第 46 卷 第 5 期　　质 谱 学 报　　Vol. 46　No. 5
2025 年 9 月　　Journal of Chinese Mass Spectrometry Society　　Sep. 2025

水中卤代消毒副产物质谱数据库的设计与实现

刘璨璠[1,2]，吴庆丰[2]，毛瑞士[2,3]，李　敏[2,3]，张雨樵[1,2]，张登红[1]
（1. 西北师范大学物理与电子工程学院，甘肃 兰州　730070；2. 中国科学院近代物理研究所，甘肃 兰州　730000；
3. 中国科学院大学，北京　100049）

摘要：在饮用水消毒过程中，含氯消毒剂与水中有机物反应会生成具有毒性的卤代消毒副产物（halogenated disinfection by-products，HDBPs），对人体健康构成威胁。为实现复杂水样中 HDBPs 的非靶向筛查，本研究基于中国科学院近代物理研究所公共技术中心的高分辨四极杆飞行时间质谱仪（Q-TOF MS），开发了一款综合性的质谱数据管理系统。该系统采用 Python 开发，使用 MySQL 构建数据库，并通过 PyQt 实现图形界面，具备质谱数据的存储、管理、查询和分析功能，且设计了高效的质谱匹配算法，能够快速鉴别目标化合物，并支持多种卤代乙酸质谱数据的录入与管理。本实验通过对水样中卤代乙酸的非靶向筛查，表明所构建的数据管理系统能够实现复杂样品场景下的高效匹配（匹配相似度达 93%以上），充分验证了该系统的准确性与可靠性。

关键词：质谱仪；数据管理系统；质谱数据存储；质谱数据可视化；匹配算法
中图分类号：O657.63；TP311　　**文献标志码：**A　　**文章编号：**1004-2997(2025)05-0615-12
DOI: 10.7538/zpxb.2025.0009　　**CSTR:** 32365.14.zpxb.2025.0009

Design and Implementation of a Mass Spectrometry Database for Halogenated Disinfection By-products in Water

LIU Li-fan[1,2], WU Qing-feng[2], MAO Rui-shi[2,3], LI Min[2,3], ZHANG Yu-qiao[1,2], ZHANG Deng-hong[1]
（1. *College of Physics and Electronic Engineering*, *Northwest Normal University*, *Lanzhou* 730070, *China*;
2. *Institute of Modern Physics*, *Chinese Academy of Sciences*, *Lanzhou* 730000, *China*;
3. *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*）

Abstract: During the disinfection of drinking water, chlorine-containing disinfectants react with organic matter in the water to form a class of toxic by-products, i.e. halogenated disinfection by-products (HDBPs). These by-products are potentially carcinogenic and mutagenic, posing a serious threat to human health. In recent years, the detection and analysis of HDBPs have gradually become a research hotspot in the fields of environmental chemistry and public health with the increased concern for drinking water safety. However, due to the complexity of water samples and the diversity of HDBPs species, traditional detection methods face many limitations. Therefore, the development of an efficient and accurate non-targeted screening tool becomes the key to solve this problem. Mass spectrometry has become an important tool for the detection of HDBPs due to its high resolution and high sensitivity. And the selection of database and the design of data management system are especially critical in mass spectrometry data analysis. Existing databases are mainly divided into two

南京大学团簇离子束磁电双聚焦质量能量分选系统横向合作项目（E440552XZ0）
本文通信作者李敏，张登红

categories. One is commercial databases, such as Bruker's data analysis, which is integrated in the instrument software with a friendly interface and supports seamless connection, but with high cost, poor customization, and only applicable to specific instruments. Another is public databases, such as NIST and PubChem, which are open and widely applicable, but with uneven data quality and insufficient coverage of HDBPs. Moreover, the traditional manual comparison method is inefficient and inaccurate when using these databases, which is difficult to meet the demand of high-throughput detection. Therefore, designing a dedicated mass spectrometry data management system to improve the analysis efficiency and realize the non-targeted screening of HDBPs has become the key to current research. To address the above problems, this study developed a comprehensive mass spectrometry data management system based on a high-resolution quadrupole time-of-flight mass spectrometer (Q-TOF MS) at the Public Technology Center of the Institute of Modern Physics, Chinese Academy of Sciences. The system was developed in Python language, using MySQL to build the background database and PyQt to build the graphical interface (GUI). Based on this technical architecture, the system realizes the storage, management, query and analysis functions of mass spectrometry data. The system was designed with an efficient mass spectrometry matching algorithm, which can identify the target compounds quickly. To validate the performance of the developed system, this study experimentally screened haloacetic acid in water samples in a non-targeted manner. The experimental results showed that this system is able to achieve efficient compound matching in complex sample scenarios with a matching similarity of more than 93%, demonstrating the advantages in terms of accuracy and reliability.

Key words: mass spectrometer; data management system; mass spectral data storage; mass spectral data visualization; matching algorithm

水质安全不仅直接关系到人们的生命健康，也是环境安全、生态可持续发展中的重要一环。随着工业化和城市化进程的加快，水体污染问题日益显著[1]。在水质检测领域，质谱仪已被广泛应用于地表水、环境水、饮用水和废水的分析，能够检测极微量化合物，如农药、药物、个人护理产品以及消毒副产物（disinfection by-products, DBPs）[2-3]。在现代水处理过程中，消毒是一个不可或缺的步骤，可有效杀死或去除病原体，但也会带来某些副作用，特别是会形成 DBPs[4]，这些副产物主要是由消毒剂（如氯、臭氧）与自然水体中的有机物和无机物反应产生的。某些类型的 DBPs，如卤代乙酸（haloacetic acids, HAAs），具有致癌性和遗传毒性。因此，对饮用水中 DBPs 的监测至关重要[5]。

传统的目标分析方法需要预先获知特定化合物的存在才能提供精确的检测结果，而质谱仪可以对复杂水样进行非靶向筛查。质谱仪的高分辨率、高通量、高灵敏度和高准确性使其能够在复杂背景下检测低浓度化合物，被广泛应用于化学、生物学、医学、药学和环境科学等领域[6-10]。中国科学院近代物理研究所公共技术中心质谱实验室的四极杆飞行时间质谱仪（Q-TOF MS）搭载高灵敏度电喷雾离子源（ESI），能够精确检测相对分子质量从几十到十几万的化合物，其卓越的性能可以满足复杂样品中多种化合物的检测需求。针对样品中化合物种类繁多且难以通过人工比对实现精准鉴定的局限性，需要建立专用质谱数据库，并通过分析软件调用数据库进行匹配，从而实现复杂样品中化合物的高通量自动化鉴定。如，Kadokami 等[11]报道了一种用于鉴定样本中微污染物的新型气相色谱-质谱数据库，共记录了近 700 种化学物质的质谱信息，可高效且低成本地分析样本中微污染物；Huang 等[12]报道了一种基于质谱的皂苷快速鉴定数据库，共收录了 4 196 种皂苷，提升了皂苷分析的效率与准确性。

