显的分离和聚集。本文通过高斯混合模型，对这些离子进行聚类。高斯混合模型是常用的聚类算法，其类别的分界线可以是非线性的曲线。本文数据复杂度较高，因此适用于高斯混合模型。聚类过程尝试设置分类数分别为3~6，最终确认分类数为4，聚类后画

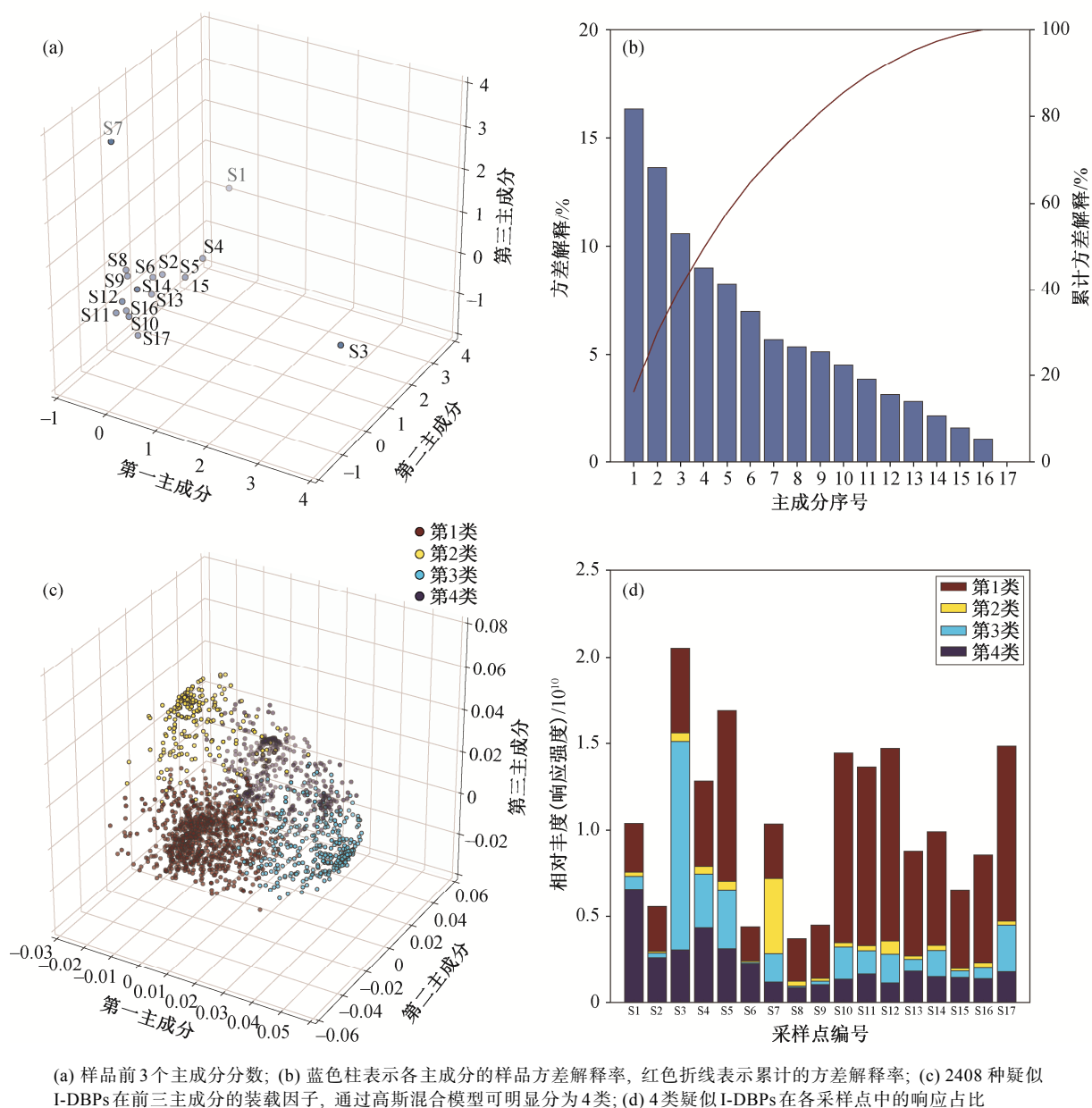


图 6 主成分分析结果

Fig. 6 Results of principal component analysis

出各采样点4类离子的响应强度(图6(d))。4类离子中, 第1类离子明显以原点(0, 0, 0)为中心聚集, 代表对前3个主成分几乎没有贡献的那些离子, 2.2节中鉴定的10种主要I-DBPs离子中,  $C_7H_3I_2O_3^-$  (0.55 mDa)、 $C_5H_2I_2NO^-$  (0.42 mDa)和 $C_6H_2I_2NO_3^-$  (0.69 mDa) 3种离子的某个同分异构体属于第1类。第一主成分上, 第3类离子是主要的贡献者, 之前鉴定的 $C_7H_3I_2O_3^-$ 和 $C_5H_2I_2NO^-$ 属于此类。第1类离子也包括了 $C_7H_3I_2O_3^-$ 在此处属于同分异构体, 其第1类和第3类离子的保留时间分别为1分钟和5分钟,

