



液质联用技术用于复杂混合物体系中小分子化合物的分析

余玲玲^{1,2}, 李兴^{2,3}, 陈燕丽², 文超^{2,3}, 胡凯锋^{2*}

1. 云南大学化学科学与工程学院, 昆明 650091

2. 中国科学院昆明植物研究所, 植物化学与西部植物资源持续利用国家重点实验室, 昆明 650201

3. 中国科学院大学, 北京 100049

*通讯作者, E-mail: kaifenghu@mail.kib.ac.cn

收稿日期: 2017-07-15; 接受日期: 2017-09-15; 网络版发表日期: 2017-11-23

国家重点研发计划(编号: 2017YFC0906900)、国家自然科学基金(编号: 21505142)、云南省高端科技人才引进计划(编号: 2012HA015)、青年千人计划和上海高校中药药效物质E-研究院资助项目

摘要 复杂混合物体系中化合物类型多样, 理化性质各异, 对其组成成分进行定性定量分析一直是分析化学研究的热点问题. 液相色谱与质谱联用(liquid chromatography-mass spectroscopy, LC-MS), 尤其是超高效液相色谱(ultra-high performance liquid chromatography, UPLC)结合高分辨质谱(high resolution mass spectrometry, HRMS)是复杂样品分析的有力工具. UPLC能够极大程度提高色谱分离能力和效率, 为复杂混合物中各组分的有效分离提供保证; HRMS能够提供高精度的分子质量, 提高定性定量分析的可靠性和准确性. 本文综述了LC-MS技术在分析典型的复杂混合物体系(尿液、血液、细胞代谢物及天然产物提取物)中小分子化合物组分及含量中的应用, 主要从样品制备、数据采集、数据(前)处理、目标化合物的定性定量分析及多样本多变量分析(代谢组学分析)用于差异代谢物(生物标志物)的发现等方面展开, 并对LC-MS技术用于复杂混合物样品分析当前存在的问题及未来的发展进行了总结和展望.

关键词 超高效液相色谱, 高分辨质谱, 复杂混合物, 定性分析, 定量分析, 多变量分析

1 引言

1999年4月, 美国*Science*杂志出版了以“复杂系统”为主题的专辑, 主要针对化学^[1]、生物^[2]、神经系统^[3]等领域的复杂性研究进行了报道. 在分析化学领域, 复杂体系分析主要包括混合物样品中目标化合物的定性定量分析及多样本的代谢组学分析等, 复杂混合物样品中包含大量已知和未知、理化性质各异的组分.

多组分复杂混合物样品的研究逐渐引起越来越多的关注, 如血液^[4,5]、细胞代谢物^[6,7]、天然产物^[8-10]等样本中的成分分析. 生物样品组成成分非常复杂. 例如, 血液、细胞提取物中既存在核酸、蛋白质、多糖、脂类等生物大分子, 也存在各种代谢物、无机盐等小分子化合物; 如何从复杂的混合物体系中发现并识别已知化合物, 鉴定未知化合物, 进一步实现对微量或痕量目标化合物的准确定性定量分析一直是近年来

引用格式: 余玲玲, 李兴, 陈燕丽, 文超, 胡凯锋. 液质联用技术用于复杂混合物体系中小分子化合物的分析. 中国科学: 化学, 2017, 47: 1379-1391
Yu L, Li X, Chen Y, Wen C, Hu K. Analysis of small molecule compounds in complex mixture system by LC-MS. *Sci Sin Chim*, 2017, 47: 1379-1391, doi: [10.1360/N032017-00118](https://doi.org/10.1360/N032017-00118)

分析化学研究的热点问题。

复杂混合物体系组成成分的分析与分离分析技术的发展密切相关。近年来,随着分离与分析技术的不断进步,结合色谱、光谱及波谱的各种联用技术在复杂混合物体系定性定量分析研究中的应用越来越广泛,如液相(气相)色谱-红外光谱、液相(气相)色谱-质谱、液相(气相)色谱-核磁共振等联用技术;其中,液相(气相)色谱-质谱联用技术在复杂样品成分分析方面的应用较为普遍。气相色谱-质谱联用技术(gas chromatography-mass spectroscopy, GC-MS)主要用于分析易挥发、热稳定的化合物;液相色谱-质谱(LC-MS)与GC-MS具有一定的互补性质,可用于分析极性、非极性、不易挥发、热不稳定的物质。LC-MS具有高效、灵敏的特点,使其在复杂混合物体系分析中发挥着尤为突出的作用,在微量或痕量物质的分析方面较其他检测技术具有明显优势;尤其是超高效液相色谱-高分辨质谱(UPLC-HRMS)技术的发展大大提高了复杂混合物的分离效率和定性定量分析的准确性。LC-MS作为分析复杂样品最有力的工具之一,在食品^[11]、医药^[12]、检测化验^[13]等与人类生活息息相关的领域都有广泛应用。运用LC-MS分析复杂样品中有机小分子类化合物主要包括以下步骤:样品制备、数据采集、数据(前)处理、定性定量分析以及多样本基于代谢组学的多变量统计分析(图1)。根据研究目的和样品性质的不同,尽可能排除干扰,选择合适的样品制备方法是LC-MS分析的前提;建立快速灵敏的数据采集方法及合理的数据处理与分析策略是获得准确、可重现分析结果的基础。复杂样品分析的目标,一方面是要实现对其组成成分的准确定性定量分析;另一方面,针对多样本分析,开发建立高效平行的数据采集方法、高通量的数据(前)处理步骤以及采用正确、合理的多变量分析方法对获得的分析结果和结论的可靠性至关重要,如代谢组学研究中代谢标志物

的发现。本文综述了LC-MS在典型的复杂混合物体系(如尿液、血液、细胞代谢物、天然产物提取物)中小分子化合物分析中的运用,主要从样品制备、数据采集、数据(前)处理、目标化合物的定性定量分析及多样本多变量分析(代谢组学分析)用于差异代谢物(代谢组学标志物)的发现等方面(图1)展开论述。

2 LC-MS分析样品的制备

样品制备过程包括样品采集、储存及处理。由于复杂混合物样品化合物类型复杂多样、基质干扰多,并且生物样品往往取样量少,待测物浓度低,在测定前除少数样品可以作简单处理直接分析外,多数需要进一步分离、纯化,排除干扰,增加目标化合物响应。样品制备在整个分析过程中是一个既耗时又容易引进误差的步骤,近年来引起越来越多科研工作者的关注。以下主要针对常用的混合物样品,如尿液、血液、细胞代谢物以及天然产物提取物的制备方法进行综述。

2.1 尿液样品的制备

随着医疗卫生技术力量和水平的提升,LC-MS检测分析已越来越广泛地应用于医学检验和临床研究中,人体血液和尿液样品是最常用的LC-MS分析对象。在实际研究中,需根据样品的性质、目标成分以及分析目的的不同,选择恰当的样品制备方法。收集生物样品后,往往需要进行除蛋白、分离与浓缩、缀合物水解等预处理^[14,15]。尿液样品组成成分主要为水、无机盐及极性小分子代谢物,由于其中某些成分的排出量受昼夜节律影响较大,常收集24 h内排出的尿液,混合进行分析。尿液样品受氧气作用,其中的尿胆原、胆红素等易氧化变质;且尿样容易生长细菌,细菌的生长繁殖可使尿素分解,产生氨,改变尿液的pH。因此,尿液样品长期保存时,应储存于-20℃环境下。为了最大限度保留尿液中的代谢物,常采用直接离心或过滤

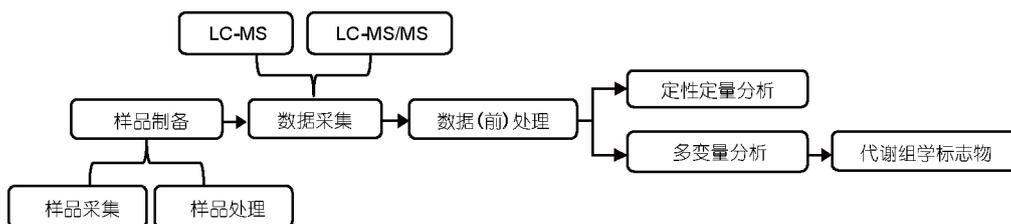


图1 复杂混合物样品LC-MS分析流程图

方法, 移除悬浮微粒. 由于尿液中存在磷酸盐、硫酸盐等各种盐类物质及微量的蛋白质, 往往会污染液相色谱柱及质谱离子源, 且会造成色谱保留时间漂移. 为降低LC-MS分析的基质效应, 常用蒸馏水或甲醇对尿液进行稀释^[16,17]; Wong等^[16]比较了直接使用0.22 μm 滤膜过滤、水稀释(1:1, v/v)、甲醇稀释(1:1, v/v) 3种不同的处理方法, 考察检出特征数目及方法的重现性, 结果表明使用甲醇稀释检出代谢物的同质性及方法重现性较好, 但直接过滤处理检出特征数目最多, 适合尿液中代谢物的全面分析(图2).

