

生物质谱在基因组学研究中的应用*

吴 晶 纪建国

(北京大学生命科学学院 北京 100871)

摘 要 综述生物质谱用于分析核糖核酸的应用进展。概述电喷雾电离质谱(ESI-MS)和基质辅助激光解吸电离质谱(MALDI-MS)的原理及它们在核酸分析中的应用。总结质谱用于单核苷酸多态性分型分析(SNP genotyping);对短的串联重复序列(STR)的分析;对寡核苷酸片断的序列分析等三个方面的研究成果。提出质谱用于基因组学研究存在的问题,并展望生物质谱未来的发展方向。

关键词 基因组学 电喷雾电离质谱(ESI-MS) 基质辅助激光解吸质谱(MALDI-MS) 生物质谱的应用

生物质谱的发展基于20世纪80年代产生软电离方式:电喷雾电离质谱(Electrospray Ionization - Mass Spectrometry, ESI-MS)和基质辅助激光解吸电离质谱(Matrix - ssisted Laser Desorption Ionization - Mass Spectrometry, MALDI-MS)¹等,使得生物质谱能用于分析高极性、难挥发和热不稳定性生物大分子。生物质谱具有快速、准确、灵敏、高通量等特征,这些特征符合对蛋白质和核酸的分析要求。因此,生物质谱在基因组学和蛋白组学中得到广泛的应用。

1 生物质谱

1.1 电喷雾质谱

电喷雾技术由Zeleny于1917年发明,1984年美国耶鲁大学John Fenn研究首次发表ESI-MS的研究成果,并于4年后报道运用ESI-MS首次成功地进行蛋白质分析,从此揭开ESI-MS用于生物大分子的研究。

ESI具有从液相中直接取样的优点,对分析物施加相当小的内能(所以完整的大分子能被检测)并有效的将样品带上极性和电荷带入气相。含有多于5个或6个残基的多肽几乎总被观测到呈多质子状态;其电荷状态与其大小和碱性氨基酸存在的数目成正比。分子量远远超过质谱仪所具有的质量范围的蛋白质通常可被检测到,因携带足够多的电荷使其产生 m/z 小于2000的“外壳”样峰。且可以方便地与高效液相色谱(HPLC)和毛细管电泳(CE)等分离技术在线联用,在分析复杂混合物时很有优势²。

1.1.1 电喷雾离子化 电喷雾电离是在液滴变成蒸汽产生离子发射过程中形成的,也称为“离子蒸发”。待测分子溶解在溶剂中,以液相方式通过毛细管到达

喷嘴,在喷嘴高电压作用下形成带电荷的微滴,随着液滴中挥发性溶剂蒸发,微滴表面的电荷体密度随半径减少而增加,到达某一临界点时,样品将以离子方式从液滴表面蒸发,进入气相,即实现样品的离子化,由于没有直接的外界能量作用于分子,因此对分子结构破坏较少,是一种典型的“软电离”方式。

在常规电喷雾中,喷雾时形成的较大液滴中,有一部分样品还来不及离子化,因而降低样品的利用率和灵敏度,近年来发展纳升级电喷雾方法(nanospray),它采用内径 $1\mu\text{m}$ 左右的喷嘴,在喷嘴前端,喷雾液滴的体积是常规电喷雾的 $1/100$,使样品充分利用,并有效离子化。这样就大大提高样品的利用率和灵敏度³。

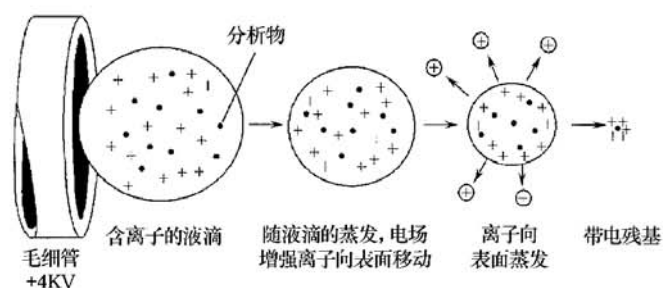


图1 电喷雾电离工作原理

1.2 基质辅助激光解吸质谱

激光解吸电离质谱(Laser desorption/ionization - mass spectrometry, LDI-MS)是分析难挥发有机物的手段之一,曾用于分析合成聚合物和热不稳定的小分子。但激光解析测定的分子量低于3kDa。直至1988年Tanaka和Hillenkamp两个研究小组分别提出使用基质辅助激光解析(matrix-assisted laser desorption ionization, MALDI)电离质谱,使LDI-MS应用于生

* 本项目由国家重点基础研究发展规划项目计划(2001CB510207)和国家重点科技攻关计划(2004BA711A18)资助

物大分子分析得到发展³。

1.2.1 基质辅助激光解吸离子化 MALDI-MS 分析生物大分子的关键是首先分析的生物大分子与合适的基质(一种小分子有机酸)形成共结晶,当激光照射到晶体上时,基质分子吸收能量,并转换成固体的激发能,从而导致固相转移,形成气相分子,此时被分析与基质分子一起被释放出来。在此过程中生物大分子与基质分子同时发生质子转移,生物大分子得到离子化。

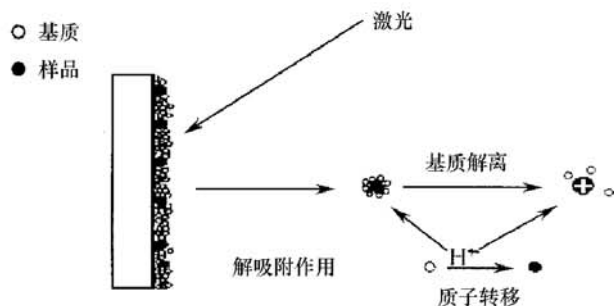


图2 MALDI工作原理图

用 MALDI 分析生物大分子的过程中,选择合适的基质很重要,一般认为基质具备三种功能:(1)能吸收能量;(2)能使生物大分子彼此分离;(3)能使被分析的生物大分子离子化。

