

基于质谱技术的药物代谢产物鉴定策略进展

王慧英, 王锐, 杨华, 李萍*

(中国药科大学天然药物活性组分与药效国家重点实验室, 江苏 南京 210009)

[摘要] 药物代谢产物鉴定是药物代谢研究的重要环节, 也是药理药效研究的物质基础。在药物代谢产物分析检测中, 质谱技术因具有适用范围广、灵敏度高优势而成为药物代谢产物的主要检测手段。基于质谱技术的药物代谢产物鉴定常用的方法有标准品比对、文献及数据库比对、质谱图谱解析、放射性同位素标记以及计算机辅助鉴定。综述近年来常用于药物代谢产物鉴定的质谱鉴定方法研究进展, 以期药物代谢产物研究提供参考。

[关键词] 药物代谢产物; 鉴定; 质谱

[中图分类号] R969.1

[文献标志码] A

[文章编号] 1001-5094 (2018) 03-0193-08

Recent Advances in Mass Spectrometry-based Identification of Drug Metabolites

WANG Huiying, WANG Rui, YANG Hua, LI Ping

(State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, China)

[Abstract] Identification of drug metabolites is an important part of drug metabolism research and the material basis of pharmacology and pharmacodynamics research. Mass spectrometry is a major detection method for drug metabolites due to its high sensitivity and extensive application. In general, the methods of drug metabolites identification include reference standards comparison, literature and database matching, comprehensive spectrometry analysis, radioisotope labeling and computer-assisted structural elucidation. In this paper, recent advances in these methods were summarized in order to provide reference for the identification of drug metabolites.

[Key words] drug metabolites; identification; mass spectrometry

药物代谢产物具有背景基质复杂、种类多样、含量较低的特点, 这要求其分析仪器具有较强的检测能力。因此, 在化合物检测方面具有应用范围广、不受检测对象物态限制、灵敏度高(可检测到pg级别)、分析速度快、信息直观等多种优势的质谱(mass spectrometry, MS)技术成为目前药物代谢产物的主要检测手段^[1]。由于药物代谢产物的背景基质较为复杂, 直接质谱进样对药物代谢产物的离子化效率影响极大, 因此在药物代谢产物分析过程中, 质谱常与液相色谱、气相色谱等分离能力较强的色谱技术串联使用。为了获取更多的代谢产物结构信息, 色谱-质谱联用技术可进一步与紫外检测器、离子淌度光谱等光谱检测技术联用, 得到代谢产物的吸收波长、碰撞截面积等参数。

从药物代谢产物的质谱图谱中可得到化合物的相对分子质量、同位素丰度比、二级质谱碎片、多级

质谱碎片等参数, 其中二级质谱碎片及多级质谱碎片中包含了较多的结构信息, 是代谢产物鉴定的主要依据。保留时间、吸收波长、碰撞截面积等色谱-光谱信息能够反映化合物的一些理化性质, 对化合物鉴定具有辅助作用。故根据化合物的质谱图谱信息及保留时间、吸收波长等参数进行标准品验证以及数据库比对, 因其简单准确、方便快捷, 是目前应用最广泛的药物代谢产物鉴定方式。但现有化学标准品及数据库中所收录的谱图信息十分有限, 不能满足复杂多样的代谢产物结构鉴定。面对这一瓶颈, 研究者们通过丰富代谢产物数据库、总结特定结构化合物的质谱行为规律、引入新方法、开发新的鉴定工具等手段促进药物代谢产物鉴定。

1 标准品验证

标准品验证是最简单准确的药物代谢产物鉴定方法, 即在相同的检测方法下, 对代谢产物与标准品的保留时间、最大吸收波长、相对分子质量、质谱二级碎片等信息进行比较鉴定。因此该方法仅适用于鉴定存在标准品的已知化合物。标准品验证因具有简单、准

接受日期: 2017-07-03

项目资助: 国家新药创制重大专项 (No. 2015ZX09101043-010)

***通讯作者:** 李萍, 教授;

研究方向: 中药药效物质与作用机制研究;

Tel: 025-83271379; **E-mail:** liping2004@126.com

确的特点, 在药物代谢物鉴定过程中得到普遍认可。在标准品易得的情况下, 该法是代谢产物鉴定的首选方法。但在实际研究过程中, 大部分代谢产物的标准品难以获取且价格高昂, 因此对于成分复杂多样的药物代谢产物的鉴定, 运用标准品进行验证较难实现^[2-4]。

2 文献及数据库比对鉴定

药物代谢产物鉴定可通过与相同检测条件下的文献或数据库中的标准图谱数据比对完成化合物的鉴定, 因此该方法仅适用于鉴定存在标准图谱信息的已知化合物。参考代谢组学中代谢产物定性分析的要求^[5], 用于比对的2个化合物需在相同检测条件下存在2个及以上相同且独立的正交参数, 方可认为是同一物质。常用的参数有保留时间、相对分子质量、质谱碎片、吸收波长等。化合物保留时间受色谱柱、检测仪器、流动相、温度等多种因素影响, 变化较大; 吸收波长只能反映化合物的结构类型; 质谱图能够反映代谢产物详细的结构信息且相关标准数据相对较多。因此,

