

高分辨电喷雾离子化质谱数据批量自动化 预鉴别软件的研制

梁乾德, 马 靖, 高 月

(军事医学科学院放射与辐射医学研究所, 北京 100850)

摘要: 为了克服高分辨电喷雾离子化质谱数据预鉴别工作量大、繁琐的技术瓶颈, 采用 Microsoft Excel 插件方式, 研制了用于高分辨电喷雾离子化质谱数据批量自动化预鉴别的软件。此软件可根据用户指定的 Excel 文件中数据库数据和待鉴别数据的位置, 最小质荷比偏差等信息, 自动鉴别并生成鉴别结果报告。实际测试表明, 该软件与人工鉴别结果完全相同, 既节省了人力, 又能够极大地提高工作效率, 并且操作简便, 容错性好, 具有实用性。

关键词: 质谱; 预鉴别; 软件; 自动化; 高分辨

中图分类号: O 657.63 **文献标识码:** A **文章编号:** 1004-2997(2012)05-0286-04

Software Development for Batch Automatic Preliminary Identification of High Resolution Electrospray Mass Spectrometric Data

LIANG Qian-de, MA Jing, GAO Yue

(*Institute of Radiation Medicine, Academy of Military Medical Science, Beijing 100850, China*)

Abstract: In order to tackle the bottleneck of heavy-workload and cumbersomeness with preliminary identification of high resolution electrospray mass spectrometric data, a software as Microsoft Excel Plug-in for batch automatic preliminary identification of high resolution electrospray mass spectrometric data was developed, which performs automatic identification and creates report of results based on information such as location of library data and data to be identified in an Excel document and mass accuracy, etc. designated by user. Test showed that the software produces same result compared with handwork, but saves manpower and greatly improves efficiency, with simple operation, high fault tolerance and practicability.

Key words: mass spectrometry; identification; software; automatic; high resolution

代谢组学 (metabolomics) 致力于生物系统
中小分子物质 (相对分子质量 < 1 500) 的定性和

定量分析^[1], 目前已运用于微生物学、动物学、植
物学、诊断标志物发现、毒理学、药物研发、食品

收稿日期: 2012-04-05; 修回日期: 2012-07-17

基金项目: 国家自然科学基金 (81073161) 资助

通信作者: 梁乾德 (1971~), 男 (汉族), 吉林农安县人, 博士, 从事药物分析、中药现代化等研究。E-mail: liangqiande@yahoo.com.cn

通信作者: 高 月 (1963~), 女 (汉族), 江苏宜兴人, 研究员, 从事药理毒理学、中药现代化等研究。E-mail: gaoyue@bmi.ac.cnc

和环境科学等诸多领域^[2-3]。核磁共振和质谱是代谢组学研究的主要技术手段,后者在灵敏度等方面拥有优势^[4-5]。近年来,高分辨质谱技术的出现,使质谱定性分析能力大大提高,从而在代谢组学中的地位越来越重要^[6-9]。其中,电喷雾离子化方式以其“软电离”特性而倍受青睐^[1]。

在代谢组学研究中,样品组成不明确且目标化合物不确定的分析称为“非靶标分析(untargeted analysis)”。目前,在基于质谱的代谢组学领域,化合物的定性鉴别已经成为非靶标分析的技术瓶颈^[6],这很大程度上因为对复杂样品中大量化合物进行定性鉴别的工作量极大^[3]。含未知成分样品的高分辨电喷雾离子化质谱数据的鉴别通常需要借助化合物数据库^[3,10-14]。将实验获得的待鉴定精确质荷比与数据库中化合物相应离子的精确质荷比进行比对,确定一种或数种可能的化合物,即预鉴别,然后用对照品进行实验验证以确定化合物的结构^[1]。但是,优质的数据库通常化合物数量庞大,例如人类代谢组学数据库 HMDB(Human Metabolome Database)目前含有超过 7 900 个代谢物^[15]。而且同一种化合物通过电喷雾离子化有可能形成多种形态的离子^[3],如正离子模式下有可能形成 $[M+H]^+$ 、 $[M+NH_4]^+$ 、 $[M+Na]^+$ 、 $[M+K]^+$ 、 $[2M+H]^+$ 、 $[2M+H+K]^{2+}$ 、 $[2M+H+Na]^{2+}$ 等,在负离子模式下有可能形成 $[M-H]^-$ 、 $[M+HCOO]^-$ 、 $[2M-H]^-$ 、 $[M+Cl]^-$ 等。必须分别对多种可能的离子形态进行计算和比较,才能较为全面地排查,减少预鉴别的疏漏和误判,可见预鉴别的工作量很大。另一方面,当所分析的样品为成分未知的复杂混合物(如中药提取物、动植物体液、组织等)时,会产生大量待鉴别的数据,从而进一步增加预鉴别的工作量。