2.2 血液样品的制备

为了获得准确可靠的LC-MS分析结果, 对于血液样品的制备常需考虑以下因素: 受试者人口学资料的收集、血液收集管、溶血、新鲜血液的处理及冻融操作. 在大规模临床血样代谢组研究中首先需要考虑收集受试者相关的人口学资料, 主要包括: (1) 受试者年龄、性别和民族; (2) 昼夜节律(对于某些激素如皮质醇、孕酮、睾酮等尤其重要); (3) 饮食和运动; (4) 个体差异^[18]. 在一项血样代谢组研究中, Minami等^[19]发现不同类型的溶血磷脂酰胆碱浓度具有显著的昼夜节律波动, 表明昼夜节律差异和生理性的生物节律对LC-MS分析结果有重要影响. 同时, 该研究结果还显示采用代谢谱筛查手段诊断新生儿代谢(如尿素循环代谢物鸟氨酸、瓜氨酸和4-胍基丁酸)缺陷, 收集血样时昼夜节律是一个非常重要的因素^[19]. 已有研究表明, 饮食和运动可影响基于LC-MS的血样代谢组结果^[20], 因而一般推荐在身体平静状态下采集受试者空

腹(8~10 h)血样用于LC-MS分析^[21]. 为了防止“化学噪音”干扰LC-MS血样代谢谱, 考虑采用同一批次的专用采血管, 因为采血管的材料组分、添加剂都可影响LC-MS分析过程中的离子化, 以至于改变甚至抑制代谢物离子化和/或引入干扰峰^[22]. Dunn等^[23]在大规模血清和血浆代谢组研究中考察了多种抗凝剂的影响, 建议避免使用柠檬酸钠, 这种添加剂可造成某些内源性代谢物的信号变化. 对于溶血样品应标记为不合格样品, 因为溶血过程中, 红细胞、白细胞和血小板等血细胞的破裂可释放某些离子(钾、镁、铁、磷)、代谢物(尿素、色氨酸、脂质)、结构蛋白(血红蛋白)、酶(天冬酰胺氨基转移酶、乳酸脱氢酶和白细胞蛋白酶), 这些物质最终都可直接影响血浆、血清的代谢物谱, 如内源性化合物的降解^[24,25]. 由于血细胞的代谢活动及胞内某些物质的缓慢释放, 因而全血样品长时间暴露于室温可影响某些物质的稳定性和浓度, 故对于新鲜血液样品应及时将血浆或血清与血细胞进行分离, 一般建议在2 h内完成. 对于不能立即进行LC-MS分析的样品, 通常要求储存于 -70°C 或更低的温度. 有研究表明, 血液样品的反复冻融操作可造成代谢物谱的显著变化^[26]. 为了避免多次冻融操作, 一般可将血样分装成小包装低温保存. 上述血液样品收集、处理和管理的不当操作都能在LC-MS分析结果中反映出来. 如果不排除这些因素的影响, 将影响LC-MS分析结论的可靠性. 减少上述误差的有效方法是建立一套血浆或血清样品LC-MS分析的标准操作规程(standard operation procedure, SOP).

为防止蛋白质及其复合物在色谱柱上沉积、堵塞, 导致分析重现性下降, 需要对血液样品进行除蛋白处理. 分析血液样品中极性较大的糖类、氨基酸时, 常使用与水相混溶的乙腈、甲醇、乙醇等有机溶剂沉淀蛋白; 将水溶性有机溶剂与血浆或血清按一定比例混合, 涡旋、离心、取上清液进行分析. Bruce等^[27]用乙腈、丙酮、甲醇和乙醇4种有机溶剂对血浆样品进行处理, 以发现标志物的数量、数据的重现性及对色谱柱寿命的影响为指标, 考察不同溶剂组成对血浆样品中小分子代谢物($\text{MW}<1500\text{ Da}$)的提取效率, 研究表明以甲醇-乙醇(1:1, v/v)或甲醇-乙腈-丙酮(1:1:1, v/v/v)混合溶剂与血浆(4:1, v/v)混合, 涡旋15 min条件最优; 同时表明, 在 4°C 下离心10 min, 沉淀蛋白效果较好. 对于样品中难离子化的目标化合物, 常采用衍生化处理;

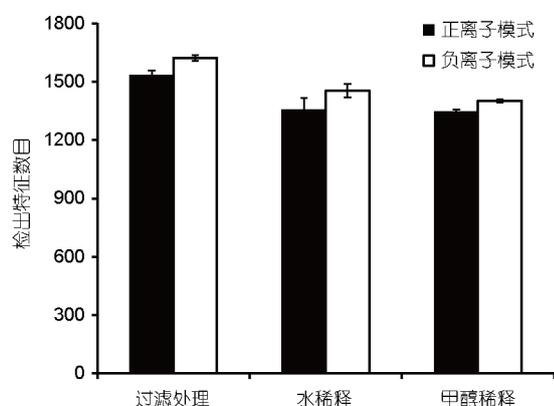


图2 不同方法处理尿液检出特征数目比较^[16]

通过与易电离的物质反应, 改变待分析物的物理和化学性质。例如, 测定血浆样品中难离子化的糖类化合物, 常采用还原胺化反应, 生成易离子化的稳定衍生物, 增强质谱响应^[28,29] (图3)。

2.3 细胞代谢物样品制备

由于细胞内的酶系活跃, 代谢物之间的转换较快, 细胞内代谢物制备方法显著影响分析结果的准确性、重现性。因此, 运用代谢组学方法研究某一时刻/特定情况下的细胞内代谢物, 建立合适的样品制备方法是取得可靠分析结果的先决条件。用于代谢组学研究的细胞代谢物样品制备步骤主要包括: 细胞淬灭和代谢物提取。

2.3.1 细胞淬灭

为了准确反映特定情况下代谢物的信息, 需要使细胞内的酶迅速失活, 终止代谢反应, 确保代谢物谱瞬间“冻结”。理想的淬灭方法应满足以下两个要求: (1) 细胞内的酶迅速失活; (2) 保持细胞完整性、不造成胞内代谢物泄露^[30]。贴壁细胞样品制备方法较悬浮细胞复杂, 在淬灭前首先需使用磷酸盐缓冲液(PBS)或等渗盐水清洗细胞, 移除培养皿中残留的培养基, 以防培养基溶解于提取剂中, 干扰后续分析; 进一步需将贴壁细胞与培养皿分离, 传统方法利用胰蛋白酶消化细胞, 由于消化时间不易控制, 时间过长会水解细胞膜上蛋白质, 造成胞内代谢物泄露, 不可避免改变细胞代谢物谱^[31]; 因此, 消化过程中需严格控制胰蛋白酶浓度及作用时间。

常用的淬灭方式有: (1) 瞬间将样品温度改变至低温(-40℃)或高温(>80℃), 如液氮速冻法^[32]; (2) 加入有机溶剂, 如冷/热甲醇、热乙醇、冷乙腈等^[33-35]。Canelas等^[36]最初用60%的冷甲醇(-40℃)淬灭酵母胞内代谢物, 该方法被广泛应用于微生物代谢组学研究领域。然而, 在此过程中, 原核细胞的细胞膜会因为冷

应激而受到破坏, 从而导致细胞内代谢物的泄露。随后, 一系列缓冲剂被添加到冷甲醇中来防止泄露^[37]。最常用的缓冲剂有羟乙基哌嗪乙硫磺酸(HEPES)^[38]、碳酸氢铵^[39]以及氯化钠^[40]。Kapoor等^[41]对5种清洗细胞和淬灭方法进行了比较, 通过检测细胞基质内的ATP和细胞形态的变化, 发现用PBS清洗细胞, 60%的甲醇(添加70 mM HEPES)淬灭细胞, 泄露程度最小。在细胞淬灭过程中代谢物泄露是在所难免的, Wurm等^[42]研究表明, 当细胞膜在采样、清洗或淬灭过程中被破坏5%时, 90%的葡萄糖会在1 s内泄露。为了尽可能减少细胞处理过程中代谢物的泄露, 简化操作步骤, Ritter等^[43]直接将萃取剂加入培养皿中淬灭细胞, 同时对代谢物进行提取。在细胞淬灭过程中应尽量将泄露的影响降到最低, 并注意平行操作、快速淬灭、迅速收集样品。