通常情况下, 文献及数据库比对主要依赖于质谱图谱信息的比对。

由于文献中谱图分散, 查找十分困难, 为了便于化合物的分析鉴定, 人们将文献中以及标准品的化学信息、代谢通路、质谱数据谱图等整理成数据库。经过长期的探索和积累, 各国的研究者们建立了多种与代谢产物鉴定相关的化合物数据库并不断地进行丰富更新^[6-8]。由于数据库综合了世界范围内的研究成果且大部分为免费开放型数据库, 因此相对于标准品比对, 数据库比对鉴定方便易行、成本较低。

在代谢产物鉴定过程中常用到的数据库有: 1) 综合性数据库, 例如: Metlin、HMDB、MMCD; 2) 鉴定专用的波谱类数据库, 例如: MassBank、GNPS; 3) 代谢通路数据库, 例如: KEGG、SMPDB、PlantCyc; 4) 化学结构数据库, 例如: PubChem、ChemSpider; 5) 针对某一类物质的小型代谢数据库, 例如: ECMDB、YMDB。代谢产物鉴定过程中, 一些常用的数据库可见表1。

表1 代谢产物鉴定中常用数据库

Table 1 Commonly used databases for metabolite identification

数据库	内容	参考文献
The Human Metabolome Database (HMDB)	人体中小分子代谢物的详细信息	[9]
The Madison Metabolomics Consortium Database (MMCD)	用于代谢研究的核磁、质谱数据库	[10]
Metlin	代谢组学数据库, 包含大量代谢物化学信息和质谱数据	[11]
Spectral Database for Organic Compounds (SDBS)	有机化合物谱图库 (MS、 ¹ HNMR、 ¹³ CNMR、FT-IR、Raman、ESR)	[12]
RIKEN tandem mass spectral database (ReSpec)	植物代谢产物的二级质谱数据库, 提供物种分类与代谢物之间的关系	[13]
The Global Natural Products Social Molecular Networking (GNPS)	天然产物的质谱数据库	[14]
Kyoto Encyclopedia of Genes and Genomes (KEGG)	综合性生物信息学数据库, 包含疾病、生物通路、基因、药物、化学结构等信息	[15]
National Institute of Standards and Technology (NIST)	美国国家标准与技术研究院化学数据库	[16]
KNAPSAcK family databases (KNAPSAcK)	用于植物学研究的植物代谢数据库	[17]
ChemSpider	用于化合物结构快速检索, 包含580 000多个化合物结构信息的数据库	[18]
PubChem	小分子有机化合物的结构生物活性信息	[19]
Golm Metabolome Database (GMD)	植物分子生理学相关的GC-MS数据, 提供普通查询和进化树搜索功能	[20]
MassBank	以质谱分析为基础的数据库	[21]
WEIZMASS	3 540多个植物代谢物的高分辨质谱数据	[22]
The DrugBank database (DrugBank)	药物及其作用靶点的化学和生物学信息	[23]
Chemical Entities of Biological Interest Ontology (ChEBI)	大量活性小分子及其生物活性信息	[24]
Bovine Metabolome Database (BMDB)	7 800多个牛体内代谢物信息	[25]
FoodDB	常见食品及未加工食品的化学信息	[26]
The Escherichia coli Metabolome Database (ECMDB)	大肠杆菌代谢组学数据 (27 000多个代谢物结构、5 000个NMR、LC-MS、GC-MS图谱以及分子生物学、化学生物学信息)	[27]
The Yeast Metabolome Database (YMDB)	酵母代谢组学数据库 (提供酵母小分子代谢产物的GC-MS、LC-MS、NMR信息)	[28]
The Small Molecule Pathway Database (SMPDB)	人体中小分子代谢通路	[29]
PlantCyc	植物代谢通路数据	[30]
The Toxin and Toxin Target Database (T3DB)	常见毒性成分及其靶点	[31]

化合物的谱图信息来自于化合物单体的分析检测,而现存化合物单体数目决定了有限的谱图信息。设备、环境、检测模式等各种因素可导致同一化合物的谱图存在一定的差异,进一步缩小了图谱比对的适用范围。随着现代分析技术及检测手段的不断进步,可被检测到的代谢产物数目日益增加,虽然相关数据库也在不断进行信息的完善和丰富,但其发展速度并不能与代谢产物鉴定所需求的信息量相匹配。故通过文献及数据库比对进行代谢产物鉴定仍具有一定的局限性。

3 质谱图谱解析

在代谢产物分析过程中,各类有机化合物的质谱裂解行为与其结构密切相关,因此质谱图谱解析适用于所有化合物的结构分析。但是仅依靠质谱无法判定化合物的精确结构,所以质谱图谱解析仅能作为化合物结构鉴定的辅助手段。随着先进的分析仪器的不断涌现,在化合物检测中可获得更精准的数据信息,更多反映代谢产物性质的参数。人们在实验过程中不断总结各类化合物的质谱规律,并将这些规律运用到质谱图谱解析当中,进一步促进了代谢产物的鉴定。质谱图谱的具体解析流程如下。