目前，常见的质谱数据库主要有两类：一类是购买仪器配套的数据库，一般集成在配套的分析软件上，如 Bruker 公司的 DataAnalysis、岛津

公司的 GCMSsolution、Agilent 公司的 MassHunter 等，这些数据库虽具有用户友好的界面和优质的技术支持，并与其品牌的质谱仪无缝连接，但成本较高且定制性较差，数据覆盖不足，只适用于固定品牌或型号的仪器[13]；另一类是独立于仪器、公开的公共数据库，如 NIST Mass Spectral Library、MassBank[14]、PubChem 等。Margolin Eren 等[15]测定了 46 种化合物的质谱数据，并将其与 NIST 质谱库进行比较。Elapavalore 等[16]在含有毒理学相关化学物质的混合物中提取出大量质谱数据，并上传至 MassBank质谱库，显著提高了在环境和暴露组学研究中非靶向小分子鉴定工作流程的可信度。上述数据库通常由研究机构、学术团体或政府机构维护和更新，数据来源复杂，存在数据格式不统一、数据质量参差不齐等问题[17-20]。

为解决上述问题，本工作开发了一款专用的数据管理软件，旨在有效管理和处理质谱仪生成的大量复杂数据。该软件支持导入 XML 格式的实验数据文件，在检查和解析后，对数据进行降噪、归一化等预处理工作，通过质谱匹配算法将实验质谱数据与数据库中的标准质谱数据进行比对，实现对卤代乙酸的非靶向筛查，以全面提升非靶向筛查工作的效率和质量。

1　质谱数据管理系统的设计与开发

1.1　系统需求分析

质谱仪的系统软件平台由一系列相关的软件组成，示于图 1，可分为两大类：一类是与仪器直接相关的软件，负责质谱仪的控制与数据采集；另一类是用于管理质谱数据。在这些软件中，质谱数据分析处理软件位于整个软件族的下游位置[21]。在水中卤代乙酸消毒副产物检测中，质谱技术能够提供准确的化合物信息。然而，质谱数据的获取只是检测过程的一个环节，面对复杂水样中大量的非靶向化合物，亟需一个全面的质谱数据管理系统，用于高效存储、组织、检索和分析，以提升数据的利用效率，实现对目标化合物的快速筛查和匹配，使整个质谱分析过程完整有序。

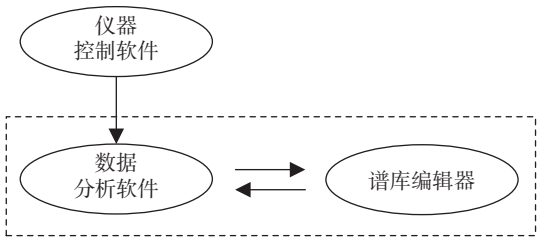

图 1　质谱仪的系统软件族

Fig. 1　System software family of mass spectrometer

在实验过程中，质谱仪会产生大量的复杂数据，如质荷比（*m*/*z*）、离子强度、碎片离子峰、同位素峰等，这些数据受离子化模式和碰撞能等参数的影响，决定了分析结果的质量和准确性。目前，自然界中已知的化合物数量超过数十万种，仅靠人工识别未知化合物的种类和结构不仅效率低，且难度大。因此，利用计算机辅助分析技术取代传统的人工鉴别方法可大幅提高鉴别的准确性和操作效率[22]。

在系统功能设计上，按照服务对象将用户分为实验人员和管理人员，收集、分析实验室在日常科研和技术应用中不同用户对质谱数据管理系统的多样化需求。实验人员的需求为注册、登录账号、修改个人信息，通过多样化的查询方式，针对性地快速检索化合物的信息和质谱图，以及对未知卤代消毒副产物成分进行匹配并导出结果。实验室管理人员负责实验数据的统筹管理和系统运行维护，除具有实验人员的需求外，还有数据格式标准化与共享、灵活更新数据库与数据扩展、确保数据的精确性、数据的长期存储与溯源、管理实验人员权限和账号信息等需求。本工作的主要目的是设计和开发一款符合上述需求、功能全面的数据管理系统。

1.2　系统总体设计

基于对系统需求的深入分析，考虑到开发效率、运行成本以及终端用户的操作便捷性，采用 Python 编程语言进行开发。此外，本系统选用轻量级的 MySQL 作为数据库管理系统，以保证数据处理的高效性和可扩展性。为进一步提升用户交互体验，使用功能强大的 PyQt 框架设计图形用户界面。质谱数据管理系统的具体功能模块示于图 2。

数据采集模块：系统支持标准格式文件（如.XML）的导入，通过脚本解析文件内容，提取 *m*/*z*、强度值和实验元数据存入数据库；管理员可通过图形界面手动输入化合物信息与实验质谱数据。该系统还支持通过 API 从 PubChem 开源库获取数据，进行格式转换后存入数据库。

数据处理模块：用户上传质谱数据，对数据

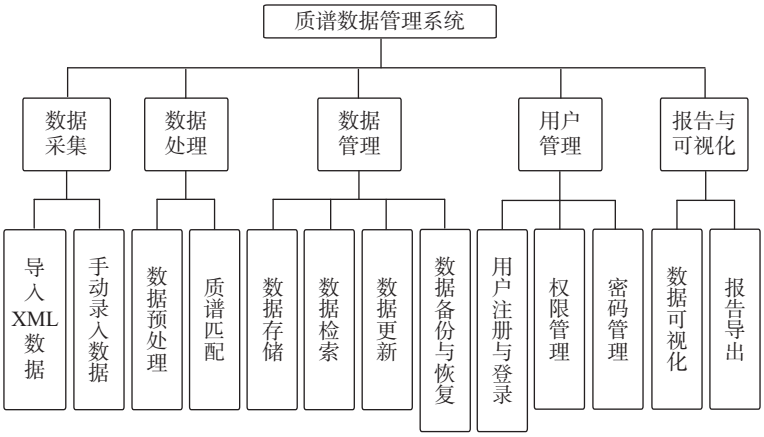

图 2 质谱数据管理系统的功能图

Fig. 2 Function diagram of mass spectrometry data management system

进行预处理；通过余弦相似度算法与数据库中存储的数据进行比对，提供匹配结果。

数据管理模块：存储处理好的数据及相关元数据；针对不同需求，提供不同化合物的多途径检索方式，如分子式、化合物名称、分子质量、CAS 编号、精确分子质量等；定期备份数据，并在必要时恢复数据。

