

# 基于碎裂模式筛查转化产物的高分辨液质方法

王麒麟, 王联红

(南京大学环境学院污染控制与资源化研究国家重点实验室, 南京 210023)

**摘要:** 合成化学品在环境中广泛检出, 在环境条件下发生转化, 可能产生毒性更强的化合物。转化产物的识别已成为制约合成化学品生态毒性评估的瓶颈之一。围绕如何鉴定磺胺嘧啶的微生物转化产物, 利用 ProteoWizard 软件将样品预处理和质谱检测获得的原始数据转化为 mzML 通用格式文件, 在 python 和 R 编译器中提取分子离子和碎片离子质量数的关键数据, 基于母体与转化产物相似的碎裂模式, 在 R 编译器中实现转化产物的筛查和可视化, 鉴定出 7 种转化产物。该方法突破了质谱型号和数据格式的屏障, 有助于发展通用的转化产物筛查方法, 助力环境分析研究。

**关键词:** 转化产物筛查; 高分辨液相色谱质谱联用; 环境分析; 开源软件

中图分类号: X592

文献标识码: B

## High resolution liquid chromatography-mass spectrometry method for screening transformed products based on fragmentation patterns

Wang Qilin, Wang Lianhong

(State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, China)

**Abstract:** Synthetic chemicals are widely detected in the environment and undergo transformation under environmental conditions, which may produce more toxic compounds. The identification of conversion products has become one of the bottlenecks restricting the evaluation of ecotoxicity of synthetic chemicals. Focusing on how to identify the microbial transformation products of sulfadiazine, ProteoWizard software was used to convert the original data obtained from sample pretreatment and mass spectrometry into mzML general format files, and key data of molecular ion and fragment ion mass numbers were extracted from python and R compilers. Based on the fragmentation mode similar to that of the parent and transformation products, The transformation products were screened and visualized in the R compiler, and 7 kinds of transformation products were identified. This method breaks through the barrier of mass spectrum type and data format,

and contributes to the development of universal screening methods for transformed products, and assists environmental analysis research.

**Key words:** Screening of conversion products; High resolution liquid chromatography mass spectrometry; Environmental analysis; Open source software

## 1 前言

各种过程会转化环境中的合成化学品（杀虫剂，药物，杀菌剂和工业原料等），因此环境暴露于母体化合物和由此产生的转化产物的混合物中。一些监管计划需要转化产物对人类和环境健康影响的信息，其中识别潜在关注物质，理解转化产物的环境行为及其对生物体的影响至关重要(Boxall et al., 2004)。但许多情况下，这些信息是缺失的。转化产物的识别是制约全面评估合成化学品健康与生态风险的瓶颈之一。

四极杆飞行时间串联质谱法或离子阱轨道阱串联质谱法是两种能够快速灵敏和可靠地检测和鉴定低分子量物质的技术。并且由于初级转化产物的碎裂模式与母体化合物往往相似，基于碎裂模式可以对样品高分辨质谱数据中的未知转化产物进行挖掘(Kern et al., 2009)。但是各质谱仪器的数据格式和分析软件存在差异，基于特定型号质谱仪开发的转化产物筛查方法缺乏跨平台移植性。通用的转化产物筛查方法已经成为环境分析研究的迫切需求。

磺胺类抗菌药是最早的合成抗菌药，已广泛应用于临床和兽医。磺胺嘧啶是使用量最大的磺胺类抗菌药之一，在环境中广泛检出(Sun et al., 2015)，是需要优先关注的污染物(Yang et al., 2022)。这项研究使用四极杆飞行时间串联质谱（AB SCIEX TripleTOF5600），以磺胺嘧啶微生物转化样品为例，利用 R 和 python 语言作为分析工具以兼容多种质谱数据格式，介绍利用碎裂模式筛查高分辨质谱数据中转化产物信息的通用方法。

## 2 实验方法

### 2.1 样品处理与质谱检测

样品处理与质谱检测的流程如图 1 所示。这项研究选择土壤微生物富集液作为反应体系，投加 10 毫克每升磺胺嘧啶，并监测体系内磺胺嘧啶浓度变化。在监测到磺胺嘧啶浓度降低到初始浓度的 80%，50%，30%时，分别取样，通过冷冻干燥除水，重新溶解在甲醇中，并氮吹浓缩。随后通过液相色谱 C18 分离柱将潜在代谢产物根据亲水性强弱分离成不同组分，并再次氮吹浓缩。如潜在代谢产物保留时间接近，可以更改流动相，对馏分再次分离浓缩。