2.3.2 细胞代谢物提取

细胞内的代谢反应通过淬灭停止后, 需要将代谢物从细胞中提取出来。由于细胞代谢组学研究要求尽可能保留和体现胞内代谢物谱信息, 一般提取时应遵循以下原则: (1) 以适当的回收率从细胞中提取大量的代谢产物; (2) 代谢物降解控制在最小范围内; (3) 防止代谢物泄漏; (4) 尽可能减少对代谢物结构的破坏。在细胞代谢组学领域中, 应用最广的是传统的溶剂提取法, 常用的提取方法有酸/碱提取、冷/热有机溶剂提取、两相萃取等; 根据所要研究代谢物的极性不同, 通常用热乙醇、高氯酸或甲醇、氯仿、环己烷等分别提取亲水性或亲脂性代谢物^[44]。目前, 以甲醇为代表的有机溶剂提取法在代谢组学分析中应用最为广泛, 具有操作简便、提取效率高、提取物易浓缩等优点。Dettmer等^[45]以结肠癌细胞株为研究对象, 比较了7种不同提取方法, 对具有代表性的胞内代谢物进行定量分析, 结果表明以80%的甲醇为溶剂提取出的代谢物最多。为了提高代谢物的提取效率和覆盖率, 常采用混合溶剂, 同时提取水溶性及脂溶性代谢物; Fei等^[46]运用甲醇-乙醇-水(2:2:1, v/v/v)提取胞内代谢物, 该方法对不同种类(革兰氏阳性/阴性菌及哺乳动物细胞)和不同方式培养的细胞(贴壁/悬浮)胞内极性和非极性代谢物均有较高的覆盖率, 相比传统亲水性和亲脂性代谢物分别提取方法, 该法具有高通量、无偏性的特点。代谢物的提取量不仅取决于提取溶剂的使用, 在提取过程中细胞的破裂程度也是决定因素,

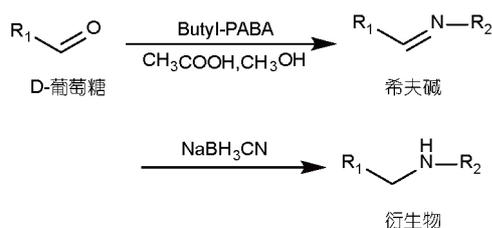


图3 D-葡萄糖还原胺化反应^[28,29]

因此可通过超声破碎、微波辅助萃取、机械匀浆等方法增加细胞的破碎程度,提高代谢物提取量.虽然目前文献报道的提取方式多样,但由于细胞内代谢物具有不同的物理化学性质(如分子量、极性、溶解性等),特定的提取条件通常仅适合某些类代谢物的提取,或只对某一类细胞有效.因此,在研究工作中,应根据研究目的选择合适的提取方法,并对提取溶剂、温度、pH等条件进行优化.

2.4 天然产物提取物样品的制备

天然产物中含有结构类型丰富的各类化学成分,有效成分的提取、分离和纯化是天然产物研究的重要环节.首先,需选择合适的方法将有效成分尽可能多得提取出来.常用的提取方法有溶剂提取法、水蒸气蒸馏法、升华法、压榨法等,其中溶剂提取法的用途最广;溶剂提取法根据“相似相溶”原理,结合目标化合物亲水性/亲脂性程度的不同,选择合适的溶剂进行提取,如糖类、氨基酸、无机盐等水溶性成分常用水或含水醇提取,蒽醌、黄酮、香豆素、游离生物碱等用氯仿、乙醚等亲脂性有机溶剂提取,挥发油、叶绿素、三萜类化合物用石油醚提取.从天然产物中寻找未知有效成分或有效部位时,常按提取溶剂极性递增的方式,用不同溶剂依次提取,得到不同部位提取物.此法得到的提取物多为混合物,常需进一步分离与纯化;常用的分离纯化技术有:系统溶剂分离法、两相溶剂萃取法、沉淀法、结晶法和色谱法等,其中液相色谱是目前应用较广泛的分离与纯化技术.传统方法研究天然产物中活性成分,通常存在制备效率低、耗时、有效成分损失多等缺点;近年来,有研究工作者用超高效液相色谱-高分辨质谱联用(UPLC-HRMS)技术,结合代谢组学多变量分析方法研究天然产物中的活性成分,避免了繁琐复杂的分离和纯化过程,粗提取物经萃取、离心或过滤等简单处理后可直接用于LC-MS分析^[47];Li等^[48]研究了姜科植物益智中抗帕金森病的活性成分,通过提取、分馏和“组学”分析,从益智仁提取物中发现抗帕金森病的先导化合物,并结合核磁共振和X射线晶体衍射对发现的此化合物进行了结构鉴定.

3 液相萃取与固相萃取

对于混合物样品中某一类化合物的靶向分析,

分离纯化步骤是必不可少的.液-液萃取(liquid-liquid extraction, LLE)或固相萃取(solid-phase extraction, SPE)是大部分混合物样品通用的前处理技术,不仅能够富集目标化合物,还可以有效降低LC-MS分析过程中基质效应的影响.LLE又称溶剂萃取,利用目标化合物在不同溶剂中溶解度或分配系数不同,将目标化合物从原溶剂转移到萃取剂中,达到分离和纯化的目的.分析血液样品中极性较小的脂质化合物时,通常使用与水互不相溶的溶剂为萃取剂,去除样本中的蛋白和一些极性较大的化合物^[49];Pellegrino等^[50]采用甲醇-甲基叔丁基醚-氯仿(1.33:1:1, v/v/v)混合溶剂为萃取剂,同时提取出血清样品中的9类脂质(甾醇类、胆固醇酯、磷脂、鞘磷脂、甘油酯等),该方法开启了人类血清脂质分析在科学研究和临床试验中大规模应用的大门.分散液-液微萃取(dispersible liquid-liquid microextraction, DLLME)是在传统LLE基础上发展起来的新型萃取技术,集萃取和富集于一体,具有成本低、富集效率高、有机溶剂消耗少等优点;通过加入分散剂,提高有机萃取剂在水相中的分散,增加萃取剂与水相的接触面积,使目标化合物快速转移到萃取剂微滴中^[51].该技术已逐渐应用于复杂样本的分析,Fischella等^[52]用DLLME对全血进行处理,以氯仿为萃取剂、甲醇为分散剂,有效提取出血液样品中的可卡因及其代谢物、安非他命及其类似物、苯二氮卓类药物及其他治疗精神疾病的药物,所有分析物基质效应相对较低(对于低浓度样品,离子抑制不超过26%,离子增强不超过28%),表明样品纯化方法的有效性.

SPE是20世纪70年代发展起来的样品预处理技术,与液-液萃取相比前处理速度更快、有机溶剂用量少、提取效率高、重现性好.SPE技术运用液相色谱分离理论,在实际操作中需根据目标化合物与固定相相互作用力的大小,选择合适的SPE柱及洗脱方式.Gonzalez-Dominguez等^[17]采用稀释法、液液萃取和固相萃取等不同方法处理尿液,比较代谢物指纹图谱的变化,表明使用50%甲醇-水(v/v)稀释和混合模式吸附剂(mixed mode sorbent)固相萃取处理,代谢物覆盖范围更广.20世纪90年代在固相萃取技术基础上发展了固相微萃取(solid-phase microextraction, SPME)技术,该技术无需提取溶剂,利用SPME装置提取头表面不同性质的吸附剂,对不同种类化合物进行萃取和浓缩;采样时只需将提取头浸入样品或置于样品上部空间,

提取目标化合物^[53,54]。近年来, SPME技术已应用于代谢组学^[55,56]、环境监测^[57]、食物营养成分分析^[58]、药物分析^[59]等领域。Ahmad等^[60]将SPME装置与质谱联用, 样品无需进行其他分离纯化处理, 提取头上吸附的药物直接被气化、电离, 整个分析过程只需5 min, 此技术成功应用于血液样品中美托洛尔与普萘洛尔的提取。为了提高分离纯化效率, 逐渐发展了在线萃取技术, 如将SPE与LC或LC-MS或核磁共振(NMR)联用, 实现自动化、高效的样品处理与分析。