很可能属于脂肪和芳香族I-DBPs, 因此其分布也有较大的差异。第3类离子按照第三主成分装载因子得分排序, 前五位的离子式分别为 $C_{11}H_{10}IN_2O_4^-$  (0.77 mDa)、 $C_5H_2I_2NO^-$  (0.42 mDa)、 $C_6H_3I_3NO_2^-$  (0.81 mDa)、 $C_8H_4I_3O_4S^-$  (1.16 mDa)和 $C_7H_5I_2O_3S^-$  (0.42 mDa), 其中4种均为芳香族I-DBPs, 且这些第3类中的显著离子在采样点3检测到更高的响应强度, 这正是采样点3在第三主成分方向上偏离其他采样点的原因。第2类和第4类离子的分析方法与第3类相似。第2类和第4类中显著的离子分别在采样

点7和1检测到更高的响应,其中第2类中的显著离子大多数带有较弱的质量亏损,导致离子式不能完全确定,如分数在前两位的质荷比为384.9579(离子式可能为 $C_{14}H_{10}IO_5^-$ , 0.61 mDa)和334.976852(离子式不确定)的离子。第4类中最显著的离子为 $C_{12}H_{12}IO_4^-$ (0.57 mDa)的两种同分异构体。如图6(b)所示,第4~6主成分仍然对方差有不小的贡献度,但在数据分析中,这些主成分并不存在突出的采样点,故不详细讨论。

不同采样点间(特别是采样点1,3和7)占主导的I-DBPs有很大的差异性,前3个主成分仅能解释40%的方差,若考虑更多的主成分,剩余14个采样点间同样存在很大的差异。

### 3 结论

1) 基于碘离子搜索的非靶向筛查,辅以建立本地数据库的方法,可以全面而快速地筛选I-DBPs,也可筛查其他卤素元素取代的消毒副产物。

2) 本文研究的样品中I-DBPs数量庞大,其中存在大量不同结构的毒性较强的酚类I-DBPs。在17个采样点中共检测到不同质荷比的I-DBPs疑似离子839种,考虑同分异构体共有2408种,远多于传统靶向方法能分析的化合物数量。酚类I-DBPs是主要的I-DBPs类型之一,也是一些采样点的特征I-DBPs离子,并且在碘离子响应强度前10的离子中,有4种属于酚类I-DBPs。酚类I-DBPs的毒性更强,但在环境中的检测报道较少或尚未报道。未来,对环境中酚类I-DBPs进行更细致的结构解析及生态效应研究十分必要。

3) I-DBPs种类繁多,结构复杂,不同采样点存在不同的主要I-DBPs类型,差异明显。通过主成分分析,可筛选出3个具有独特污染特征的采样点,进而筛选出特征I-DBPs,显示这些采样点的主导I-DBPs有很大的差异。本文将I-DBPs疑似离子按采样点特征分为4类,其中3类分别属于这3个采样点的特征污染物。另外,每一类都包括不同结构的酚类I-DBPs,说明这些物质在自然环境中的存在形式具有复杂性,急需更多的检测数据来明晰I-DBPs在环境中的赋存形式。

### 参考文献

[1] Wang X S, Song H, Zhang J, et al. Chlorination decreases acute toxicity of iodophenols through the

formation of iodate and chlorinated aliphatic disinfection byproducts. *Water Research*, 2021, 194: 116951

- [2] Andrzejczyk N E, Greer J B, Nelson E, et al. Novel disinfection byproducts formed from the pharmaceutical gemfibrozil are bioaccumulative and elicit increased toxicity relative to the parent compound in marine polychaetes (*Neanthes arenaceodentata*). *Environmental Science & Technology*, 2020, 54(18): 11127–11136
- [3] Han J, Zhang X. Evaluating the comparative toxicity of DBP mixtures from different disinfection scenarios: a new approach by combining freeze-drying or rotoevaporation with a marine polychaete bioassay. *Environmental Science & Technology*, 2018, 52(18): 10552–10561
- [4] Yang J, Li W, Zhu Q, et al. Identification, formation, and predicted toxicity of halogenated DBPs derived from tannic acid and its biodegradation products. *Environmental Science & Technology*, 2019, 53(22): 13019–13030
- [5] Tang H, Zhong H, Pan Y, et al. A new group of heterocyclic nitrogenous disinfection byproducts (DBPs) in drinking water: role of extraction pH in unknown DBP exploration. *Environmental Science & Technology*, 2021, 55(10): 6764–6772
- [6] Dong H, Qiang Z, Richardson S D. Formation of iodinated disinfection byproducts (I-DBPs) in drinking water: emerging concerns and current issues. *Accounts of Chemical Research*, 2019, 52(4): 896–905
- [7] Ding S, Deng Y, Li H, et al. Coagulation of iodide-containing resorcinol solution or natural waters with ferric chloride can produce iodinated coagulation byproducts. *Environmental Science & Technology*, 2019, 53(21): 12407–12415
- [8] Hu J, Xu Y, Chen Y, et al. Formation of carbonaceous and nitrogenous iodinated disinfection byproducts from biofilm extracellular polymeric substances by the oxidation of iodide-containing waters with lead dioxide. *Water Research*, 2021, 188: 116551
- [9] Pan Y, Zhang X, Li Y. Identification, toxicity and control of iodinated disinfection byproducts in cooking with simulated chlor(am)inated tap water and iodized table salt. *Water Research*, 2016, 88: 60–68
- [10] Hollender J, Schymanski E L, Singer H P, et al. Non-target screening with high resolution mass spectrometry in the environment: ready to go?. *Environmental Science & Technology*, 2017, 51(20): 11505–