对于如此巨大的工作量,利用计算机实现自动化鉴别是必然的选择。目前国外已经研制了一些自动化预鉴别软件^[3],但国内还未见这方面的研究报道。为了向国际水平靠拢,以及能够根据自身需要开发具有特定功能的软件,有必要开展在本领域的研究探索,获得自主研发这类软件的能力。本工作研制了一种用于高分辨电喷雾离子化质谱数据批量自动化预鉴别的软件,定名为“Searcher”。旨在用计算机代替人完成繁琐的预鉴别工作,以期节省人力,提高工作效率。

1 试验部分

1.1 仪器与试剂

Microsoft WindowXP™ 操作系统, Microsoft Excel™ 软件(2003 或 2007 版本),普通个人电脑。Waters Acquity™ 超高效液相色谱系统(UPLC), Waters SYNAPT HD MS™ 四极杆飞行时间质谱系统(使用电喷雾电离源), Waters HSS T3(100 m×2.1 mm×1.8 μm)色谱柱。乙腈(HPLC 级):美国 Fisher Scientific 公司产品;甲酸(HPLC 级):德国 CNW Technologies GmbH 公司产品;水:由美国 Millipore 超纯水系统制备。

1.2 Searcher 软件的工作原理及操作流程

Searcher 软件采取 Excel 插件的方式工作。安装后在任意打开的 Excel 工作簿中生成一个启动按钮。首先将数据库数据和待鉴别数据拷贝到任意一个 Excel 工作簿中。数据库数据必须包含化合物各种离子形态下的精确质荷比、化合物名称等信息,排列方式为:每种化合物占一行,每种信息(各种离子的精确质荷比、化合物名称等)占一列。待鉴别数据必须排成一列。两种数据可以拷贝到同一个工作表,也可以拷贝到不同工作表。点击 Searcher 启动按钮后出现软件界面,示于图 1。点击界面上的“指定标准数据(库数据)的范围”按键,用鼠标选择数据库中用于预鉴别的精确质荷比数据区域;点击界面上的“指定标准数据(库数据)的标识列”按键,用鼠标选择数据库中化合物标识(如化合物名称)所在列;点击界面上的“指定受检数据(实际观测数据)的范围”按键,用鼠标选择待鉴别质荷比数据所在列。根据需要修改质荷比准确度(mass accuracy)的上限(单位是 ppm),然后点击“检索键”,启动鉴别。鉴别时,计算机对每一个待鉴别数据,都会在全部分指定的数据库数据范围内逐一比对,当二者差异小于等于指定的匹配准确度上限时,就会将该数据认定为匹配数据。鉴别完成后,自动生成一个鉴别结果报告表,示于图 2。报告表中列出全部待鉴别质荷比,其中在库中有匹配数据者所在单元格填充红色,并在右侧列出匹配化合物个数、全部匹配化合物的标识及其匹配准确度值(单位是 ppm)。



图 1 软件工作界面示例

Fig. 1 Example of the software's working interface

图 2 鉴别结果报告表示例

Fig. 2 Example of identification results report sheet

1.3 软件功能的验证

通过文献检索,建立中药地黄 *Rehmannia glutinosa* LIBOSCH 的化合物数据库,共含有 113 种化合物。对其中所有化合物,依其分子式,计算出 $[M-H]^-$ 、 $[M+HCOO]^-$ 、 $[2M-$

$H]^-$ 3 种常见负离子的精确质荷比。待鉴别数据是本实验室某次地黄成分分析获得的部分精确质荷比数据,共 62 个,分析系统为 UPLC-TOF MS,负离子模式,此处省略分析过程。所有数据在 Excel 表格中排成一列。分别以人工

方式和软件方式进行预鉴别,然后比较鉴别结果。

2 结果与讨论

2.1 软件功能验证结果

经比较,人工方式和软件方式预鉴别结果完全一致,共得到 14 个匹配数据。人工方式耗时数小时,而软件方式耗时仅 1 s 左右。在节省人力的同时,工作效率得到极大提高。