4 LC-MS数据采集

4.1 液相色谱条件选择与优化

液相色谱在复杂混合物体系分离分析中发挥着极其重要的作用, 根据流动相和固定相极性及其作用机制的不同, 液相色谱可分为正相色谱、反相色谱、离子交换色谱等^[61]。目前分析混合物样品, 常采用UPLC进行分离, UPLC是在传统的HPLC基础上发展起来的新型分离技术; 通过使用填料粒径更小的色谱柱(粒径 $< 2 \mu\text{m}$), 超高压输液泵, 提高分离效率、分离度, 增加分析通量、灵敏度及色谱峰容量^[62]。

4.1.1 固定相(色谱柱)的选择

色谱柱是液相色谱的核心部件, 在分离分析中起关键作用。目前应用最广的是反相液相色谱柱, 它是表面键合疏水基团的硅胶为固定相, 适合于中等极性或弱极性化合物的分离; 正相色谱柱通常以硅胶、氧化铝、氨基键合硅胶等为固定相, 适用于极性化合物的分离; 对于大部分极性和亲水性物质的分离, 亲水作用色谱(hydrophilic interaction liquid chromatography, HILIC)是正相色谱的有力补充。在复杂样品分离中需根据目标化合物的亲水/疏水性质选择合适的色谱柱。例如, 天然产物粗提物中多为亲脂性成分, 常采用反相柱对其进行分离; Vlase等^[63]运用Zorbax SB-C₁₈ (Agilent, USA)色谱柱从豆科植物红三叶草和染料木提取物中分离出7种异黄酮类化合物; 本课题组也利用CORTECS[®]UPLC[®]C18 (Waters, USA)色谱柱对海棠果中降胆固醇成分进行分离。血液样品和细胞提取物中多为亲水性小分子代谢物, 常采用HILIC色谱柱分离此类物质。Prinsen等^[64]运用HILIC色谱柱对血浆中氨基酸进行快速分离, 在18 min内有效分离出36个氨基酸, 并

对其中24个氨基酸进行定量分析, 此法对于研究患有先天性氨基酸代谢缺陷症的病人体内氨基酸的代谢具有重要意义。在代谢组学研究中常联合使用不同性质的色谱柱, 以提高可分析代谢物的覆盖率^[65]。例如, 本课题组利用具有亲水性表面的Zorbax SB-Aq (Agilent, USA)色谱柱和Zorbax HILIC Plus (Agilent, USA)非键合硅胶色谱柱分离血浆和细胞提取物中小分子代谢物。

4.1.2 流动相优化

在液相色谱分离中, 流动相组成和洗脱模式是影响分离效果的关键因素, 需要根据样品性质(如极性、酸碱性、溶解度、稳定性等)和色谱柱类型选择合适的洗脱条件。反相色谱常采用甲醇、乙腈、水为流动相, 其中乙腈洗脱能力较强; 亲水作用色谱使用与反相色谱的流动相组成相似的水-有机溶剂体系为流动相, 初始条件通常为高比例的有机相(如95%的乙腈), 物质的洗脱顺序与反相色谱相反, 随着水相比比例的增加, 洗脱能力增强。对于复杂样品的分离, 为了使目标物与杂质有效分离, 常采用梯度洗脱方式, 在一个分析周期内, 按一定程序改变流动相的浓度配比。Ma等^[66]采用反相洗脱系统, 以乙腈-水为流动相, 选择线性梯度洗脱方式, 实现3 min时间内有效分离大鼠血浆中安非他命、美托洛尔、非那西汀、奥美拉唑等6种药物。

分析样品中酸性或碱性物质时, 常在流动相中加入酸、碱或缓冲盐作为改性剂(modifier); 通过调节流动相pH, 抑制目标化合物解离, 提高待测物分离过程中的均一性, 能够有效改善色谱峰形, 提高分析的重现性。在液相分离中常以离子对试剂三氟乙酸(trifluoroacetic acid, TFA)作为改性剂改善拖尾、克服峰展宽, 但由于其极易电离, 在负离子模式下抑制目标分析物负离子的形成, 降低离子化效率^[67]; 在正离子模式下, 带正电的分析离子易与TFA形成离子对, 也会降低离子化程度^[68]。因此, 在LC-MS分析中应尽可能降低TFA的浓度或用甲酸/乙酸代替; 在使用缓冲盐作改性剂时, 常使用甲酸铵、醋酸铵等挥发性的盐, 且需严格控制浓度(一般不超过20 mM); 不易挥发的盐(如磷酸盐、氯化钠等)需慎用, 由于其不易气化, 容易造成质谱离子源污染。

4.2 质谱条件优化与数据采集

质谱仪以离子源、质量分析器、检测器为主要

部件. 常用的大气压电离技术包括电喷雾电离(electrospray ionization, ESI)、大气压化学电离(atmospheric pressure chemical ionization, APCI)和大气压光致电离(atmospheric pressure photoionization, APPI), 这3种离子化方式相互补充; ESI为软电离技术, 能够提供分子离子峰或准分子离子峰信息, 适合于分析中等极性和极性化合物, 是目前适用范围最广的离子化方式; APCI和APPI主要适合于分析非极性小分子化合物.

根据质谱分辨率的不同, 可将其分为高分辨和低分辨质谱仪, 其中高分辨质谱指的是分辨率大于10000的质谱^[69], 常用的高分辨质谱有飞行时间质谱(time-of-flight, TOF)、静电场轨道阱质谱(orbitrap)、傅里叶变换离子回旋共振质谱(Fourier transform ion cyclotron resonance, FTICR). 在复杂样品分析中, 常使用高分辨串联质谱(tandem mass spectrometry)技术, 采集二级(MS/MS)或多级质谱(MSⁿ)数据, 既可以为目标化合物及碎片离子提供精确质荷比(误差小于10 ppm)和可能的元素组成信息, 又能够获得目标物的裂解信息, 用于已知化合物的定量分析(MRM)和未知化合物的结构解析. 常用的串联质谱包括: 四极杆飞行时间质谱(quadrupole time-of-flight, Q-TOF)、三重四极杆质谱(triple quadrupole, QqQ)、四极杆离子阱质谱(quadrupole ion trap, QIT)等. 与传统的LC-MS分析方法相比, LC-MS/MS具有较高的特异性, 能更好地消除基质的干扰, 获得高信噪比(S/N)的质谱信号. 本实验室使用的MicrOTOF-QII质谱仪(Bruker, Germany), 可选择多反应监测(multiple reaction monitoring, MRM)模式进行MS/MS定量分析. MRM技术是在选择反应监测(selective reaction monitor, SRM)技术的基础上演化而来的, 可设定多个母离子, 交替进行二级质谱信号采集; 此外, 自动二级质谱(Auto MS/MS)模式可自动选择信号强度超过设定阈值的前体离子进行MS/MS分析; 在已知目标化合物保留时间的情况下, 编辑Auto MS/MS标签下的预设前体列表(scheduled precursor list, SPL), 能够实现不同保留时间针对特定的母离子, 自动进行一级和二级质谱数据采集. 对于包含许多重叠信号的复杂生物样品中低丰度的前体离子, SPL的应用能够提高MS/MS扫描覆盖率^[70].

质谱数据采集时首先要针对目标化合物优化质谱条件. 由于不同品牌和型号质谱仪的物理条件如真空度、离子传输管路尺寸、电压等不同, 参数设置往

往存在差异, 难以相互直接借鉴, 需根据实际情况优化调节质谱参数. 针对MicrOTOF-QII质谱仪, 质谱参数优化项主要包括离子源和离子传输部分各种参数: 雾化气压、干燥气流速及温度、离子传输部分的电压以及碰撞能等. 离子源部分参数主要影响样品的离子化效率, 即信号的强度; 离子传输部分参数影响传输过程中对离子的选择性. 通常以对照品溶液为分析对象, 采用注射泵恒速直接进样方式, 首先扫描一级质谱, 确定母离子质荷比; 以目标离子的响应强度及稳定性为标准, 通过实时监测, 调节各参数对信号强度的影响, 确定参数设置. 一般情况下, 被分析物在正、负离子模式下的信号响应会有所不同, 通常选择正、负离子模式同时采集以得到更多质谱信息. 进行MS/MS分析时, 碰撞能是需要优化的关键参数, 碰撞能设置过低, 化合物不能有效裂解, 碎片离子峰信号强度偏低; 反之, 碰撞能设置过高, 会产生离子过度碎片化现象, 目标碎片离子容易进一步裂解, 同样会导致目标碎片离子峰信号强度偏低. 碰撞能优化过程中通常设定多个扫描段、每个扫描段设定不同的碰撞能, 获得不同能量下的碎片离子信息. 由于不同品牌或型号仪器真空度的不同, 会引起离子传输速度的差异, 使得最终到达碰撞池的离子动能不尽相同; 此外, 碰撞池氦气气压的微小变化, 也影响碰撞裂解过程, 这些因素均会导致最终碰撞裂解产生的子离子碎片指纹的差异. 在实际实验过程中, 通常需根据特定的质谱仪器进行个性化的碰撞能优化.