用户管理模块：区分管理员和普通用户，提供安全的注册、登录功能，支持修改账户密码；根据用户角色设置不同的权限，确保系统的安全性和操作的规范化。管理员拥有数据增、删、改、查权限和普通用户的所有权限；普通用户仅可查询数据库中的化合物信息、上传实验数据及匹配分析。

报告和可视化模块：为用户提供数据分析结果的图表，生成分析报告并导出为 PDF 格式。

2 数据库设计

2.1 数据库 E-R 模型

在质谱数据管理系统中，核心实体包括实验记录、样品信息、化合物信息、质谱数据、实验仪器及用户，这些实体通过明确的关系相互关联。例如，数据来源、样品和用户分别与实验记录呈一对多的关系，一个数据来源可设计多个实验，质谱仪可用于生成多张谱图，谱图、化合物与匹配结果均为一对多的关系。该设计清晰地呈现了实体间的内在关系，支持复杂数据的高效管理。

从样本采集和实验操作到色谱与质谱的详细数据分析，各级数据层层依赖，确保数据库设计的规范性与完整性。通过外键约束有效减少数据冗余，同时支持复杂查询功能，如按化合物检索相关样品与标准品，或查看特定仪器生成的质谱数据。此外，该设计便于实现基于用户角色的权限控制，灵活限定数据访问范围。综上，最终构建的本系统总 E-R 图示于图 3。

2.2 数据库表设计

在明确实体及其属性后，依据实体关系设计了 MySQL 数据库表，严格遵循设计原则，定义字段属性并分析表间关联性。通过数据库逻辑设计，将概念设计具体化为表结构和逻辑操作。详细的质谱数据管理系统表分别列于表 1～4。

3 算法实现

质谱匹配算法是质谱数据管理系统的核心，用于分析实验质谱数据并与数据库中标准质谱数据进行比对，实现对未知化合物的定性分析，广泛应用于非靶向化合物的鉴定，为环境科学和化学研究提供了有力工具[23]。质谱匹配的基本原理基于相似性计算原理，通过比较实验质谱图与标准质谱图之间的相似度进行匹配。

3.1 质谱匹配流程

首先，加载未知化合物的质谱数据文件，解析并提取文件中的谱图数据、碰撞能和前体离子 m/z 等信息；然后，根据碰撞能和前体离子 m/z 设定的阈值范围，初步筛选出参考质谱图，以消除不同仪器和实验条件带来的误差，提高匹配效率和准确性。提取实验质谱图和参考质谱图的前 5 个峰值作为匹配的关键特征峰，基于余弦相似度算法，计算两者的余弦相似度评分，返回相似度大于 40%且排名前 5 的数据。质谱匹配算法流程示于图 4。

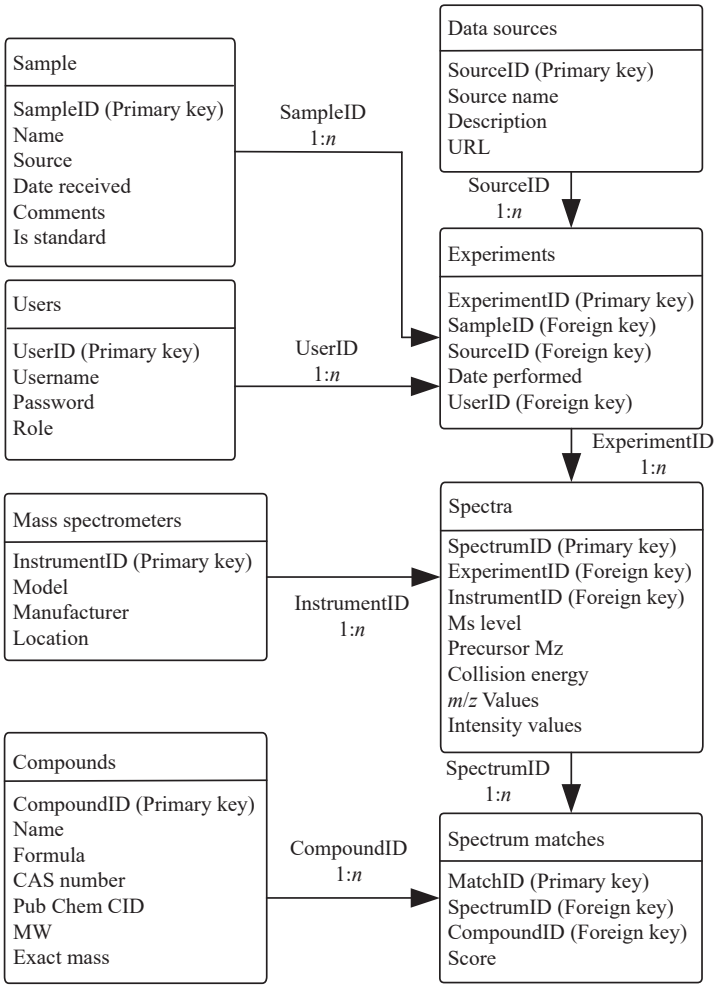

图 3　系统总 E-R 图

Fig. 3　E-R diagram of system

表 1　质谱数据表

Table 1　Data table of mass spectrometry

字段名称 Column name	字段类型 Data type	是否为空 Allow null	主键 Primary Key	外键 Foreign Key	说明 Description
SpectrumID	INT	否	是	否	质谱ID
ExperimentID	INT	否	否	是	实验ID
InstrumentID	INT	否	否	是	仪器ID
MsLevel	INT	否	否	否	质谱级别
PrecursorMz	FLOUAT	是	否	否	前体离子质荷比
CollisionEnergy	FLOUAT	是	否	否	碰撞能
m/*z*Values	VARCHAR	否	否	否	质荷比列表
IntensityValues	VARCHAR	否	否	否	强度列表

3.2　数据预处理

匹配前，需要对质谱数据进行预处理，以提高数据质量、减少噪音、增强算法的准确性和效率[5]。

3.2.1　去噪处理　设置强度阈值、移除低强度噪声峰，根据 *m*/*z* 分布的统计特性剔除异常峰。

3.2.2　归一化处理　原始的质谱数据强度为绝对强度，为确保比较的一致性，通常需要对质谱

表 2 化合物表

Table 2 Table of compounds

字段名称 Column name	字段类型 Data type	是否为空 Allow null	主键 Primary Key	外键 Foreign Key	说明 Description
CompoundID	INT	否	是	否	化合物ID
Name	VARCHAR	否	否	否	化合物名称
Formula	VARCHAR	否	否	否	化合物化学式
CASNumber	VARCHAR	否	否	否	CAS号
PubChemCID	VARCHAR	否	否	否	PubChem的CID
MW	DECIMAL	否	否	否	相对分子质量
ExactMass	DECIMAL	否	否	否	精确分子质量