3.1 分子式的确定

代谢产物的分子式可先根据一级质谱得到的相对分子质量,确定化合物的候选分子式。高分辨质谱的相对分子质量偏差极小,可大幅度缩小匹配分子式数目,同时可应用氮律、同位素丰度比^[32]对所得分子式进一步筛选。Kind等^[33]提出7条黄金法则用于精确质谱分子式判定,涵盖元素数量限制、Lewis和Senior规则、碳氢元素比、同位素比、氮氧硫磷与碳元素比、元素比可能性检查、三甲基硅烷检查7个方面,有效缩减候选分子式数目、排除假阳性。夏兵等^[34]在此基础上,采用分子式数据库进行分子式预测的方法,通过分子式出现频率和同位素分布相似度评价等综合评分,进一步缩小候选分子式范围。

3.2 结构推断

根据分子式计算不饱和度,结合最大吸收波长、保留时间等可推测化合物类型;根据质谱特征碎片,碎片差值可推测化合物的裂解过程以及可能存在的官能团。同类化合物具有相同的基本结构,常常能够发生固定的裂解,例如乌头碱型生物碱的二级质谱的主要裂解方式是去除C8位上的取代基^[35];黄酮在负离子

模式下容易发生RDA (Diels-Alder reaction) 裂解,产生以 $^{1,3}A^-$ 为主的特征碎片离子^[36]。在实际应用中,这些特殊碎片还可作为一类化合物的诊断离子进行化合物寻找和结构解析^[37]。

药物代谢常常发生一些固定的代谢反应,代谢产物中存在较多结构相似的同分异构体,其质谱行为相似,主要碎片离子相同,难以区分鉴定。研究者在实验中发现部分位置异构体虽然有相同的MS/MS碎片离子,但却有不同的碎片离子丰度比。例如,Ostrowski等^[38]通过绘制3~30 eV碰撞电压下的各碎片离子丰度-电压曲线,成功区分MS/MS主要碎片相同的3种位置异构体:3-甲氧基-4-羟基苯甲酸、2-羟基-3-甲氧基苯甲酸、4-甲氧基-3-羟基苯甲酸;Cuyckens等^[39]通过质谱联用技术对糖基连在不同位置的18种黄酮苷进行了研究,总结发现糖基连接位置与各碎片离子的相对丰度相关。另外,随着多级质谱的逐渐应用,在化合物二级碎片的基础上可得到更详细的碎片离子的信息,使得原本在二级质谱上具有十分相似裂解方式的化合物得到进一步的区分。例如,相同苷元的黄酮多糖氧苷的二级质谱行为极其相似,Kite等^[40]利用离子阱质谱对黄酮多糖苷进行检测,发现山柰酚-3-O-洋槐二糖苷和山柰酚-3-O-芸香糖苷的二级碎片十分相似,但其三级质谱 $[(M+Na)-\text{山柰酚}]^+$ 的碎片离子却不同。

对于同分异构体化合物的鉴定,还可以采用离子淌度质谱 (ion mobility mass spectrometry, IMMS)。IMMS由离子淌度光谱与质量分析器串联组成,其原理是离子在漂移管中与缓冲气体碰撞时的碰撞截面不同,因此可将离子按大小和形状进行分离,在同分异构体分离以及化合物鉴定方面具有独特优势^[41-42]。IMMS现广泛应用于代谢组学、结构生物学、脂质组学、蛋白质组学以及临床研究中^[43-45]。化合物的碰撞截面积与化合物的构型和空间大小有关且相对稳定,因此碰撞截面积的引入有效提高了药物代谢产物鉴定的准确度^[46-48],甚至仅根据化合物的碰撞截面积和相对分子质量等信息与标准数据比对即可完成对化合物的鉴定^[48]。

以上是质谱图谱解析的一般步骤。对于药物代谢产物的鉴定,若只有质谱谱图,普适性的规律难以精确判定化合物结构,特殊规律例如碎片离子丰度比具有很强的经验性且建立在化合物的基本结构已知的基础之上,因此通过质谱图谱解析进行药物代谢产物鉴定具有一定的局限性。

4 放射性同位素标记鉴定

为了更直观快速地锁定代谢产物, 获取更多的代谢产物结构信息, 人们引入了放射性同位素标记鉴定。该法是寻找、鉴定代谢产物非常直观有效的方法, 适用于所有有机化合物的分析, 其常用方法分为2类, 一类是直接标记, 另一类是间接标记。

4.1 直接标记法

放射性同位素直接标记是在进行代谢反应前对所研究化合物进行放射性同位素标记, 药物经代谢反应后, 直接得到含有放射性同位素标记的代谢产物。该方法能够非常直观地将代谢原型、代谢产物与复杂基质区分开来, 快速而准确地找到代谢产物, 同时根据放射性同位素的存在对代谢产物的结构鉴定具有一定的辅助作用。Jaganath等^[49]在研究芦丁的肠道菌群代谢产物过程中, 将¹⁴C标记的槲皮素与肠道菌群共孵, 用HPLC-RC-MS²对其进行检测, 发现黄酮醇迅速降解, 根据其放射性元素, 检测得到槲皮素的单氧化物、双氧化物及亚甲基醌中间体等代谢产物。放射性同位素标记法虽然准确、直观, 但同位素标记位点直接影响代谢产物获取的全面性, 且实验过程中需考虑放射性同位素本身的半衰期等性质特点, 因此在放射性同位素的选择上需要根据放射性同位素本身的性质、实验条件与目的、实验人员的安全性等多方面进行考虑。在药物代谢研究中¹⁴C与³H是最常用的放射性同位素。