5 LC-MS数据(前)处理

针对复杂混合物中的靶向成分进行定量分析, 即从混合物体系中检测分析某些特定的已知化合物, 常采用提取离子色谱图(extracted ion chromatograms, EIC), 通过从总离子色谱图(total ion chromatogram, TIC)中提取某个目标离子的色谱图, 排除不同组分之间及背景的干扰, 即使在色谱峰未能完全分离的情况下, 利用质谱的高分辨性能也可以对特定的组成进行较为准确的定量分析. 一般可直接利用商业质谱仪器自带的软件(如DataAnalysis、Xcalibur、MassLynx等)完成相关数据前处理, 如背景扣除、峰平滑、积分等.

各种组学分析方法, 如蛋白组学、代谢组学等分析方法, 常被用于复杂混合物体系多个样本的LC-MS数据的分析, 用于生物标志物或差异代谢物的发现.

“组学”技术一般数据较为复杂, 包含的信息量较大, 因而数据分析与信息挖掘是组学技术的关键点之一. 本文主要阐述基于LC-MS的非靶向代谢组学数据处理和分析. 常见的质谱仪器自带的商业LC-MS数据分析软件大多不具备信号峰对齐功能, 且无法实现自动化批量数据处理, 因此, 非靶向代谢组学数据分析多采用专业的代谢组学数据处理软件, 如XCMS^[71,72]、MZmine 2^[73]、ApLCMS^[74]、MAVEN^[75]等. 代谢组学数据处理主要包括数据前处理和数据变换步骤. 数据前处理是代谢组学数据处理的一个关键步骤, 由降噪与基线校正(背景扣除)、峰识别与积分、峰对齐、峰补齐构成. 噪音和背景信号主要来自仪器干扰(如检测器)和分析系统, 如流动相和缓冲盐的组成和浓度的变化引起的基线漂移. 降噪与基线校正(背景扣除)是将分析物信号与噪音及背景信号分开. 随后需要从降噪与扣除背景信号的数据中识别分析物信号, 并获取相关的定量信息(峰强度, 即积分面积或高度). 不同数据处理软件包所采用的峰识别算法各不相同. 例如, XCMS有两种峰识别算法: 匹配过滤(matched filter)和连续小波变换(centWave)^[71,72]. 匹配过滤算法先在设定的质荷比范围内提取离子基峰色谱图(extracted ion base-peak chromatogram, EIBPC), 在色谱维度, 再以高斯函数二阶导数为峰模型进行匹配和过滤和色谱峰识别. centWave算法采用连续小波变换检测不同峰宽的峰, 确定峰范围(峰的起点和终点), 取质心峰(centroid peak)的最大值作为峰的强度. MZmine 2的峰识别算法包括精确质量测定、构建色谱图、色谱峰去卷积^[73]. Gurdeniz等^[76]运用MarkerLynx、MZmine、XCMS 3种不同软件处理血浆样品LC-MS数据, 比较正/负离子模式下获得特征的数目, 由于不同软件峰识别算法不同, 仅有37%~46%的共同提取特征(图4).

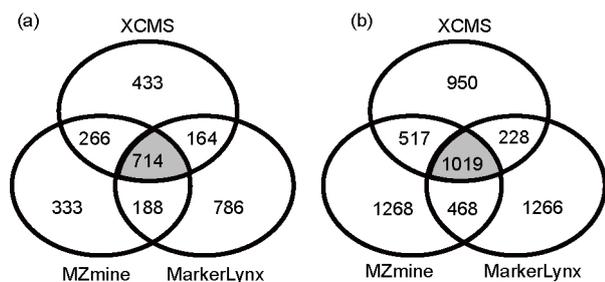


图4 (a) 负离子模式3种软件提取特征数目; (b) 正离子模式3种软件提取特征数目^[76]

由于各种环境、仪器条件的波动以及样品条件的变化, 用于代谢组学研究的多个样本的LC-MS信号峰通常会在色谱维度发生样品间保留时间相对漂移现象. 因此, 进行代谢组学数据多变量分析前一般需要先对多个样本的LC-MS数据进行信号峰匹配和保留时间对齐处理. 通常可先利用特定内标统一校正所有样品的保留时间^[77,78]. 针对某些信号峰保留时间发生局部非线性漂移^[71,79], 已开发有许多不同的对齐算法用于保留时间的非线性校正^[80,81]; Lange等^[82]比较了6种不同的保留时间对齐算法, 其中XCMS处理代谢组数据的效果较好. XCMS的对齐算法首先鉴别数百(千)个性质良好的峰组(peak group)作为对齐标记(markers)用于其他检出特征(质荷比/保留时间)的对齐. 一般从代谢组样品中鉴别的标记峰组的数目可足以覆盖样品的色谱图并正确校正非线性保留时间漂移^[71]. 在实际的代谢组数据处理过程中, 部分样品可能不能检测到某些LC-MS数据特征(质荷比/保留时间), 导致代谢组学数据矩阵出现空缺数据点. 为了实现代谢组学数据矩阵的完整性, 需要对多个样本的数据进行峰补齐处理. 通常可根据设定的阈值(50%、80%, 即所谓的“五五”规则、“二八”规则)来筛选判定是否保留某些峰组(特征)并进行补齐处理. XCMS默认阈值为50%, 即50%或以上的样品都检出同一特征时, 将保留该峰组并进行峰补齐处理. 峰补齐(gap-filling)处理对峰组中未检测到该特征的样品在相应的峰区域对原始数据进行积分创建一个峰强度值(面积), 以获得完整的数据矩阵, 用于后续的多变量统计分析. 经上述前处理步骤后, 可获得一个完整的代谢组数据矩阵, 其行表示样品(即谱图), 列表示变量(特征). 值得强调的是, 在数据前处理过程中需要结合实验数据及相应的数据处理方法, 先对相关参数进行调试优化, 方可进行批量化数据处理.

6 目标化合物的定性定量分析

针对特定目标化合物的定性分析, 首先可根据高分辨质谱提供精确分子量预测其元素组成和分子式, 结合其同位素分布模式信息(包含各同位素峰的精确分子量以及相对强度比例)进一步加以确证. 以MicroTOF-Q II为例, 测定化合物分子量的精度可以达到 10^{-4} Da, DataAnalysis 4.1 (Bruker)数据分析软件包中的SmartFormula通过联合利用精确分子量和同位素

分布模式信息, 计算得到许多可能的分子式, 可进一步结合氮规则、电子数以及先验的化学知识等信息进行进一步筛选和排除, 得到候选分子式. 精确的分子量和真实同位素分布模式(true isotopic pattern, TIP)是用以判断候选分子式似然性的关键因素; 实际测定的同位素分布模式与基于候选分子式的理论同位素分布模式的差距用mSigma值来表示, 该值越小, 说明二者越接近, 候选分子式的似然性越高^[83]. 单纯依靠高分辨质谱获得的精确分子量, 预测结果往往会得到数十个(甚至更多)的可能分子式; 结合mSigma值判定, 可很大程度上减少候选分子式的数目. 进一步利用MS/MS碎片离子信息, 运用SmartFormula3D™分析可进一步缩小候选分子式的范围甚至得到唯一的分子式. 利用CompoundCrawler™ (Bruker)对分子式进行数据库检索, 可获得该化合物可能的分子结构.