表 3 样品表

Table 3 Table of samples

字段名称 Column name	字段类型 Data type	是否为空 Allow null	主键 Primary Key	外键 Foreign Key	说明 Description
SampleID	INT	否	是	否	样品ID
Name	VARCHAR	否	否	否	样品名称
Source	VARCHAR	否	否	否	样品来源
DateReceived	DATE	否	否	否	接收日期
Comments	TEXT	是	否	否	注释
IsStandard	BOOLEAN	否	否	否	是否为标准品

表 4 质谱仪表

Table 4 Table of mass spectrometer

字段名称 Column name	字段类型 Data type	是否为空 Allow null	主键 Primary Key	外键 Foreign Key	说明 Description
InstrumentID	INT	否	是	否	仪器ID
Model	VARCHAR	否	否	否	仪器型号
Manufacturer	VARCHAR	否	否	否	制造商
Location	VARCHAR	否	否	否	位置

强度进行归一化。本文采用在简单要素缩放的方法上进行相对强度归一化，使其相对强度范围为 0%～100%，强度值具有可比性，以消除绝对强度差异的影响，从而只需比较形状而非绝对强度值[24]。归一化公式示于式(1)：

$$I_N = \frac{I_O}{I_{max}} \times 100\% \quad (1)$$

式中，I_O 为质谱图的绝对强度值，I_{max} 为质谱图的绝对强度最大值。为使质谱图的相对强度归一化结果 I_N 在 0%～100%范围内，对简单要素缩放的结果乘 100%。

3.2.3 峰选择 选取特征性强、具有代表性的前 5 个信号峰进行匹配。

3.3 余弦相似度匹配算法

为提高谱库检索的效率，研究人员提出多种相似性算法，如果想要得到 2 个质谱图之间的评价指标，可以将它们表示为向量，然后计算相似性系数[25]。例如，南开大学律祥俊等[26]提出了内积相似性算法；吉林大学扈庆[22]在此基础上进一步改进，提出了一种新的算法。此外，天津大学宋爽[27]提出了一种基于 P 范数的相似性算法；李宏彬等[28]提出了皮尔逊相似性算法。尽管上述算法在一定程度上提高了质谱检索效率，但在处理高维向量的稀疏性、权重分布不均等问题时仍存在局限性。

余弦相似度是一种衡量 2 个向量方向相似度的度量，在质谱匹配中，每个质谱图可以被视为 1 个向量，其中每个维度对应 1 个特定的 *m*/*z*，该维度值是该 *m*/*z* 处的信号强度。对 2 个向量的夹角余弦值进行计算，得到的值即为余弦相似

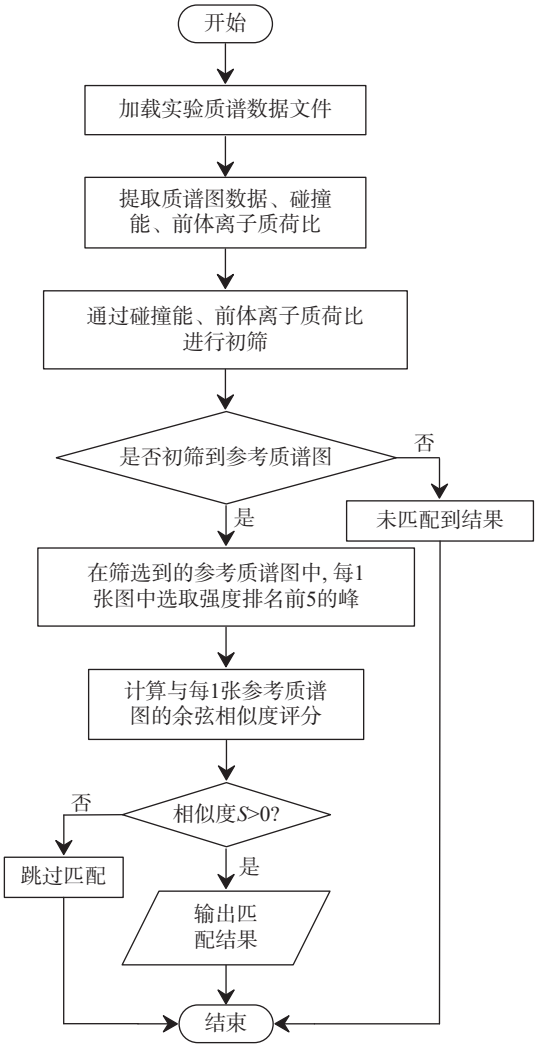

图 4　质谱匹配流程图

Fig. 4　Matching flowchart of mass spectrometry

度值，其范围为 0～1。该值越接近 1，表示 2 个质谱图越相似；越接近 0，表示越不相似[29]。通过计算未知质谱图与已知质谱图的余弦相似度值，可以找到匹配度最高的化合物，从而推测未知样本的可能结构。

余弦相似度 S 的计算公式示于式(2)：

$$S=\frac{\sum_i (I_{1i}\cdot I_{2i})}{\sqrt{\sum_i I_{1i}^2}\cdot\sqrt{\sum_i I_{2i}^2}} \quad (2)$$

式中，I_{1i} 和 I_{2i} 分别是已知质谱图和未知质谱图第 i 个匹配峰的强度。

余弦相似度算法具有较强的鲁棒性，对峰强度的绝对值变化不敏感，能够有效应对因实验条件差异导致的强度波动问题。同时，该算法效率高，通过稀疏向量化表示和优化预处理，大幅降低了计算复杂度，提升了匹配效率，适合在大规模数据库中进行快速比对。此外，该算法适用性广泛，不仅适用于目标化合物筛选，还能用于复杂样品的非靶向分析，尤其在快速定位特定化合物方面表现出色。

在余弦相似度的基础上，还可为不同 m/z 值的峰赋予不同的权重，增强重要峰对相似度的影响[30]。但由于权重的设定具有一定的主观性，可能会导致结果出现偏差，故本研究仅采用余弦相似度算法。

3.4　算例分析

为对比化合物 A 和 B 的质谱特征差异，选取二者的质谱图示于图 5。首先，对这 2 种化合物的原始质谱数据进行预处理；然后，提取质谱图中具有较强特征性和代表性的信号峰作为分析数据的核心特征；随后，将提取的信号峰数据导入余弦相似度算法程序中；最终得到化合物 A 与 B 之间的余弦相似度为 93.9%。