4.2 间接标记法

常用的放射性同位素间接标记法有放射性同位素标记捕获剂法^[50-51]、H/D交换法等。例如, Mezzine等^[50]将曲格列酮和冷冻人肝脏细胞一起共同孵育, 同时加入同位素标记的谷胱甘肽与谷胱甘肽的混合物, 应用LC-HRMS进行检测, 可快速检测到单同位素标记的谷胱甘肽与活性代谢产物的结合物。

H/D交换法与LC-MS、GC、NMR等分析技术相结合, 是近年发展起来用于化合物的结构鉴定的新方法, 其利用氘原子容易与周围的氢原子发生交换的特点^[52-54], 实现对代谢产物的结构鉴定。Shah等^[52]利用柱后注入D₂O, 比较进行H/D交换前后的质谱图谱, 计算不稳定氢数目。使用柳氮磺胺吡啶对整个过程的各个参数进行优化, 优化后的方法适用于各类化合物的分析, 可在各类质谱设备上应用。同时成功应用于普伐他汀内酯和地氯雷他定的氮氧化物的分析和代谢产

物的鉴定。相比用氘代试剂作为流动相的方法, 该方法所消耗的氘代试剂量极少, 且有望应用于仿制品的鉴别、质谱碎片离子分布、化学反应的中间过程分析等方面。

5 计算机辅助鉴定

随着分析仪器的发展, 单个代谢样品中即可检测到成千甚至上万个化合物, 如此庞大的数据亟需无监督的自动化鉴定过程。计算机的引入有效解决了大数据处理难题, 同时加速了无标准谱图化合物的解析过程, 在一定程度上实现了药物代谢产物的自动化鉴定。计算机辅助鉴定方法多种多样, 根据其工作原理, 现分为化学结构数据库基础上的自动化鉴定、碎片树、基于代谢反应预测的代谢产物鉴定3类, 其中化学结构数据库基础上的自动化鉴定适用于无标准谱图的已知化合物的鉴定, 碎片树以及基于代谢反应预测的代谢产物鉴定为普适性方法, 同时适用于未知代谢产物的分析鉴定。

5.1 化学结构数据库基础上的自动化鉴定

相对于存在标准质谱图的化合物, 数据库收录的具有结构信息的化合物在数目上具有绝对优势, 例如质谱专用数据库MassBank中收录了3 127个化合物的26 296张质谱图, 而化学结构数据库ChemSpider中收录了580 000多个化合物结构。如果在化合物结构数据库基础上, 根据化合物结构即可推测出化合物的主要离子碎片, 或者根据离子碎片即可寻找匹配的化合物结构, 那么药物代谢产物可鉴定范围将进一步扩大。研究人员基于这些思考, 将化学信息学与计算机科学相结合, 开发出多种二级质谱、多级质谱的检索与相似度匹配的计算机程序, 实现了基于化合物结构数据库的自动化鉴定过程, 扩大了药物代谢产物的可鉴定范围。目前相关的方法工具可分为以下3类。

5.1.1 计算机模拟MS/MS谱图的代谢产物鉴定 该类方法是根据现有数据库中化合物的MS/MS标准谱图规律建立算法, 并利用现有数据库中的数据对算法不断优化。代谢鉴定工具通过这些算法对实际谱图进行分析, 完成化合物预测、图谱匹配、评分排序等化合物自动化鉴定过程。例如, CFM-ID^[55]是基于二级质谱并运用此原理的网络服务器, 其主要通过概率模型预测裂解过程, 生成候选化合物的裂解图谱, 然后对预测的裂解图谱和实际图谱进行相似性评分。CFM-ID实际

应用过程中具有3方面的作用:1)对已知化合物的二级质谱解析;2)对未知化合物的二级质谱预测;3)通过对特定图谱的候选结构进行预测排序,进而鉴定代谢产物。

5.1.2 基于裂解方式预测的代谢产物鉴定 此类方法是根据化合物结构,对其所有可能裂解方式进行预测,将所有裂解方式产生的离子碎片组合与实际质谱谱图进行匹配评分,得到最佳候选化合物结构。应用此原理的代谢产物鉴定软件很多,例如MS-Finder^[56]、MetFrag^[57]、MIDAS^[58]、MAGMa^[59]。MIDAS是一种通过将MS/MS谱图与数据库中的代谢产物的预测碎片进行匹配完成代谢产物鉴定的数据库检索方法。为了计算匹配度,MIDAS首先根据化学键断裂一般规律列举出代谢产物的可能碎片并计算其合理性,然后对实验数据的MS/MS谱图与代谢产物预测的MS/MS谱图进行匹配评分。Ridder等^[60]结合多级质谱谱图树,构建了一种以子结构为基础的拓展算法。该算法通过建立候选化合物的子结构分层树与各级碎片离子进行匹配,计算匹配分数,根据其匹配分数对从PubChem数据库得到的候选离子进行排序,获取代谢产物的分子结构。以上2种方法对质谱均要求较高的质量精度。MetFrag则以某一特定的相对分子质量对PubChem、KEGG、ChemSpider等数据库进行检索,然后对合理碎片进行计数,结合碎片峰评分完成对候选化合物的进一步筛选。