复杂系统的研究不仅需要对其组分进行定性分析, 还需要对某些目标化合物进行定量测定. LC-MS是基于质谱信号峰强度进行定量分析. 目标化合物质谱信号峰强度除了与其浓度相关外, 还会受到基质效应的影响, 即与目标化合物共洗脱的其他物质成分(包括流动相溶剂等)会影响其离子化几率, 不同基质会对目标化合物的电离起抑制或增强作用, 影响定量分析的准确性^[84-86]. 因此, 液质联用定量分析常采用内标法, 选用与目标化合物具有类似结构的化合物或具有相同分子结构的稳定同位素(如¹³C、¹⁵N、¹⁷O等)标记的化合物作为内标物. 以同位素作内标能够显著性提高定量测定的准确性和精确度. 对于复杂体系中已知目标化合物的定量分析常使用同位素稀释质谱法(isotope dilution mass spectrometry, IDMS), 根据目标化合物与其同位素标记物质荷比的差别, 选择提取离子色谱峰进行分析, 分别对目标化合物和同位素标记物色谱峰进行积分, 计算色谱峰面积比值, 确定目标物含量^[87,88]. IDMS方法可以有效降低基质效应的影响, 同时可以减少样品处理及仪器不稳定性造成的影响, 并且IDMS用于复杂体系中微量、痕量成分的定量测定, 具有灵敏度高、准确度好的优点^[87]. 但同位素标记物作为内标, 成本较高且不易获得. 为了解决这个难题, 研究者设计、合成了一系列稳定同位素编码衍生化(isotope-coded derivatization, ICD)试剂, 采用同位素编码衍生技术, 使具有某一类官能团的物质通过衍生化反应同时获得同位素内标物. 此外, 同位素编码衍生

化法通常能够改善目标化合物的色谱分离效果, 提高质谱检测灵敏度^[89]; Zhang等^[90]以自主合成的非标与重氢标记的10-甲基-吡啶酮-2-磺酰氯(*d*₀/*d*₃-MASC)为衍生化试剂, 对环境和生物样品中雌激素类化合物进行定量分析. LC-MS定量分析常采用二级质谱MRM模式扫描, 选择丰度最强、最稳定的子离子进行定量分析; MRM相比LC-MS一级质谱定量分析具有更高的选择性, 使用此方法进行定量时, 碰撞能的设置显得尤为重要, 应避免碰撞能过低, 由于裂解不够导致子离子强度过低, 或碰撞能过高引起过度碎片化导致目标子离子强度过低, 子离子信号强度将影响定量分析的准确性和线性范围. 已有报道研究者运用IDMS结合MRM数据采集模式, 实现对地高辛^[91]、洋地黄毒苷^[91]、氨基酸^[92]的准确定量.

7 多样本数据多变量分析

多样本代谢组学研究通常采用多变量分析以揭示多个样本间的相互关系(分组)以及各个不同变量对观察结果的贡献或与观察结果的相关性. 在多变量分析前, 一般需要对数据进行变换处理, 主要包括归一化(normalization)和标准化(scaling)等. 仪器稳定性、基质效应等因素造成的电喷雾离子化过程中离子化几率的变化, 或样品制备过程中样品浓度的差异, 均会改变检测到的质谱信号强度. 为了鉴别发现真实的与实验处理因素相关的代谢物含量的变化, 一般需要对数据作归一化处理. 归一化主要对数据矩阵的行作处理, 校正整体信号强度的变化或波动, 已有文献报道了代谢组数据各种不同的归一化方法和策略^[93,94]. 标准化处理主要对数据矩阵的列作处理, 如自标准化(autoscaling), 即每一列变量减去其平均值并除以该列的标准偏差^[95], 标准化处理可以改变或校正不同变量对观察结果贡献或相关性的权重. 经数据前处理和数据转化后, 为了降低数据的复杂性(维度)和鉴别数据中显著相关性的变量, 通常对数据进行多变量分析. 两种常用的多变量分析方法为主成分分析(principal component analyses, PCA)和偏最小二乘法(partial least squares, PLS)^[96,97]. PCA对数据不作分组, 通过散点得分(score)图和荷载(loadings)图可观察样本数据的分布趋势及逸出值(奇异样本点), 辨识数据中的最大方差来源. PLS方法假定观察结果的变化是由一组隐(藏)变量所驱动, 隐变量可抽象为实验的处理因素/刺激条

件等. PLS分析可鉴别数据中用于描述用户定义分组之间最大协方差的特征矢量^[98], 该特征矢量在原始测量数据的变量(特征, 质荷比/保留时间)空间轴上的投影大小可表示原始特征(质荷比/保留时间)与分组的相关性, 可用于指导生物标志物或差异代谢物的发现. 两种处理方法的根本区别为: PLS是用户定义分组的有监督分析方法, 而PCA是一种无监督的单纯减少变量(降维)方法.

8 总结与展望

随着中药学、生命科学、食品科学、营养科学和环境科学等的快速发展, 越来越多的领域涉及复杂混合物体系的定性定量分析. 样品制备是复杂混合物分析的首要 and 关键环节, 也是制约分析化学发展与应用的瓶颈问题; 为了减小分析误差、提高分析的重现性、保证分析结果的准确性, 应尽可能简化样品处理步骤, 发展程序化与标准化的样品制备方法, 开发自动化的样品制备平台. 现代分离与分析技术的发展, 使复杂混合物体系的分析有了质的飞跃. 例如, 固相萃取-液相色谱-质谱联用(SPME-LC-MS)技术, 集采样、萃取、检测于一体, 简化样品制备过程, 大大缩短整体分析时间^[99]; 针对复杂样品的快速分析, 可采用二维液相色谱(2D-LC)技术, 通过联合使用不同分离机理的色谱柱, 显著提高复杂混合物的分离能力及峰容量, 降低色谱峰重叠, 提高色谱峰鉴定的可靠性^[100]; 对于热不稳定和低挥发性化合物的分离, 超临界流体色谱(supercritical fluid chromatography, SFC)是液相色谱和气相色谱的有力补充, 并成为手性化合物立体异构体拆分的首选方法^[101].

由于复杂体系组成成分多样, 包含许多微量甚至痕量的化合物成分, 通常难以被检测到, 微量或痕量未知化合物的结构难以确定. 质谱(尤其结合纳喷技术)一方面具有较高的检测灵敏度, 另一方面, 基于高

分辨质谱获得的精确分子量以及同位素分布模式信息, 结合串联质谱获得的二级(或多级)质谱碎片指纹包含的丰富的结构信息, 能够帮助微量或痕量未知化合物的结构鉴定. 但由于不同品牌及型号质谱仪之间各种物理条件(如传输管路真空度、各种加速电压、碰撞池氦气气压等)的不同, 导致真实的离子碰撞能的绝对大小难以精确控制, 采集的二级(或多级)质谱碎片指纹往往存在差异. 由于不同质谱仪器产生特异性的质谱碎片信息, 限制了二级(或多级)质谱数据(指纹图谱)的相互借鉴和利用, 不利于建立一套通用性的二级(或多级)质谱数据库用于未知化合物的鉴定. 通过设定梯度碰撞能(collision energy ramp)^[102], 在二级(或多级)质谱数据采集过程中通过扫描改变碰撞能的大小并对数据进行加和, 不同仪器能够获得较为一致性的二级(或多级)质谱数据, 可用于建立一套通用性的二级(或多级)质谱数据库用于未知化合物的鉴定. 离子淌度质谱技术(ion mobility spectrometry-mass spectrometry, IMS-MS)利用具有相同分子式但结构不同的同分异构体的离子与缓冲气体(buffer gas)的碰撞截面积(collision cross-section, CCS)不同能够对同分异构体进行分离和识别鉴定, 为化合物的结构确证提供更为精细的结构信息^[103].

近年来发展起来的敞开式离子化(ambient ionization, AI)技术, 在常压、开放环境下对样品进行离子化. 例如, 质谱成像技术, 具有无需样品前处理、实时原位、低损耗等优点, 展现了质谱分析技术发展的重要趋势, 已逐渐应用于复杂体系样品的分析, 适合于微量目标化合物高通量鉴别和检测^[104]. 随着高时空分辨的质谱仪器以及相关联用技术的发展, 实验数据量通常越来越大, 包含的信息量也越来越多, 如何从大量原始数据中获取有效信息是分析化学面临的一项新的挑战, 这将推动复杂体系数据分析方法的开发及化学计量学的发展和在复杂混合物体系分析中的应用.