根据母体化合物基团及潜在转化产物的预测结构，选择合适的质谱参数。磺胺嘧啶经前期预实验，选用 ESI+ 模式，去簇电压 100，碰撞能 35，碰撞能范围 15，IDA 模式下检测 50-500 道尔顿范围的单电荷离子。在该参数条件下，质谱运行稳定，结果具有良好的重现性。

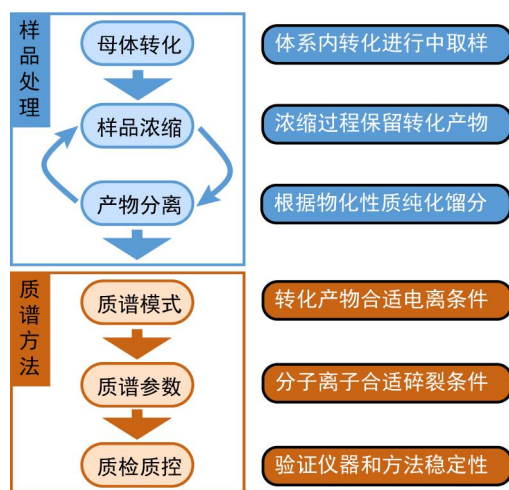


图 1 采样，样品预处理与质谱检测

## 2.2 数据处理

质谱原始数据的处理如图 2 所示。原始数据备份并经过 MD5 检验后妥善保存于离线硬盘中。由于各质谱厂家的数据格式不同，为了统一下游处理流程，使用 ProteoWizard 软件将格式转换为统一格式 mzML，保存为 input.mzML (Chambers et al., 2012)。该格式可直接在 python 中以文本格式读入或通过 mzR 编辑包读入 R 中。

使用 python 编写脚本从 input.mzML 文件中提取一级质谱数据，包括文件内所有分子离子的编号，质荷比与保留时间，导出为 precursor.csv 文件。随后，在 R 或 Rstudio 编译器内，调用 mzR 编辑包 peaks 函数逐个读取分子离子的二级质谱数据，包括碎片离子质荷比及信号强度（信号强度可归一化为百分比）(Chambers et al., 2012)。由于二级质谱采集数据中的背景噪声和微弱碎片信号并不能提供信息用于下游分析中的产物筛查和结构鉴定，反而造成干扰并增加工作量，因此归一化碎片离子信号强度后，删除低于 10% 的碎片离子数据，以减少计算量。

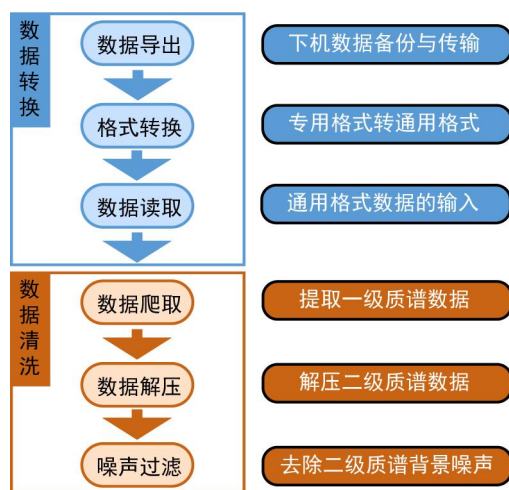


图 2. 质谱数据的处理流程

## 3 结果

### 3.1 转化产物筛查

因为母体化合物和转化产物存在高度的结构相似性,在二级质谱中也具有相似的碎裂模式。因此,母体化合物与转化产物共享相同的碎片离子或中性损失。基于此,在完成 2.2 的数据清洗后,通过编写的 R 脚本检索与母体共享相同碎片离子或中性损失的转化产物。为减小计算量,每个分子离子仅保留 3 个最强信号的碎片离子。并且考虑到质量数检测存在一定误差,将每个碎片离子设置为范围为 0.001 道尔顿的区间。如图 3 所示,从磺胺嘧啶的碎片离子[156.010, 156.011], [108.043, 108.044]和[92.048, 92.049]以及中性损失[95.048, 95.049], [143.015, 143.016]和[159.009, 159.010]分别检索 input.mzML 文件中提取出的各分子离子及其碎片离子是否存在相同的质量数。结果检索到质量数为 293.0695, 279.0550, 267.0548, 252.0436, 236.0481, 110.0606 的 6 个转化产物,分别与磺胺嘧啶共享 0~2 个相同质量数的碎片离子和中性损失。这 6 个转化产物根据二级质谱图给出可能的结构。此外,分别以各转化产物作为新的母体进一步检索其他代谢产物,转化产物 293.0695, 279.0550 和 267.0548 均存在[96.055, 96.056]的碎片离子,根据该碎片离子检索到分子离子 96.0555, 并给出可能的结构。