值得注意的是，化合物 A 和 B 的质谱图分别为同一物质在 2 次质谱实验中获得的结果。使用匹配算法对这 2 组数据进行分析得出高相似度值，充分验证了余弦相似度算法的有效性和可靠性。此外，为进一步评估该算法在更广泛数据集中的表现，对大量质谱实验数据进行验证。结果表明，无论数据复杂度如何变化，该算法在各种数据集中的匹配准确率均较高，能够满足复杂质

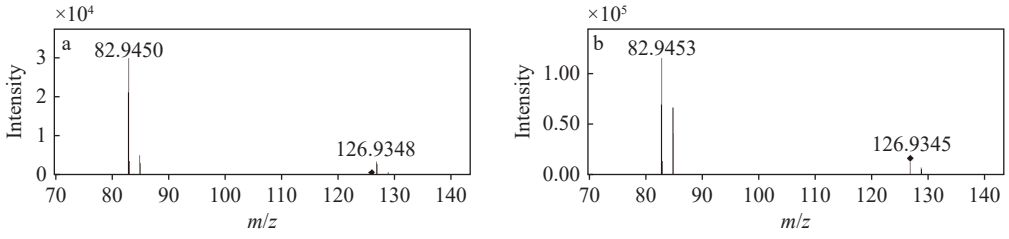

图 5　化合物 A(a)、B(b)的质谱图

Fig. 5　Mass spectra of compounds A (a) and B (b)

谱数据分析的要求。由于本工作基于高分辨质谱仪，并且采用同一仪器对标准品和样品进行实验，数据差异性不大，因此，将余弦相似度算法选定为质谱数据库中数据匹配过程的核心算法。

4 实验与验证

卤代乙酸是饮用水消毒过程中产生的一类消毒副产物，其对人体健康的影响已受到广泛关注。色谱法的分离能力强、选择性高、能够同时测定多种物质，在卤代乙酸测定中得到广泛应用[31]。为实现对水中卤代乙酸的高效筛查，本研究采用高分辨质谱技术结合液相色谱分离，该方法能够提供复杂样品中多种化合物的详细分析，且具有高灵敏度和高选择性[32]。

4.1 主要仪器与装置

Bruker maXis Plus 高分辨率四极杆飞行时间质谱仪：德国 Bruker 公司产品，配有 ESI 源和大气压化学电离源（APCI），并搭载 Compass 和 DataAnalysis 数据处理系统；Ultimate 3000 超高效液相色谱仪：美国 Thermo Fisher 公司产品，配有二极管阵列检测器（DAD）和自动进样器，搭载 Chromeleon 色谱数据处理系统；UPHW 系列超纯水系统：中国西安优普仪器设备有限公司产品；Sartorius BSA224S 电子天平（最大称量 220 g，精度 0.1 mg）：德国 Sartorius 公司产品；移液枪：德国 Eppendorf 公司产品。

4.2 主要材料与试剂

一氯乙酸（MCAA）、一溴乙酸（MBAA）、二氯乙酸（DCAA）、三氯乙酸（TCAA）、溴氯代乙酸（BCAA）、一溴二氯乙酸（BDCAA）、二溴乙酸（DBAA）、二溴一氯乙酸（CDBAA）、三溴乙酸（TBAA）等甲醇中 9 种卤代乙酸混标（1 000 mg/L），抗坏血酸（含量＞99%）：河南标准物质研发中心产品；甲酸、甲醇：均为质谱级，美国 Thermo Fisher 公司产品；超纯水：由 UPHW 系列超纯水系统制备。

4.3 实验条件

4.3.1 色谱条件 Waters ACQUITY UPLC BEH C18 色谱柱（2.1 mm×50 mm, 1.7 μm）；流动相：A 相为含 0.1%甲酸（*V*/*V*）的水溶液，B 相为甲醇；梯度洗脱程序：0～3 min（5%～40%B），3～5 min（40%～46%B），5～12 min（90%B），然后用 5%A 平衡 5 min；流速 200 μL/min；柱温 10 ℃；进样量 10 μL。

4.3.2 质谱条件 ESI 源，负离子扫描方式，多反应监测模式（MRM），干燥气流速 8.0 L/min，离子源温度 220 ℃，质量扫描范围 *m*/*z* 50～2 200。

4.4 样品采集、保存与前处理

首先，先打开自来水水龙头放水 5 min，再用 100 mL 带聚四氟乙烯衬垫的棕色玻璃瓶进行采样，添加 0.70 g/L 抗坏血酸作为脱氯剂，密封避光冷藏，保存时间为一周。由于自来水 HAAs 本底值较低，故加标浓度按照 10 μg/L 进行分析[31]。检测前过 0.45 μm 微孔滤膜，分别获取标准样品的一级和二级质谱图，得到准分子离子峰和碎片离子峰信息[19]。

4.5 质谱数据库建立与系统应用

利用 BRUKER maXis Plus 超高分辨飞行时间质谱仪采集卤代乙酸标准品的多级质谱数据，采集过程示于图 6。

在 Compass Data Analysis 软件中选取 1 个色谱峰，导出其二级质谱图，并将该化合物的其他信息，如名称、分子式、CAS 编号、相对分子质量、精确分子质量等存入数据库中，建立目标化合物卤代乙酸的标准质谱数据库[33]。以 2 种最重要的卤代乙酸（DCAA 和 TCAA）为例，其相关信息列于表 5。

采用液相色谱-质谱联用技术对样品进行分离与检测，并利用上述建立的卤代乙酸质谱数据库进行质谱匹配，实现对样品中未知化合物的定性分析。匹配参数包括：前体离子 *m*/*z* 的质量误差＜5×10^{-6}，碎片离子 *m*/*z* 的质量误差＜0.02 u，余弦相似度＞0.4+。

在 ESI⁻模式下，固定液相色谱条件并进样分析，首先进行母离子扫描，根据离子信号强度选出 9 种卤代乙酸的目标母离子，经 *m*/*z* 推测 9 种母离子形式均为[M−H]⁻；然后对母离子施加一定的碰撞能量获得碎片离子，其中 DCAA 和 TCAA 的子离子碎片分别为 *m*/*z* 82.9 和 116.9[34]。利用卤代乙酸的标准质谱数据库对采集的加标样品进行筛查鉴定，部分筛查结果列于表 6。可知，实测值与理论值的质量误差较小，余弦相似度接近 100%，表示它们高度相似，均符合匹配参数，表明该样品中含有 DCAA 和 TCAA。