参考文献

- 1 Whitesides GM, Ismagilov RF. *Science*, 1999, 284: 89–92
- 2 Weng G, Bhalla US, Iyengar R. *Science*, 1999, 284: 92–96
- 3 Koch C, Laurent G. *Science*, 1999, 284: 96–98
- 4 Mohamed KM, Cromarty D, Steenkamp V. *J Chromatogr B*, 2015, 997: 1–6
- 5 Han J, Liu Y, Wang R, Yang J, Ling V, Borchers CH. *Anal Chem*, 2015, 87: 1127–1136
- 6 Roci I, Gallart-Ayala H, Schmidt A, Watrous J, Jain M, Wheelock CE, Nilsson R. *Anal Chem*, 2016, 88: 2707–2713

- 7 Mi S, Zhao YY, Dielschneider RF, Gibson SB, Curtis JM. *J Chromatogr B*, 2016, 1031: 50–60
- 8 Yang Z, Wu Y, Wu S. *J Chromatogr A*, 2016, 1431: 184–196
- 9 de Villiers A, Venter P, Pasch H. *J Chromatogr A*, 2016, 1430: 16–78
- 10 Li Z, Wang Y, Ouyang H, Lu Y, Qiu Y, Feng Y, Jiang H, Zhou X, Yang S. *J Chromatogr B*, 2015, 988: 45–52
- 11 Al-Taher F, Cappozzo J, Zweigenbaum J, Lee HJ, Jackson L, Ryu D. *Food Control*, 2017, 72: 27–35
- 12 Lendoiro E, Jiménez-Morigosa C, Cruz A, Páramo M, López-Rivadulla M, de Castro A. *Drug Test Anal*, 2017, 9: 96–105
- 13 Leung KSY, Fong BMW. *Anal Bioanal Chem*, 2014, 406: 2289–2301
- 14 Tulipani S, Llorach R, Urpi-Sarda M, Andres-Lacueva C. *Anal Chem*, 2013, 85: 341–348
- 15 Nouman EG, Al-Ghobashy MA, Lotfy HM. *J Chromatogr B*, 2015, 989: 37–45
- 16 Wong MCY, Lee WTK, Wong JSY, Frost G, Lodge J. *J Chromatogr B*, 2008, 871: 341–348
- 17 González-Domínguez R, Castilla-Quintero R, García-Barrera T, Gómez-Ariza JL. *Anal Biochem*, 2014, 465: 20–27
- 18 Yin P, Lehmann R, Xu G. *Anal Bioanal Chem*, 2015, 407: 4879–4892
- 19 Minami Y, Kasukawa T, Kakazu Y, Iigo M, Sugimoto M, Ikeda S, Yasui A, van der Horst GTJ, Soga T, Ueda HR. *Proc Natl Acad Sci USA*, 2009, 106: 9890–9895
- 20 Brauer R, Leichtle AB, Fiedler GM, Thiery J, Ceglarek U. *Metabolomics*, 2011, 7: 344–352
- 21 Simundic AM, Cornes M, Grankvist K, Lippi G, Nybo M. *Clin Chim Acta*, 2014, 432: 33–37
- 22 Ellervik C, Vaught J. *Clin Chem*, 2015, 61: 914–934
- 23 Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB, Goodacre R, Goodacre R. *Nat Protoc*, 2011, 6: 1060–1083
- 24 Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, Lucio M, Zell A, Haring HU, Xu G, Lehmann R. *Clin Chem*, 2013, 59: 833–845
- 25 Lippi G, Salvagno GL, Montagnana M, Brocco G, Guidi GC. *Clin Chem Lab Med*, 2006, 44: 311–316
- 26 Fliniaux O, Gaillard G, Lion A, Cailleu D, Mesnard F, Betsou F. *J Biomol NMR*, 2011, 51: 457–465
- 27 Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S, Guy PA. *Anal Chem*, 2009, 81: 3285–3296
- 28 Zhang H, Stevens RD, Young SP, Surwit R, Georgiades A, Boston R, Millington DS. *Clin Chem*, 2009, 55: 527–532
- 29 Dalpathado DS, Jiang H, Kater MA, Desaire H. *Anal Bioanal Chem*, 2005, 381: 1130–1137
- 30 Kapoore RV, Coyle R, Staton CA, Brown NJ, Vaidyanathan S. *Metabolomics*, 2015, 11: 1743–1755
- 31 Teng Q, Huang W, Collette TW, Ekman DR, Tan C. *Metabolomics*, 2009, 5: 199–208
- 32 Panopoulos AD, Yanes O, Ruiz S, Kida YS, Diep D, Tautenhahn R, Herreras A, Batchelder EM, Plongthongkum N, Lutz M, Berggren WT, Zhang K, Evans RM, Siuzdak G, Izipisua Belmonte JC. *Cell Res*, 2012, 22: 168–177
- 33 Paglia G, Hrafnisdóttir S, Magnúsdóttir M, Fleming RMT, Thorlacius S, Palsson BÖ, Thiele I. *Anal Bioanal Chem*, 2012, 402: 1183–1198
- 34 Stuani L, Lechaplais C, Salminen AV, Ségurens B, Durot M, Castelli V, Pinet A, Labadie K, Cruveiller S, Weissenbach J, de Berardinis V, Salanoubat M, Perret A. *Metabolomics*, 2014, 10: 1223–1238
- 35 Wu Y, Li L. *Anal Chem*, 2013, 85: 5755–5763
- 36 Canelas AB, Ras C, ten Pierick A, van Dam JC, Heijnen JJ, van Gulik WM. *Metabolomics*, 2008, 4: 226–239
- 37 Faijes M, Mars AE, Smid EJ. *Microb Cell Fact*, 2007, 6: 27
- 38 Taymaz-Nikerel H, de Mey M, Ras C, ten Pierick A, Seifar RM, van Dam JC, Heijnen JJ, van Gulik WM. *Anal Biochem*, 2009, 386: 9–19
- 39 Sellick CA, Hansen R, Maqsood AR, Dunn WB, Stephens GM, Goodacre R, Dickson AJ. *Anal Chem*, 2009, 81: 174–183
- 40 Dietmair S, Timmins NE, Gray PP, Nielsen LK, Krömer JO. *Anal Biochem*, 2010, 404: 155–164
- 41 Kapoore RV, Coyle R, Staton CA, Brown NJ, Vaidyanathan S. *Analyst*, 2017, 142: 2038–2049
- 42 Wurm M, Zeng AP. *Metabolomics*, 2012, 8: 1081–1089
- 43 Ritter JB, Genzel Y, Reichl U. *Anal Biochem*, 2008, 373: 349–369
- 44 Sellick CA, Knight D, Croxford AS, Maqsood AR, Stephens GM, Goodacre R, Dickson AJ. *Metabolomics*, 2010, 6: 427–438
- 45 Dettmer K, Nürnberger N, Kaspar H, Gruber MA, Almstetter MF, Oefner PJ. *Anal Bioanal Chem*, 2011, 399: 1127–1139
- 46 Fei F, Bowdish DME, McCarry BE. *Anal Bioanal Chem*, 2014, 406: 3723–3733
- 47 Cuperlovic-Culf M, Culf AS. *Expert Opin Drug Dis*, 2016, 11: 759–770
- 48 Li G, Zhang Z, Quan Q, Jiang R, Szeto SSW, Yuan S, Wong W, Lam HHC, Lee SMY, Chu IK. *J Proteome Res*, 2016, 15: 2595–2606
- 49 Cajka T, Fiehn O. *TrAC Trends Anal Chem*, 2014, 61: 192–206
- 50 Pellegrino RM, Di Veroli A, Valeri A, Goracci L, Cruciani G. *Anal Bioanal Chem*, 2014, 406: 7937–7948
- 51 Spietelun A, Marcinkowski Ł, de la Guardia M, Namieśnik J. *Talanta*, 2014, 119: 34–45
- 52 Fisichella M, Odoardi S, Strano-Rossi S. *MicroChem J*, 2015, 123: 33–41