2.2 软件特点

此软件使用简单方便,容错度高。鉴于研究工作中通常以 Microsoft Excel 文件为质谱实验数据和数据库数据的载体,该软件采取 Excel 插件的方式工作。在指定数据库数据范围和待鉴别数据范围时,无需用键盘输入,只需用鼠标将 Excel 表上相应范围框住,然后点击确定即可。在指定数据库数据标识列时,只需用鼠标点击该列任意单元格,然后点击确定即可。匹配准确度限值可以根据需要任意改变。当发现不合法数据(非数值、空单元格等)时会自动跳过。

3 结论

Searcher 软件能代替人工完成高分辨电喷雾离子化质谱数据批量自动化预鉴别。既节省人力,又能够极大地提高工作效率,并且操作简便,容错性好,具有实用性。

参考文献:

- [1] XIAO J F, ZHOU B, RESSOM H W. Metabolite identification and quantitation in LC-MS/MS-based metabolomics[J]. Trends Anal Chem, 2012, 32: 1-14.
- [2] SUGIMOTO M, KAWAKAMI M, ROBERT M, et al. Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis [J]. Curr Bioinform, 2012, 7(1): 96-108.
- [3] DUNN W, ERBAN A, WEBER R Z, et al. Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics [J]. Metabol, 2012, published online: 26 May.
- [4] THEODORIDIS G A, GIKA H G, WANT E J, et al. Liquid chromatography-mass spectrometry based global metabolite profiling: A review[J]. Anal Chim Acta, 2012, 711: 7-16.
- [5] COY S L, CHEEMA A K, TYBURSKI J B, et al. Radiation metabolomics and its potential in biodosimetry[J]. Int J Radiat Biol, 2011, 87(8): 802-823.
- [6] KUEGER S, STEINHAUSER D, WILLMITZER L, et al. High-resolution plant metabolomics: From mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions[J]. Plant J, 2012, 70(1): 39-50.
- [7] HNATYSHYN S, SHIPKOVA P, et al. Automated and unbiased analysis of LC-MS metabolomic data[J]. Bioanal, 2012, 4(5): 541-554.
- [8] BURGESS K, CREEK D, DEWSBURY P, et al. Semi-targeted analysis of metabolites using capillary-flow ion chromatography coupled to high-resolution mass spectrometry [J]. Rapid Commun Mass Spectrom, 2011, 25(22): 3 447-3 452.
- [9] REGAL P, ANIZAN S, ANTIGNAC J P, et al. Metabolomic approach based on liquid chromatography coupled to high resolution mass spectrometry to screen for the illegal use of estradiol and progesterone in cattle[J]. Anal Chim Acta, 2011, 700(1/2): 16-25.
- [10] TSUTSUI H, MAEDA T, MIN J Z, et al. Biomarker discovery in biological specimens (plasma, hair, liver and kidney) of diabetic mice based upon metabolite profiling using ultra-performance liquid chromatography with electrospray ionization time-of-flight mass spectrometry [J]. Clin Chim Acta, 2011, 412(11/12): 861-872.
- [11] MOHAMED R, VARESEIO E, IVOSEV G, et al. Comprehensive analytical strategy for biomarker identification based on liquid chromatography coupled to mass spectrometry and new candidate confirmation tools[J]. Anal Chem, 2009, 81(18): 7 677-7 694.
- [12] YOO B C, KONG S Y, JANG S G, et al. Identification of hypoxanthine as a urine marker for non-Hodgkin lymphoma by low-mass-ion profiling[J]. BMC Cancer, 2010, 10: 55.
- [13] LOFTUS N, BARNES A, ASHTON S, et al. Metabonomic investigation of liver profiles of nonpolar metabolites obtained from alcohol-dosed rats and mice using high mass accuracy MSn analysis[J]. J Proteome Res, 2011, 10(2): 705-713.
- [14] XU F, ZOU L, LIN Q, et al. Use of liquid chromatography-tandem mass spectrometry and online databases for identification of phosphocholines and lysophosphatidylcholines in human red blood cells[J]. Rapid Commun Mass Spectrom, 2009, 23(19): 3 243-3 254.
- [15] WISHART D S, KNOX C, GUO A C, et al. HMDB: A knowledgebase for the human metabolome[J]. Nucleic Acids Res, 2009, 37 (Database issue): D603-610.