完成匹配后，系统支持将匹配结果通过报告形式导出，方便实验人员直观查看匹配情况。经系统匹配后，导出的实验报告包含报告导出时间、仪器名称及匹配度大于 40%的 5 个匹配结果。本实验导出的报告示于图 7，可知，该化合

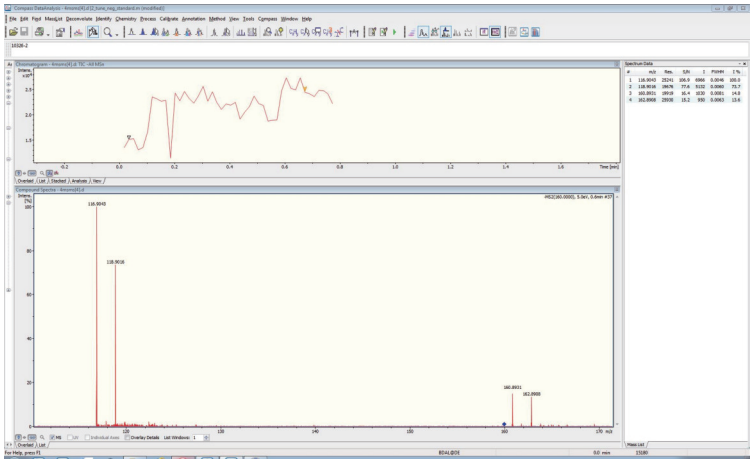

图 6　标准品质谱采集过程

Fig. 6　Acquisition process of standard mass spectra

表 5　二氯乙酸、三氯乙酸数据库数据

Table 5　Database data of dichloroacetic acid and trichloroacetic acid

序号 No.	化合物名称 Compound name	分子式 Formula	离子模式 Ion mode	准分子离子 Quasi-molecular ion (*m*/*z*)	碎片离子 Fragment ion (*m*/*z*)
1	二氯乙酸	$C_2HCl_3O_2$	$[M-H]^-$	126.9348	82.9450, 84.9419
2	三氯乙酸	$C_2H_2Cl_2O_2$	$[M-H]^-$	160.8931	116.9043, 118.9016

表 6　二氯乙酸、三氯乙酸的筛查结果

Table 6　Screening results of dichloroacetic acid and trichloroacetic acid

化合物名称 Compound name	理论值 Theoretical value (*m*/*z*)	实测值 Measured value (*m*/*z*)	质量误差 Mass error	实测碎片离子 Measured fragment ion (*m*/*z*)	余弦相似度 Cosine similarity/%
二氯乙酸	126.9348	126.9345	0.0003	82.9453、84.9424	98.0
三氯乙酸	160.8931	160.8949	0.0018	116.9056、118.9026	99.2

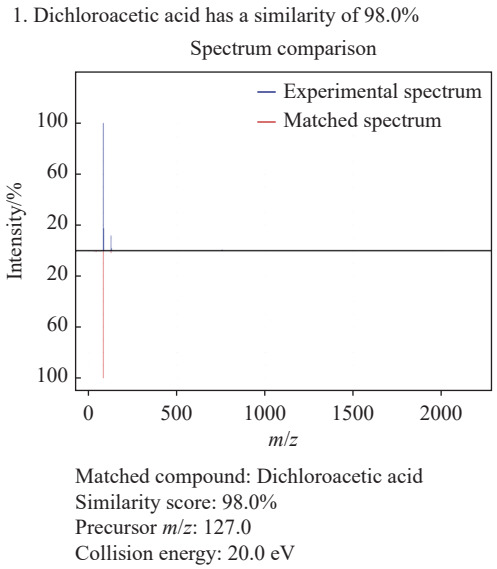

图 7　报告中匹配结果 1

Fig. 7　Matching result 1 in the report

物为二氯乙酸，相似度 98.0%，表明该系统能够较好地进行未知化合物的匹配检索，其结果及准确性与预期相符。

5　结论

本工作针对水中卤代消毒副产物的分析需求，通过整合开源数据库与实验室测得的标准品质谱数据，构建了一款针对高分辨质谱数据的综合型质谱数据管理系统，涵盖数据存储、查询、匹配和结果展示等功能。在系统开发中，重点解决了数据存储和匹配算法的准确率问题。实验验证表明，该系统在卤代乙酸非靶向筛查中表现优异，能够准确识别复杂水样中的目标化合物。另外，通过引入余弦相似度算法，显著提高了匹配效率，为非靶向筛查技术提供支持。

未来，本系统将通过扩展数据库覆盖范围，采集更多实验数据来增强目标化合物的筛查能

力，同时规范数据库格式，兼容 mzML 等通用存储格式，还将优化用户界面功能，提高查询效率并改进匹配结果的可视化表现。

参考文献：

[1] 刘洁. 环境水质分析中重金属检测技术的应用探析[J]. 黑龙江环境通报, 2025, 38(2): 133-135.
LIU Jie. Application of heavy metal detection technology in environmental water quality analysis[J]. Heilongjiang Environmental Journal, 2025, 38(2): 133-135 (in Chinese).

[2] 李颖. 试论城市生活饮用水的水质检测与分析[J]. 化工管理, 2017(2): 250.
LI Ying. Discussion on water quality detection and analysis of urban drinking water[J]. Chemical Enterprise Management, 2017(2): 250(in Chinese).

[3] 梁锦汉, 刘则华, 刘婉琼, 董燕珊. 固相萃取-超高效液相色谱-串联三重四极杆质谱法测定水体中 18 种氨基酸[J]. 质谱学报, 2025, 46(2): 242-253.
LIANG Jinhan, LIU Zehua, LIU Wanqiong, DONG Yanshan. Determination of 18 amino acids in water by solid phase extraction-ultra performance liquid chromatography-tandem triple quadrupole mass spectrometry[J]. Journal of Chinese Mass Spectrometry Society, 2025, 46(2): 242-253(in Chinese).

[4] 王栋, 贾瑞宝, 孙韶华, 宋艳, 王明泉, 赵清华, 王锐敏. 在线液液萃取-气相色谱-质谱法测定生活饮用水中卤代乙腈[J]. 中国环境监测, 2022, 38(5): 175-181.
WANG Dong, JIA Ruibao, SUN Shaohua, SONG Yan, WANG Mingquan, ZHAO Qinghua, WANG Ruimin. Determination of haloacetonitriles in drinking water using automated liquid-liquid extraction-gas chromatography-mass spectrometry[J]. Environmental Monitoring in China, 2022, 38(5): 175-181(in Chinese).

[5] 梁语韬. 自来水中卤代消毒副产物的液相色谱高分辨质谱非靶向分析[D]. 东莞: 东莞理工学院, 2023.