5.1.3 基于分子指纹特征的代谢产物鉴定 该类算法是根据化合物的碎片图谱对分子指纹特征进行预测,然后根据这些特征在结构数据库中进行检索获取候选化合物。Heinonen等^[61]运用机器学习模型(基于支持向量机原理)完成对分子特征的预测和串联质谱中的代谢产物鉴定。该方法首先根据未知化合物的质谱信息预测出化合物多种特征,然后将预测的化合物特征与大型数据库如PubChem进行匹配获取候选化合物结构。其常用的方法有FingerID、CSI:FingerID^[62],其中CSI:FingerID结合了“碎片树”的思想,详细介绍见“碎片树”部分。

5.2 碎片树

Böcker等^[63]于2008年首次提出“碎片树”(fragmentation trees, FTs)的概念,这是一种基于质谱碎片的从头鉴定方法。该法不需要化合物的精确分子结构、化学数据库、质谱数据库等信息,以碎片离子为节点,以丢失的基团为连线,表征化合物的裂解途径,

运用固定算法计算子树的最佳得分,预测化合物分子式。为了验证方法的可行性,研究人员运用此方法对32个化合物的分子式进行预测,26个化合物的正确分子式位于候选分子式列表的第1位,其余5个化合物的正确分子式也位于候选分子式列表的前五名。该方法提出后,Böcker团队不断地将其完善^[64-65],并将碎片树的思想与机器学习相结合开发出一种基于MS/MS数据对小分子结构进行数据库搜索的新方法——CSI:FingerID^[62],在应用过程中该方法与现阶段其他的代谢产物鉴定工具如MIDAS、MetFrag、MAGMa、CFM-ID、FingerID相比,均表现出更高的准确性。2016年Böcker等^[66]对“碎片树”算法——SIRIUS系列进行了新的优化,提出了一种碎片树的新计算方法——SIRIUS 3,将原本的“组合优化”转换为“最大后验估计”。相较于前面的方法,SIRIUS 3解决了碎片树的计算难题,在未知化合物分子式的从头鉴定和数据库检索寻找结构相似化合物方面更具突出优势。此外,应用于碎片树的软件还有MoleculePuzzle、ISIS等^[2]。

5.3 基于代谢反应预测的代谢产物鉴定

药物代谢产物来源于药物的代谢反应,代谢产物的结构与代谢反应的类型、前体化合物结构性质密切相关。因此研究者利用这些特征,结合化学信息学与化学计量学,开发出一系列代谢产物的预测软件和代谢产物筛选方法,辅助药物代谢产物鉴定。

代谢产物预测软件根据前体化合物的空间结构、化学键的强弱、常见代谢途径,对代谢产物进行预测。目前根据此方法常用的代谢产物预测软件有:Meteor、MetaSite、StarDrop和MetaDrug^[67-69]。Meteor主要根据现有代谢数据中的代谢规律和推理算法对化合物的代谢反应进行预测且能够覆盖主要的代谢反应,灵敏度高,但其缺点是预测化合物数量大于代谢产物数量,存在一定的假阳性,准确性低。MetaSite是一种反应自动对接模型,主要用于预测P450的I相代谢反应。StarDrop则主要运用量子力学的方法预测所要查询化合物代谢反应中CYP3A4、CYP2D6和CYP2C9的参与程度。以上预测软件基于不同的原理,多个预测软件的结合使用可提高化合物预测的准确性。

面对复杂的代谢体系,这些软件预测能力十分有限。因此,一些研究者在对复杂体系代谢产物的鉴定过程中,首先应用化学计量学手段对代谢数据进行预处理,如背景扣除、主成分分析(PCA)、质量亏损过

滤(MDF)、子离子或中性丢失过滤,然后综合代谢反应及其图谱特征进行鉴定。例如,MassMetaSite^[67]就是在MetaSite基础上,将代谢产物预测功能与上述预处理过程相结合的化合物自动化鉴定软件。另外,在寻找和鉴定药物肠道菌群代谢产物过程中,研究者们常常根据原产物以及可能发生的代谢反应,应用MetaboLynx软件,对样品和基质图谱进行扫描,寻找发生固定代谢反应的代谢产物^[70]。或者根据代谢反应直接对代谢产物的相对分子质量和分子式进行推算,结合MDF,完成代谢产物的寻找。但MDF仅仅能够实现少数步骤的代谢反应的搜索,容易造成代谢产物的丢失。Wang等^[71]通过使用MATLAB软件将任意2个代谢产物之间的差值与可能发生的代谢反应造成的相对分子质量改变相匹配,减少了代谢产物的丢失。同时计算化合物反应对之间相同的中性丢失、碎片离子、碎片离子差,三者之和与碎片离子总数的比值,得到每个化合物的匹配系数,保留每一个化合物的最大

值,以达到对代谢产物进行辅助鉴定的目的。

6 结语

随着仪器的精密度与灵敏度逐渐提升,可检测到的代谢产物越来越多,冗繁复杂的大数据及一些未知代谢产物的鉴定难题亟需解决。化学信息学、化学计量学、统计学、计算机科学等与药学的多学科交叉,在一定程度上实现了鉴定的自动化过程,部分解决了传统鉴定过程中的标准图谱有限、工作量大、无标准图谱化合物难以鉴定等难题。但相比于人工鉴定,目前自动化的鉴定结果尚不够精准,存在一定的假阳性和假阴性,因此计算机辅助鉴定与质谱谱图解析、数据库比对、标准品验证等多种方法相结合是目前代谢产物鉴定最有效手段;而建立更精准的计算机鉴定策略,实现化合物的自动化鉴定是未来的发展方向。