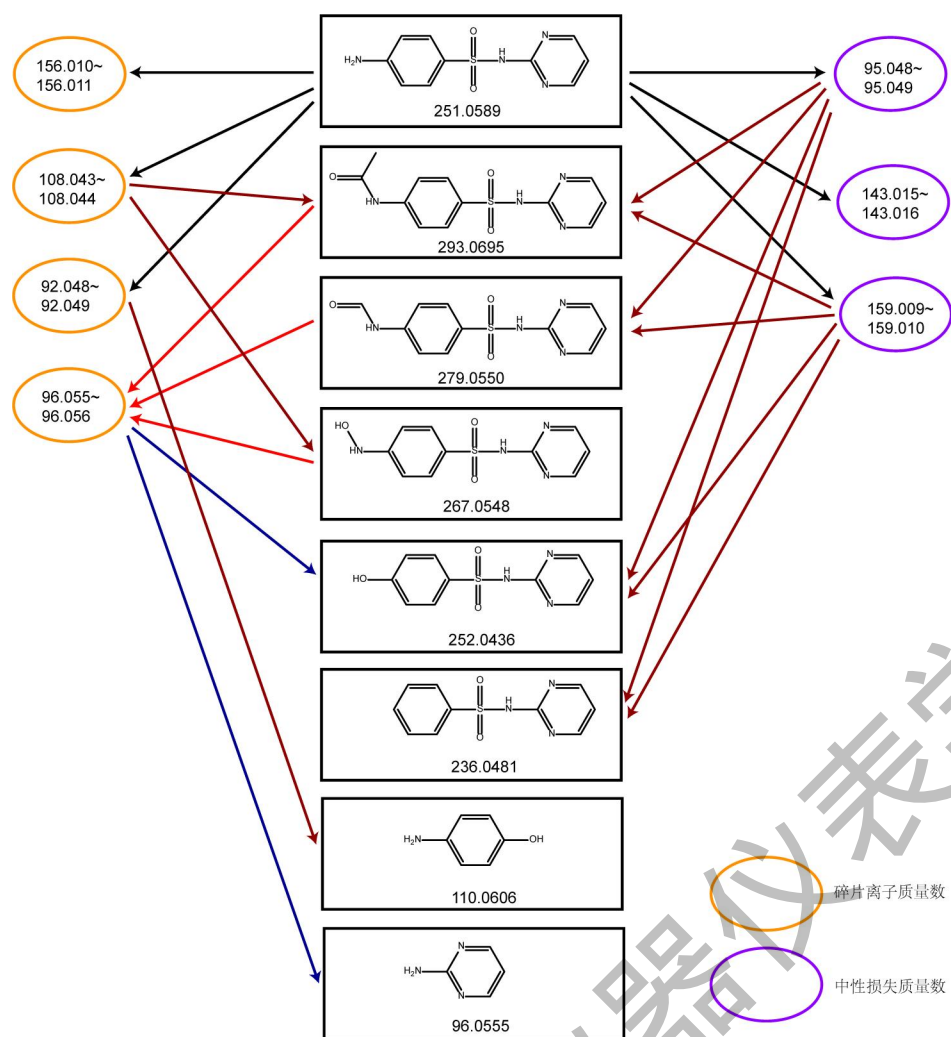


图3 根据磺胺嘧啶碎片离子和中性损失质量数筛查转化产物

### 3.2 谱图数据可视化

如图 4.1 所示，可以在 R 或 Rstudio 编译器中绘制各分子离子的二级质谱图。以磺胺嘧啶为例，调用 R 语言 ggplot2 程辑包，使用 geom\_linerange 函数可绘制出垂线图（图 4.2）。根据需要进行选择具有较强信号（如>30%最大信号）的碎片离子标注质量数（图 4.3）。对于各个碎片离子可根据质量数推测合理的元素组成（图 4.4），并进一步推测相应结构式（图 4.5），最终得到产物分子结构。