[6] 王一茹. 高分辨质谱在食品农药残留检测中的应用与展望[J]. 中国食品, 2023(20): 92-94.
WANG Yiru. Application and prospect of high resolution mass spectrometry in the detection of pesticide residues in food[J]. China Food, 2023(20): 92-94(in Chinese).

[7] SAUER S, KLIEM M. Mass spectrometry tools for the classification and identification of bacteria[J]. Nature Reviews Microbiology, 2010, 8(1): 74-82.

[8] 许中石, 陈涛, 江柯成, 姚伟宣, 王继业, 吴元钊, 王斌杰, 李国军. 大气压固体分析探针结合单四极杆质谱仪快速检测 18 种合成大麻素[J]. 质谱学报, 2024, 45(2): 292-300.
XU Zhongshi, CHEN Tao, JIANG Kecheng, YAO Weixuan, WANG Jiye, WU Yuanzhao, WANG Binjie, LI Guojun. Rapid screening of 18 synthetic cannabinoids using atmospheric pressure solids analysis probe coupled with single-quadrupole mass spectrometer[J]. Journal of Chinese Mass Spectrometry Society, 2024, 45(2): 292-300 (in Chinese).

[9] PEREZ-RIVEROL Y, BAI J, BANDLA C, GARCÍA-SEISDEDOS D, HEWAPATHIRANA S, KAMATCHINATHAN S, KUNDU D J, PRAKASH A, FRERICKS-ZIPPER A, EISENACHER M, WALZER M, WANG S, BRAZMA A, VIZCAÍNO J A. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences[J]. Nucleic Acids Research, 2022, 50(D1): D543-D552.

[10] HALDER A, VERMA A, BISWAS D, SRIVASTAVA S. Recent advances in mass-spectrometry based proteomics software, tools and databases[J]. Drug Discovery Today: Technologies, 2021, 39: 69-79.

[11] KADOKAMI K, TANADA K, TANEDA K, NAKAGAWA K. Novel gas chromatography-mass spectrometry database for automatic identification and quantification of micropollutants[J]. Journal of Chromatography A, 2005, 1 089(1/2): 219-226.

[12] HUANG F Q, DONG X, YIN X, FAN Y, FAN Y, MAO C, ZHOU W. A mass spectrometry database for identification of saponins in plants[J]. Journal of Chromatography A, 2020, 1 625: 461 296.

[13] 周义. 质谱数据处理算法的研究与应用设计[D]. 宁波: 宁波大学, 2017.

[14] HORAI H, ARITA M, KANAYA S, NIHEI Y, IKEDA T, SUWA K, OJIMA Y, TANAKA K, TANAKA S, AOSHIMA K, ODA Y, KAKAZU Y, KUSANO M, TOHGE T, MATSUDA F, SAWADA Y, HIRAI M Y, NAKANISHI H, IKEDA K, AKIMOTO N, MAOKA T, TAKAHASHI H, ARA T, SAKURAI N, SUZUKI H, SHIBATA D, NEUMANN S, IIDA T, TANAKA K, FUNATSU K, MATSUURA F, SOGA T, TAGUCHI R, SAITO K, NISHIOKA T. MassBank: a public repository for sharing mass spectral data for life sciences[J]. Journal of Mass Spectrometry, 2010, 45(7): 703-714.

[15] MARGOLIN EREN K J, ELKABETS O, AMIRAV A. A comparison of electron ionization mass spectra obtained at 70 eV, low electron energies, and with cold

EI and their NIST library identification probabilities[J]. Journal of Mass Spectrometry, 2020, 55(12): e4646.

[16] ELAPAVALORE A, KONDIĆ T, SINGH R R, SHOEMAKER B A, THIESSEN P A, ZHANG J, BOLTON E E, SCHYMANSKI E L. Adding open spectral data to MassBank and PubChem using open source tools to support non-targeted exposomics of mixtures[J]. Environmental Science Processes & Impacts, 2023, 25(11): 1 788-1 801.

[17] 张加余, 屠鹏飞. 天然产物液相色谱-质谱-数据库(LC-MS-DS)的建立与应用[J]. 药学学报, 2012, 47(9): 1 187-1 192.
ZHANG Jiayu, TU Pengfei. Construction and application of natural products LC-MS-DS[J]. Acta Pharmaceutica Sinica, 2012, 47(9): 1 187-1 192(in Chinese).

[18] 王志斌. 农药高分辨质谱数据库的建立及应用研究[D]. 秦皇岛: 燕山大学, 2017.

[19] 田宏哲, 周艳明, 刘文娥. 农产品中 50 余种农药 LC-MS/MS 质谱数据库的建立及应用[J]. 食品科学, 2010, 31(4): 218-222.
TIAN Hongzhe, ZHOU Yanming, LIU Wene. Construction and application of liquid chromatography-tandem mass spectral database of pesticides in agricultural products[J]. Food Science, 2010, 31(4): 218-222(in Chinese).

[20] 赵佳宇, 吕悦广, 辛通, 郭倩倩, 张文赢, 李金鸿, 薛宏宇, 马强. 中药复方九味汤化学成分鉴定及质谱裂解规律研究[J]. 质谱学报, 2025, 46(1): 11-25.
ZHAO Jiayu, LYU Yueguang, XIN Tong, GUO Qianqian, ZHANG Wenying, LI Jinhong, XUE Hongyu, MA Qiang. Identification and mass spectrometric fragmentation pathways of chemical components in the traditional Chinese medicine formula of Jiuwei decoction[J]. Journal of Chinese Mass Spectrometry Society, 2025, 46(1): 11-25(in Chinese).

[21] 张宝林. 质谱仪软件平台的研究与开发[D]. 长春: 吉林大学, 2006.

[22] 扈庆. 分析仪器数据格式及质谱检索系统的研究与应用[D]. 长春: 吉林大学, 2006.

[23] 骆瑜, 刘志斌. 基于 NIST 数据库的农药残留裂解谱库的建立及应用[J]. 食品安全导刊, 2019(21): 161-162, 164.
LUO Yu, LIU Zhibin. Establishment and application of pesticide residue pyrolysis spectrum library based on NIST database[J]. China Food Safety Magazine, 2019(21): 161-162, 164(in Chinese).

[24] 张海强. 基于深度神经网络与信息熵的蛋白质质谱数据分析模型[D]. 昆明: 昆明理工大学, 2023.

[25] 周义, 俞建成, 张俊良, 吴焕铭. 一种基于新的向量空间模型的谱库检索算法[J]. 真空科学与技术学报, 2016, 36(12): 1 450-1 454.
ZHOU Yi, YU Jiancheng, ZHANG Junliang, WU Huanming. Novel vector space model and algorithm for search of mass spectral library[J]. Chinese Journal of Vacuum Science and Technology, 2016, 36(12): 1 450-1 454(in Chinese).