[参考文献]

- [1] Saurina J, Sentellas S. Strategies for metabolite profiling based on liquid chromatography[J]. *J Chromatogr B*, 2017, 1044/1045: 103-111.
- [2] Vaniya A, Fiehn O. Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics[J]. *Trends Analyt Chem*, 2015, 69: 52-61.
- [3] Erratico C, Negreira N, Norouzizadeh H, et al. In vitro and in vivo human metabolism of the synthetic cannabinoid AB-CHMINACA[J]. *Drug Test Anal*, 2015, 7(10): 866-876.
- [4] Wurita A, Hasegawa K, Minakata K, et al. Identification and quantification of metabolites of AB-CHMINACA in a urine specimen of an abuser[J]. *Leg Med*, 2016, 19: 113-119.
- [5] Sumner L W, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)[J]. *Metabolomics*, 2007, 3(3): 211-221.
- [6] 赵丽梅, 谭宁华. 国外天然产物化学成分实物库及数据库建设概况[J]. *中国中药杂志*, 2015, 40(1): 29-35.
- [7] 申国安, 段礼新, 漆小泉. 植物代谢组学数据分析和数据库[J]. *生命科学*, 2015, 27(8): 995-999.
- [8] 张小蒙, 孙明慧, 李艳萍, 等. 几种用于寻找新型天然活性物质的开放型代谢组学数据库介绍[J]. *生物技术通报*, 2016, 32(12): 29-33.
- [9] Wishart D S, Jewison T, Guo A C, et al. HMDB 3.0-The Human Metabolome Database in 2013[J]. *Nucleic Acids Res*, 2013, 41(Database issue): D801-D807.
- [10] Cui Q, Lewis I A, Hegeman A D, et al. Metabolite identification via the Madison Metabolomics Consortium Database[J]. *Nat Biotechnol*, 2008, 26(2): 162-164.
- [11] Smith C A, O'Maille G, Want E J, et al. METLIN: a metabolite mass spectral database[J]. *Ther Drug Monit*, 2005, 27(6): 747-751.
- [12] NIOAI Science. Spectral Database for Organic Compounds, SDBS[DB/OL]. [2017-03-03]. <http://sdb.sdb.aist.go.jp>.
- [13] Suzuki M, Sato M, Sakata A, et al. RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database[J]. *Phytochemistry*, 2012, 82(2): 38-45.
- [14] Wang M, Carver J J, Phelan V V, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking[J]. *Nat Biotechnol*, 2016, 34(8): 828-837.
- [15] Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs[J]. *Nucleic Acids Res*, 2017, 45(Database issue): D353-D361.
- [16] Averill J D, Reneke P, Peacock R D. The National Institute of Standards and Technology (NIST)[DB/OL]. (2017-02-25)[2017-03-03]. <https://www.nist.gov/>.
- [17] Afendi F M, Okada T, Yamazaki M, et al. KnapSack family databases: integrated metabolite-plant species databases for multifaceted plant research[J]. *Plant Cell Physiol*, 2012, 53(2): 1-12.
- [18] Anon. Editorial: ChemSpider-a tool for natural products research[J]. *Nat Prod Rep*, 2015, 32(8): 1163-1164.
- [19] Kim S, Thiessen P A, Bolton E E, et al. PubChem substance and compound databases[J]. *Nucleic Acids Res*, 2016, 44(Database issue): D1202-D1213.
- [20] The German Research Foundation (Deutsche Forschungsgemeinschaft,