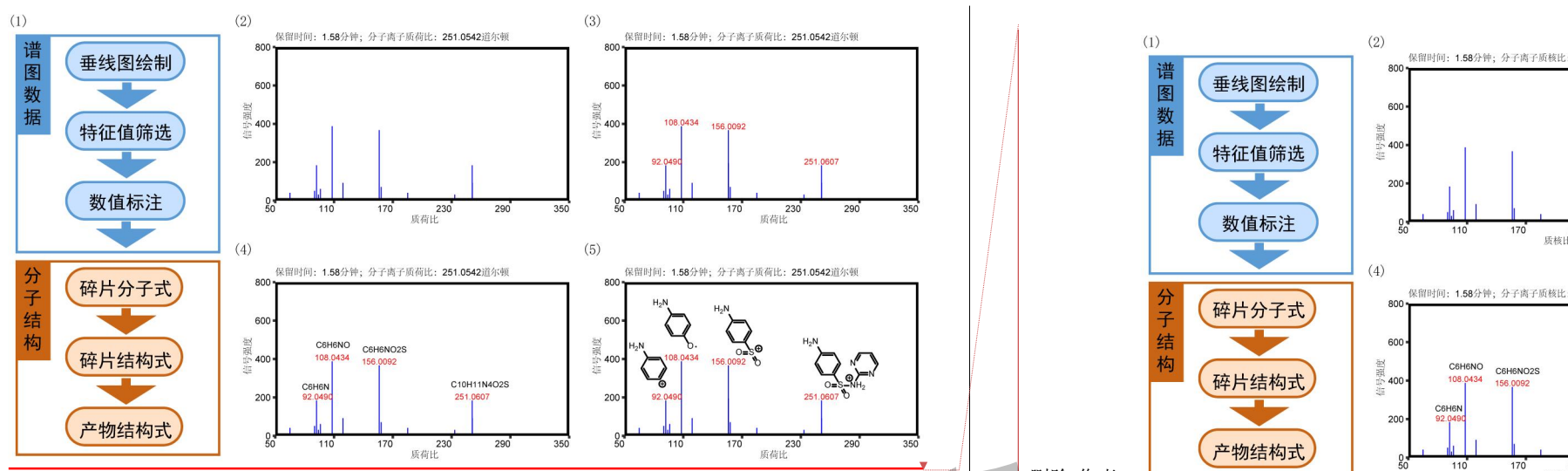


图4 谱图数据可视化及产物结构的鉴定

## 4 结论

从加入磺胺嘧啶的微生物富集液中取样，经过浓缩和纯化，利用四极杆飞行时间串联质谱法获取各馏分的原始质谱数据。通过 ProteoWizard 软件将原始质谱数据转化为 mzML 通用文件格式。利用 R 和 python 脚本，从 mzML 文件中提取数据集用于后续分析。基于转化产物碎裂模式与母体化合物相似的假设，基于磺胺嘧啶的碎片离子和中性损失检索到 7 种可能的转化产物。此外，还介绍了 R 编译器生成可视化二级谱图的方法。

该转化产物筛查方法不受质谱仪器型号、软件专利和数据格式的限制，可以在开源软件上进行转化产物筛查和可视化。在后续研究中，将程序脚本封装和发布有助于降低分析门槛，助力环境中合成化学品的转化产物研究。

### 参考文献:

- [1]Boxall A B A, Sinclair C J, Fenner K, *et al.* When synthetic chemicals degrade in the environment[J]. *Environmental Science & Technology*, 2004, 38(19): 368A-375A.
- [2]Chambers M C, Maclean B, Burke R, *et al.* A cross-platform toolkit for mass spectrometry and proteomics[J]. *Nature Biotechnology*, 2012, 30(10): 918-920.
- [3]Kern S, Fenner K, Singer H P, *et al.* Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry[J]. *Environmental Science & Technology*, 2009, 43(18): 7039-7046.

[4]Sun M M, Ye M, Wu J, *et al.* Positive relationship detected between soil bioaccessible organic pollutants and antibiotic resistance genes at dairy farms in Nanjing, Eastern China[J].

Environmental Pollution, 2015, 206: 421-428.

[5]Yang Y, Zhang X R, Jiang J Y, *et al.* Which Micropollutants in Water Environments Deserve

More Attention Globally?[J]. Environmental Science & Technology, 2022, 56(1): 13-29.

中国仪器仪表表学会