- 53 Spietelun A, Marcinkowski Ł, de la Guardia M, Namieśnik J. *J Chromatogr A*, 2013, 1321: 1–13
- 54 Boyaci E, Rodriguez-Lafuente Á, Gorynski K, Mirnaghi F, Souza-Silva ÉA, Hein D, Pawliszyn J. *Anal Chim Acta*, 2015, 873: 14–30
- 55 Bojko B, Reyes-Garcés N, Bessonneau V, Goryński K, Mousavi F, Souza Silva EA, Pawliszyn J. *TrAC Trends Anal Chem*, 2014, 61: 168–180
- 56 Zhang QH, Zhou LD, Chen H, Wang CZ, Xia ZN, Yuan CS. *TrAC Trends Anal Chem*, 2016, 80: 57–65
- 57 Xu J, Chen G, Huang S, Qiu J, Jiang R, Zhu F, Ouyang G. *TrAC Trends Anal Chem*, 2016, 85: 26–35
- 58 Xu CH, Chen GS, Xiong ZH, Fan YX, Wang XC, Liu Y. *TrAC Trends Anal Chem*, 2016, 80: 12–29
- 59 Kataoka H, Ishizaki A, Saito K. *Anal Methods*, 2016, 8: 5773–5788
- 60 Ahmad S, Tucker M, Spooner N, Murnane D, Gerhard U. *Anal Chem*, 2015, 87: 754–759
- 61 Qiu H, Liang X, Sun M, Jiang S. *Anal Bioanal Chem*, 2011, 399: 3307–3322
- 62 Ramachandra B. *Crit Rev Anal Chem*, 2017, 47: 24–36
- 63 Vlase L, Popa DS, Tero-Vescan A, Olah N. *Acta Chromatogr*, 2011, 23: 509–520
- 64 Prinsen HCMT, Schiebergen-Bronkhorst BGM, Roeleveld MW, Jans JJM, de Sain-van der Velden MGM, Visser G, van Hasselt PM, Verhoeven-Duif NM. *J Inherit Metab Dis*, 2016, 39: 651–660
- 65 李兴, 余玲玲, 胡凯锋. *生命科学仪器*, 2016, 14: 3–9
- 66 Ma J, Wang S, Zhang M, Zhang Q, Zhou Y, Lin C, Lin G, Wang X. *Biomed Chromatogr*, 2015, 29: 1203–1212
- 67 Gustavsson SÅ, Samskog J, Markides KE, Långström B. *J Chromatogr A*, 2001, 937: 41–47
- 68 Shou WZ, Naidong W. *J Chromatogr B*, 2005, 825: 186–192
- 69 Zacs D, Rjabova J, Pugajeva I, Nakurte I, Viksna A, Bartkevics V. *J Chromatogr A*, 2014, 1366: 73–83
- 70 Hoffmann T, Krug D, Hüttel S, Müller R. *Anal Chem*, 2014, 86: 10780–10788
- 71 Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. *Anal Chem*, 2006, 78: 779–787
- 72 Benton HP, Wong DM, Trauger SA, Siuzdak G. *Anal Chem*, 2008, 80: 6382–6389
- 73 Pluskal T, Castillo S, Villar-Briones A, Oresic M. *BMC BioInf*, 2010, 11: 395
- 74 Yu T, Park Y, Johnson JM, Jones DP. *Bioinformatics*, 2009, 25: 1930–1936
- 75 Melamud E, Vastag L, Rabinowitz JD. *Anal Chem*, 2010, 82: 9818–9826
- 76 Gürdeniz G, Kristensen M, Skov T, Dragsted LO. *Metabolites*, 2012, 2: 77–99
- 77 Dettmer K, Aronov PA, Hammock BD. *Mass Spectrom Rev*, 2007, 26: 51–78
- 78 Frenzel T, Miller A, Engel KH. *Eur Food Res Technol*, 2003, 216: 335–342
- 79 Podwojski K, Fritsch A, Chamrad DC, Paul W, Sitek B, Stühler K, Mutzel P, Stephan C, Meyer HE, Urfer W, Ickstadt K, Rahnenführer J. *Bioinformatics*, 2009, 25: 758–764
- 80 Perera V, De Torres Zabala M, Florance H, Smirnov N, Grant M, Yang ZR. *Metabolomics*, 2012, 8: 175–185
- 81 Bork C, Ng K, Liu Y, Yee A, Pohlscheidt M. *Biotechnol Prog*, 2013, 29: 394–402
- 82 Lange E, Tautenhahn R, Neumann S, Gröpl C. *BMC BioInf*, 2008, 9: 375
- 83 Pelzing M, Neuss C, Macht M. *LC GC Eur*, 2004, 17: 38–39
- 84 Awad H, Khamis MM, El-Aneed A. *Appl Spectrosc Rev*, 2015, 50: 158–175
- 85 Hewavitharana AK, Tan SK, Shaw PN. *LC GC NAM*, 2014, 32: 54–64
- 86 Rogatsky E, Tomuta V, Jayatillake H, Cruikshank G, Vele L, Stein DT. *J Sep Sci*, 2007, 30: 226–233
- 87 González-Antuña A, Domínguez-Romero JC, García-Reyes JF, Rodríguez-González P, Centineo G, García Alonso JI, Molina-Díaz A. *J Chromatogr A*, 2013, 1288: 40–47
- 88 Dizdaroglu M, Reddy PT, Jaruga P. *J Proteome Res*, 2011, 10: 3802–3813
- 89 Bruheim P, Kvitvang HFN, Villas-Boas SG. *J Chromatogr A*, 2013, 1296: 196–203
- 90 Zhang S, You J, Ning S, Song C, Suo YR. *J Chromatogr A*, 2013, 1280: 84–91
- 91 Kaiser P, Akerboom T, William GW, Reinauer H. *Clin Lab*, 2006, 52: 37–42
- 92 Jeong JS, Kim SK, Park SR. *Electrophoresis*, 2012, 33: 2112–2121
- 93 Wu Y, Li L. *J Chromatogr A*, 2016, 1430: 80–95
- 94 Mizuno H, Ueda K, Kobayashi Y, Tsuyama N, Todoroki K, Min JZ, Toyooka T. *Biomed Chromatogr*, 2017, 31: e3864
- 95 van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. *BMC Genomics*, 2006, 7: 142
- 96 Worley B, Powers R. *Curr Metabolomics*, 2013, 1: 92–107
- 97 Fonville JM, Richards SE, Barton RH, Boulange CL, Ebbels TMD, Nicholson JK, Holmes E, Dumas ME. *J Chemometr*, 2010, 24: 636–649
- 98 Abdi H. *WIREs Comp Stat*, 2010, 2: 97–106
- 99 Goryński K, Kiedrowicz A, Bojko B. *J Pharmaceut Biomed Anal*, 2016, 127: 147–155

- 100 Ling YS, Liang HJ, Lin MH, Tang CH, Wu KY, Kuo ML, Lin CY. *Biomed Chromatogr*, 2014, 28: 1284–1293
- 101 Speybrouck D, Lipka E. *J Chromatogr A*, 2016, 1467: 33–55
- 102 Kind T, Tsugawa H, Cajka T, Ma Y, Lai Z, Mehta SS, Wohlgemuth G, Barupal DK, Showalter MR, Arita M, Fiehn O. *Mass Spectrom Rev*, 2017, 56, doi: 10.1002/mas.21535
- 103 Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. *Chem Res Toxicol*, 2016, 29: 1956–1975
- 104 Klampfl CW, Himmelsbach M. *Anal Chim Acta*, 2015, 890: 44–59

Analysis of small molecule compounds in complex mixture system by LC-MS

Lingling Yu^{1,2}, Xing Li^{2,3}, Yanli Chen², Chao Wen^{2,3}, Kaifeng Hu^{2*}

¹ School of Chemical Science and Technology, Yunnan University, Kunming 650091, China

² State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author (email: kaifenghu@mail.kib.ac.cn)

Abstract: It has been a challenging task of analytical chemistry to qualitatively and quantitatively analyze the diverse compounds in the complex mixture. Liquid chromatography-mass spectroscopy (LC-MS), especially, ultra-high performance liquid chromatography (UPLC) coupled with high resolution mass spectrometry (HRMS), is a powerful tool for the analysis of complex samples. UPLC can greatly enhance the ability and efficiency of chromatographic separation. On the other hand, HRMS can provide more precise molecular weight and can thus improve the reliability and accuracy of qualitative and quantitative analysis. In this paper, we summarized the application of LC-MS technology to analysis of small molecule compounds in typical complex mixtures (urine, blood, cell metabolites and natural product extracts), concentrating on sample preparation, data acquisition, data (pre)processing, qualitative and quantitative analysis, multivariate analysis. Furthermore, technical challenges and developments in the analysis of complex samples by LC-MS were summarized and prospected.

Keywords: UPLC, HRMS, complex mixtures, qualitative analysis, quantitative analysis, multivariate analysis

doi: [10.1360/N032017-00118](https://doi.org/10.1360/N032017-00118)