- DFG). Golm Metabolome Database[DB/OL]. [2017-03-03]. <http://gmd.mpimp-golm.mpg.de/>.
- [21] Horai H, Arita M, Kanaya S, *et al*. MassBank: a public repository for sharing mass spectral data for life sciences[J]. *J Mass Spectrom*, 2010, 45(7): 703-714.
- [22] Nir S, Ilana R, Uwe H, *et al*. The WEIZMASS spectral library for high-confidence metabolite identification[J]. *Nat Commun*, 2016, 7: 12423-12435.
- [23] Law V, Knox C, Djoumbou Y, *et al*. DrugBank 4.0: shedding new light on drug metabolism[J]. *Nucleic Acids Res*, 2014, 42(Database issue): D1091-D1097.
- [24] Hastings J, Owen G, Dekker A, *et al*. ChEBI in 2016: improved services and an expanding collection of metabolites[J]. *Nucleic Acids Res*, 2016, 44(Database issue): D1214-1219.
- [25] Genome Alberta & Genome Canada. Bovine Metabolome Database[DB/OL]. [2017-03-03]. <http://www.cowmetdb.ca/cgi-bin/browse.cgi>.
- [26] The Metabolomics Innovation Centre (TMIC). FooDB[DB/OL]. [2017-03-03]. <http://foodb.ca/>.
- [27] Sajed T, Marcu A, Ramirez M, *et al*. ECMDB 2.0: a richer resource for understanding the biochemistry of *E. coli*[J]. *Nucleic Acids Res*, 2016, 44(Database issue): D495-D501.
- [28] Jewison T, Knox C, Neveu V, *et al*. YMDB: the Yeast Metabolome Database[J]. *Nucleic Acids Res*, 2012, 40(Database issue): D815-D820.
- [29] Jewison T, Su Y, Disfany F M, *et al*. SMPDB 2.0: big improvements to the Small Molecule Pathway Database[J]. *Nucleic Acids Res*, 2014, 42(Database issue): D478-D484.
- [30] Carnegie Institution for Science, Department of Plant Biology. Plant Metabolic Pathways Databases [DB/OL]. [2017-03-03]. <http://www.plantcyc.org/>.
- [31] Lim E, Pon A, Djoumbou Y, *et al*. T3DB: a comprehensively annotated database of common toxins and their targets[J]. *Nucleic Acids Res*, 2010, 38(Database issue): D781-D786.
- [32] Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm[J]. *BMC Bioinformatics*, 2006, 7(1): 234-243.
- [33] Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry[J]. *BMC Bioinformatics*, 2007, 8(1): 105-124.
- [34] 夏兵, 周燕, 丁立生, 等. 一种高分辨质谱准确分子式预测方法: 中国, 103792275A[P]. 2014-05-14.
- [35] Jing Z, Zhi H H, Xiao H Q, *et al*. Neutral fragment filtering for rapid identification of new diester-diterpenoid alkaloids in roots of *aconitum carmichaeli* by ultra-high-pressure liquid chromatography coupled with linear ion trap-orbitrap mass spectrometry[J]. *PLoS One*, 2012, 7(12): 1-12.
- [36] Qing Z X, Zhao H, Tang Q, *et al*. Systematic identification of flavonols, flavonol glycosides, triterpene and siraitic acid glycosides from *Siraitia grosvenorii* using high-performance liquid chromatography/quadrupole-time-of-flight mass spectrometry combined with a screening strategy[J]. *J Pharm Biomed Anal*, 2017, 138: 240-248.
- [37] Qiao X, Li R, Song W, *et al*. A targeted strategy to analyze untargeted mass spectral data: rapid chemical profiling of *Scutellaria baicalensis* using ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry and key ion filtering[J]. *J Chromatogr A*, 2016, 1441: 83-95.
- [38] Ostrowski W, Wojakowska A, Grajzer M. Mass spectrometric behavior of phenolic acids standards and their analysis in the plant samples with LC/ESI/MS system[J]. *J Chromatogr B*, 2014, 967(6): 21-27.
- [39] Cuyckens F, Claeys M. Determination of the glycosylation site in flavonoid mono-O-glycosides by collision-induced dissociation of electrospray-generated deprotonated and sodiated molecules[J]. *J Mass Spectrom*, 2005, 40(3): 364-372.
- [40] Kite G C, Veitch N C. Identification of common glycosyl groups of flavonoid O-glycosides by serial mass spectrometry of sodiated species [J]. *Rapid Commun Mass Spectrom*, 2011, 25(18): 2579-2590.
- [41] Lanucara F, Holman S W, Gray C J, *et al*. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics[J]. *Nat Chem*, 2014, 6(4): 281-294.
- [42] Li H, Bendiak B, Siems W F, *et al*. Carbohydrate structure characterization by tandem ion mobility mass spectrometry (IMMS)2[J]. *Anal Chem*, 2013, 85(5): 2760-2769.
- [43] Lu W, Xu Y, Zhao Y, *et al*. Emerging technologies, recent developments, and novel applications for drug metabolite identification[J]. *Curr Drug Metab*, 2014, 15(9): 865-874.
- [44] Paglia G, Angel P, Williams J P, *et al*. Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification[J]. *Anal Chem*, 2014, 87(2): 1137-1144.
- [45] Chouinard C D, Wei M S, Beekman C R, *et al*. Ion mobility in clinical analysis: current progress and future perspectives[J]. *Clin Chem*, 2016, 62(1): 124-133.
- [46] Regueiro J, Negreira N, Berntssen M H G. Ion mobility-derived collision cross section as an additional identification point for multi-residue screening of pesticides in fish feed[J]. *Anal Chem*, 2016, 88(22): 11169-11177.
- [47] Zhou Z, Shen X, Tu J, *et al*. Large-Scale Prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry[J]. *Anal Chem*, 2016, 88(22): 11084-11091.
- [48] Stephan S, Hippler J, Köhler T, *et al*. Contaminant screening of wastewater with HPLC-IM-qTOF-MS and LC + LC-IM-qTOF-MS using a CCS database[J]. *Anal Bioanal Chem*, 2016, 408(24): 6545-6555.
- [49] Jaganath I B, Mullen W, Lean M E, *et al*. *In vitro* catabolism of rutin by human fecal bacteria and the antioxidant capacity of its catabolites[J]. *Free Radic Biol Med*, 2009, 47 (8): 1180-1189.
- [50] Mezine I, Bode C, Raughley B, *et al*. Application of exogenous mixture of glutathione and stable isotope labeled glutathione for trapping reactive metabolites in cryopreserved human hepatocytes. Detection