- 11512
- [11] Wang L, Kang Q, Jia Y, et al. Identification of three novel chloroalkyl organophosphate triesters in house dust using halogenation-guided nontarget screening combined with suspect screening. *Environmental Science & Technology*, 2021, 55(4): 2482–2490
- [12] Qiu M, Hu A, Huang Y M, et al. Elucidating degradation mechanisms of florfenicol in soil by stable-isotope assisted nontarget screening. *Journal of Hazardous Materials*, 2021, 403: 123974
- [13] Wu H, Chen J, Peng J, et al. Nontarget screening and toxicity evaluation of diol esters of okadaic acid and dinophysistoxins reveal intraspecies difference of *Proocentrum lima*. *Environmental Science & Technology*, 2020, 54(19): 12366–12375
- [14] Washington J W, Rosal C G, McCord J P, et al. Nontargeted massspectral detection of chloroperfluoropolyether carboxylates in New Jersey soils. *Science*, 2020, 368: 1103–1107
- [15] Renaguli A, Fernando S, Hopke P K, et al. Nontargeted screening of halogenated organic compounds in fish fillet tissues from the Great Lakes. *Environmental Science & Technology*, 2020, 54(23): 15035–15045
- [16] Hollender J, van Bavel B, Dulio V, et al. High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. *Environmental Sciences Europe*, 2019, 31(1): 1–11
- [17] Pan Y, Li W, An H, et al. Formation and occurrence of new polar iodinated disinfection byproducts in drinking water. *Chemosphere*, 2016, 144: 2312–2320
- [18] De Vijlder T, Valkenburg D, Lemiere F, et al. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: the practical art of structural elucidation. *Mass Spectrometry Reviews*, 2018, 37(5): 607–629
- [19] Liu Y, D’Agostino L A, Qu G, et al. High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of poly- and perfluoroalkyl substances (PFASs) in environmental and human samples. *TrAC Trends in Analytical Chemistry*, 2019, 121: 115420
- [20] Gilca A F, Teodosiu C, Fiore S, et al. Emerging disinfection byproducts: a review on their occurrence and control in drinking water treatment processes. *Chemosphere*, 2020: 127476
- [21] Chambers M C, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology*, 2012, 30(10): 918–920
- [22] Smith C A, Want E J, O’Maille G, et al. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 2006, 78(3): 779–787
- [23] Kuhl C, Tautenhahn R, Bottcher C, et al. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry*, 2012, 84(1): 283–289
- [24] Zhang H, Zhang D, Ray K, et al. Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry. *Journal of Mass Spectrometry*, 2009, 44(7): 999–1016
- [25] Want E J, Masson P, Michopoulos F, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nature Protocols*, 2013, 8(1): 17–32
- [26] Patiny L, Borel A. ChemCalc: a building block for tomorrow’s chemical infrastructure. *Journal of Chemical Information and Modeling*, 2013, 53(5): 1223–1228
- [27] Washington J W, Rosal C G, Ulrich E M, et al. Use of carbon isotopic ratios in nontargeted analysis to screen for anthropogenic compounds in complex environmental matrices. *Journal of Chromatography A*, 2019, 1583: 73–79
- [28] Schymanski E L, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environmental Science & Technology*, 2014, 48(4): 2097–2098
- [29] Kutarna S, Tang S, Hu X, et al. Enhanced nontarget screening algorithm reveals highly abundant chlorinated azo dye compounds in house dust. *Environmental Science & Technology*, 2021, 55(8): 4729–4739
- [30] Purschke K, Vosough M, Leonhardt J, et al. Evaluation of nontarget long-term LC-HRMS time series data using multivariate statistical approaches. *Analytical Chemistry*, 2020, 92(18): 12273–12281
- [31] Yang M, Zhang X, Liang Q, et al. Application of (LC)/MS/MS precursor ion scan for evaluating the occurrence, formation and control of polar halogenated DBPs in disinfected waters: a review. *Water Research*, 2019, 158: 322–337