[26] 律祥俊, 林少凡, 张金碚, 张法义. 一种有机质谱谱图的库检索新算法[J]. 高等学校化学学报, 1994, 15(5): 678-680.
LÜ Xiangjun, LIN Shaofan, ZHANG Jinbei, ZHANG Fayi. A new algorithm for library searching of organic mass spectrum[J]. Chemical Journal of Chinese Universities, 1994, 15(5): 678-680(in Chinese).

[27] 宋爽. 气相色谱-质谱联用仪的纯净谱图提取与检索算法的研究[D]. 天津: 天津大学, 2012.

[28] 李宏彬, 赫光中, 果秋婷. 基于皮尔逊相关系数的有机质谱相似性检索方法[J]. 化学分析计量, 2015, 24(3): 33-37.
LI Hongbin, HE Guangzhong, GUO Qiuting. Similarity retrieval method of organic mass spectrometry based on the Pearson correlation coeffi cient[J]. Chemical Analysis and Meterage, 2015, 24(3): 33-37(in Chinese).

[29] WATROUS J, ROACH P, ALEXANDROV T, HEATH B S, YANG J Y, KERSTEN R D, van der VOORT M, POGLIANO K, GROSS H, RAAIJMAKERS J M, MOORE B S, LASKIN J, BANDEIRA N, DORRESTEIN P C. Mass spectral molecular networking of living microbial colonies[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(26): E1743-E1752.

[30] WANG M, CARVER J J, PHELAN V V, SANCHEZ L M, GARG N, PENG Y, NGUYEN D D, WATROUS J, KAPONO C A, LUZZATTO-KNAAN T, PORTO C, BOUSLIMANI A, MELNIK A V, MEEHAN M J, LIU W T, CRÜSEMANN M, BOUDREAU P D, ESQUENAZI E, SANDOVAL-CALDERÓN M, KERSTEN R D, PACE L A, QUINN R A, DUNCAN K R, HSU C C, FLOROS D J, GAVILAN R G, KLEIGREWE K, NORTHEN T, DUTTON R J, PARROT D, CARLSON E E, AIGLE B, MICHELSEN C F, JELSBAK L, SOHLENKAMP C, PEVZNER P, EDLUND A, McLEAN J, PIEL J, MURPHY B T, GERWICK L, LIAW C C, YANG Y L, HUMPF H U, MAANSSON M, KEYZERS R A, SIMS A C, JOHNSON A R, SIDEBOTTOM A M, SEDIO B E, KLITGAARD A, LARSON C B, TORRES-MENDOZA D, GONZALEZ D J, SILVA

D B, MARQUES L M, DEMARQUE D P, POCIUTE E, O'NEILL E C, BRIAND E, HELFRICH E J N, GRANATOSKY E A, GLUKHOV E, RYFFEL F, HOUSON H, MOHIMANI H, KHARBUSH J J, ZENG Y, VORHOLT J A, KURITA K L, CHARUSANTI P, McPHAIL K L, NIELSEN K F, VUONG L, ELFEKI M, TRAXLER M F, ENGENE N, KOYAMA N, VINING O B, BARIC R, SILVA R R, MASCUCH S J, TOMASI S, JENKINS S, MACHERLA V, HOFFMAN T, AGARWAL V, WILLIAMS P G, DAI J, NEUPANE R, GURR J, RODRÍGUEZ A M C, LAMSA A, ZHANG C, DORRESTEIN K, DUGGAN B M, ALMALITI J, ALLARD P M, PHAPALE P, NOTHIAS L F, ALEXANDROV T, LITAUDON M, WOLFENDER J L, KYLE J E, METZ T O, PERYEA T, NGUYEN D T, VanLEER D, SHINN P, JADHAV A, MÜLLER R, WATERS K M, SHI W, LIU X, ZHANG L, KNIGHT R, JENSEN P R, PALSSON B O, POGLIANO K, LININGTON R G, GUTIÉRREZ M, LOPES N P, GERWICK W H, MOORE B S, DORRESTEIN P C, BANDEIRA N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking[J]. Nature Biotechnology, 2016, 34(8): 828-837.

[31] 赵璇, 骆春迎, 张静, 杨觅, 罗新月, 赵心悦, 王炼, 邹晓莉. 反相超高效液相色谱-高分辨质谱法测定自来水消毒副产物卤代乙酸[J]. 四川大学学报(医学版), 2022, 53(3): 504-510.
ZHAO Xuan, LUO Chunying, ZHANG Jing, YANG Mi, LUO Xinyue, ZHAO Xinyue, WANG Lian, ZOU Xiaoli. Determination of haloacetic acids, disinfection byproducts, in tap water with reversed-phase ultra-performance liquid chromatography-high resolution mass spectrometry[J]. Journal of Sichuan University (Medical Sciences), 2022, 53(3): 504-510(in Chinese).

[32] 唐晓琴, 赵舰, 贺丽迎, 甘源, 周春艳, 程莉, 覃梅. 72 种生物碱高分辨质谱数据库建立与应用[J]. 中国食品卫生杂志, 2020, 32(3): 228-233.
TANG Xiaoqin, ZHAO Jian, HE Liying, GAN Yuan, ZHOU Chunyan, CHENG Li, QIN Mei. Establishment and application of 72 alkaloids database with high resolution mass spectrometry[J]. Chinese Journal of Food Hygiene, 2020, 32(3): 228-233(in Chinese).

[33] 寿晨超, 娜仁高娃, 高素芸, 刘剑, 赵丰. 天然染料质谱数据库的建立与应用[J]. 纺织学报, 2023, 44(11): 120-131.
SHOU Chenchao, NAREN Gaowa, GAO Suyun, LIU Jian, ZHAO Feng. Establishment and application of mass spectral database for natural dyes[J]. Journal of Textile Research, 2023, 44(11): 120-131(in Chinese).

[34] 雷颖, 八十岛诚, 王凌云, 范小江, 陶益, 张锡辉. 液相色谱-串联质谱法同时检测自来水中 9 种卤乙酸[J]. 中国给水排水, 2013, 29(20): 124-129.
LEI Ying, YASOJIMA Makoto, WANG Lingyun, FAN Xiaojiang, TAO Yi, ZHANG Xihui. Determination of nine haloacetic acids in tap water by liquid chromatography-tandem mass spectrometry[J]. China Water & Wastewater, 2013, 29(20): 124-129(in Chinese).

(收稿日期: 2025-01-23; 修回日期: 2025-03-19)