- of the glutathione conjugates using high resolution accurate mass spectrometry[J]. *Chem Biol Interact*, 2013, 204 (3): 173-184.
- [51] Yamaoka T, Kitamura Y. Characterization of a highly sensitive and selective novel trapping reagent, stable isotope labeled glutathione ethyl ester, for the detection of reactive metabolites[J]. *J Pharmacol Toxicol Methods*, 2015, 76(2015): 83-95.
- [52] Shah R P, Garg A, Putlur S P, *et al.* Practical and economical implementation of online H/D exchange in LC-MS[J]. *Anal Chem*, 2013, 85(22): 10904-10912.
- [53] Chandak M S, Nakamura T, Takenaka T, *et al.* The use of spin desalting columns in DMSO-quenched H/D-exchange NMR experiments[J]. *Protein Sci*, 2013, 22(4): 486-491.
- [54] Liu T, Du F, Zhu F, Xing J. Metabolite identification of artemether by data-dependent accurate mass spectrometric analysis using an LTQ-Orbitrap hybrid mass spectrometer in combination with the online hydrogen/deuterium exchange technique[J]. *Rapid Commun Mass Spectrom*, 2011, 25(21): 3303-3313.
- [55] Allen F, Pon A, Wilson M, *et al.* CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra[J]. *Nucleic Acids Res*, 2014, 42(Web Server issue): W94-W99.
- [56] Tsugawa H, Kind T, Nakabayashi R, *et al.* Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software[J]. *Anal Chem*, 2016, 88(16): 7946-7958.
- [57] Ruttkies C, Schymanski E L, Wolf S, *et al.* MetFrag relaunched: incorporating strategies beyond in silico fragmentation[J]. *J Cheminform*, 2016, 8(1): 1-16.
- [58] Wang Y, Kora G, Bowen B P, *et al.* MIDAS: a database-searching algorithm for metabolite identification in metabolomics[J]. *Anal Chem*, 2014, 86(19): 9496-9503.
- [59] Verdegem D, Lambrechts D, Carmeliet P, *et al.* Improved metabolite identification with MIDAS and MAGMa through MS/MS spectral dataset-driven parameter optimization[J]. *Metabolomics*, 2016, 12(6): 1-16.
- [60] Ridder L, van der Hooft J J, Verhoeven S, *et al.* Substructure-based annotation of high-resolution multistage Ms^n spectral trees[J]. *Rapid Commun Mass Spectrom*, 2012, 26(20): 2461-2471.
- [61] Heinonen M, Shen H, Zamboni N, *et al.* Metabolite identification and molecular fingerprint prediction through machine learning[J]. *Bioinformatics*, 2012, 28(18): 2333-2341.
- [62] Dührkop K, Shen H, Meusel M, *et al.* Searching molecular structure databases with tandem mass spectra using CSI: FingerID[J]. *Proc Natl Acad Sci USA*, 2015, 112(41): 12580-12585.
- [63] Böcker S, Rasche F. Towards de novo identification of metabolites by analyzing tandem mass spectra[J]. *Bioinformatics*, 2008, 24(16): 149-155.
- [64] Dührkop K, Scheubert K, Böcker S. Molecular formula identification with SIRIUS[J]. *Metabolites*, 2013, 3(2): 506-516.
- [65] Shen H, Dührkop K, Böcker S. Metabolite identification through multiple kernel learning on fragmentation trees[J]. *Bioinformatics*, 2014, 30(12): i157-i164.
- [66] Böcker S, Dührkop K. Fragmentation trees reloaded[J]. *J Cheminform*, 2016, 8(1): 1-26.
- [67] Pähler A, Brink A. Software aided approaches to structure-based metabolite identification in drug discovery and development[J]. *Drug Discov Today Technol*, 2013, 10(1): 207-217.
- [68] T'jolllyn H, Boussery K, Mortishire-Smith R J. Evaluation of three state-of-the-art metabolite prediction software packages (Meteor, MetaSite, and StarDrop) through independent and synergistic use[J]. *Drug Metab Dispos*, 2011, 39(11): 2066-2075.
- [69] Ekins S, Andreyev S, Ryabov A, *et al.* Computational prediction of human drug metabolism[J]. *Expert Opin Drug Metab Toxicol*, 2005, 1(2): 303-324.
- [70] Zhao M, Du L, Tao J, *et al.* Ultra-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry for rapid analysis of the metabolites of morroniside produced by human intestinal bacteria[J]. *J Chromatogr B*, 2015, 976/977(2015): 61-67.
- [71] Wang L, Ye H, Sun D, *et al.* Metabolic pathway extension approach for metabolomic biomarker identification[J]. *Anal Chem*, 2017, 89(2): 1229-1237.



【专家介绍】李萍: 博士, 教授, 博士研究生导师, 国家自然科学基金委创新研究群体学术带头人、国家杰出青年基金获得者、教育部长江学者特聘教授。现任天然药物活性组分与药效国家重点实验室主任。兼任国家药典委员会委员、美国USP-HMC东亚专家组成员、世界中医药联合会中药分析委员会副主任委员等职。

长期从事中药活性成分群发现与质量评价研究。主要研究领域包括: 中药活性成分(群)发现新理论、新方法研究; 中药质量评价新方法研究; 中药创新药物研究等。已在 *Natural Product Reports*、*Trends Analytical Chemistry*、*Green Chemistry*、*Chemical Communication*、*Journal of Chromatography A* 等本领域国际期刊发表SCI论文300余篇。获授权发明专利20余项。获得国家科技进步二等奖、教育部自然科学一等奖等奖励。主编十一五、十二五普通高等教育本科国家级规划教材《生药学》以及《中药分析学》等教材与《现代生药学》、《中国药典》配套的《中华人民共和国药典中药材显微鉴别彩色图鉴》等著作。