2 生物质谱在基因组学中的应用

2.1 生物质谱分析前样品的纯化

因为核酸分子带有磷酸基,极性和电负性大,容易吸收大量的碱金属离子,形成碱性离子加合物。导致离子速度空间和能量的分散,引起核酸分子离子峰变宽,不易准确测定其质量数⁴。因此,样品的纯化对得到好的质谱分析结果至关重要。

2.1.1 应用 ESI-MS 分析前样品的纯化 电喷雾电离产生多电荷离子,非挥发性的阳离子(如 Na^+ , K^+)是影响 ESI-MS 对寡核苷酸和核酸分析的主要原因。

一般所采用的样品纯化和脱盐的方法主要有⁵:(1)通过离子交换的方式使阳离子(如 Na^+ , K^+)与 H^+ 或 NH_4^+ 相交换。(2)从浓的醋酸铵溶液中用冰冻乙醇将寡核苷酸沉淀也是简单有效的脱盐方法。但这种方法比较耗时而且寡核苷酸的分子量不能太大(小于 25-mer)。(3)C18 反相 HPLC 可应用于酵母 tRNA 的脱盐。(4)用 N-三乙氨基乙酸盐(Buffer A)和乙腈(Buffer B)对样品进行梯度洗脱能减少盐的影响,而得到较高信噪比的质谱图。

用电喷雾质谱测定分子量的精确度主要依赖于质量分析器的选择。一般小于 50-mer 的寡核苷酸

选用三极杆和离子阱质量分析器,产生误差为正负 0.005% 或更好,ESI-TOF(time of flight)质谱仪测定分子量的误差可低于 ppm(百万分之一),傅立叶变换离子回旋共振质谱误差率低于 ppm。

尽管质量分析器的选择决定质谱分析的精确度,但是质谱分析结果的好坏最终决定于样品的纯度,充分脱盐的样品相对于含较高盐离子浓度的样品,可得到较精确的质谱信息。所以我们可以想象,较高分子质量的寡核苷酸,由于完全脱盐较难,所以他们的质量分析误差较大。

2.1.2 应用 MALDI-MS 分析前样品的纯化 MALDI-TOF-MS 虽在分析大分子物质如蛋白质、多肽、小 DNA 链中取得较好的结果,但用来分析 30 个碱基以上的 DNA 片段则不能取得理想的效果,其主要原因是灵敏度和分辨率达不到要求。所以进行较长链 DNA 的 MALDI-TOF-MS 分析之前,往往需对样品进行纯化和预处理,才能得到满意的分析结果。样品的纯化是提高分辨率和灵敏度的关键因素。

所采用的样品纯化方法主要包括为两大类⁵:一是去除碱离子法,二是分析前 DNA 样品的纯化法。去除碱离子法包括:(1)在基体中加入一定量的柠檬酸铵或柠檬酸氢二铵。 NH_4^+ 可排斥 Na^+ 、 K^+ 等离子优先与 DNA 中的 (O^-) 作用生成 ONH_4 。(2)使用 NH_4^+ 阳离子交换树脂珠,把溶液中的碱金属离子置换出来,此法适于少量样品溶液。(3)使用高效液相色谱(HPLC)置换溶液中的碱金属离子,此法适于溶液量大的情况。(4)使用电泳方法分离溶液中的碱金属离子。

分析前 DNA 样品的纯化常采用生物素-抗生物素蛋白链菌素体系。它是一种常用于生物活性物质检测和分析前进行纯化的手段。抗生物素蛋白链菌素能与连有 DNA 的生物素以极强的非共价键结合形成加合物。一种较有效的纯化方法是:采用将末端 5' 位结合有光切生物素的 DNA 样品捕捉于涂有抗生物素蛋白链菌素的磁珠上的方法,可直接在激光的照射下得纯化的 DNA 样品而进行分析。

用 MALDI-MS 分析核苷酸的分子量,能否得到高的分辨率和精确度仍然取决于样品的纯化步骤和质量分析器的选择。飞行时间质量分析器引入延迟提取(delayed extraction)技术和离子反射器(reflector)大大提高质谱分析的质量精确度和分辨率⁶。

2.2 生物质谱在单核苷酸多态性分型中的应用

单核苷酸多态性(Single nucleotide polymorphism, SNP)是指在基因组 DNA 某个位置处存在有单个碱基

的差异,其在人群中发生几率超过1%,人类基因组DNA序列90%多态性是由SNP造成的。随着大量的单核苷酸多态位点的发现,人们想知道不同人群的兴趣。SNP位点处等位基因频率(包括疾病人群和正常人群、不同种族人群等等),这往往需要对成千上万个样本进行分析,工作量非常巨大,因此急需一些高通量、低成本的SNP分型技术,以满足需求。

生物质谱技术正好符合上述需求。特别是准确性方面,由于它检测的是寡核苷酸本身的一个物理性质—分子量,而不像其它检测方法是通过间接的方式(如荧光的强度、荧光的极化、寡核苷酸的迁移率等)来检测。此外,由于它检测寡核苷酸片段分子量,所以不需要对这些片段进行特殊修饰;另外,还可以同时对多个SNP位点进行分析,这些都大大降低分析成本。

在单核苷酸多态性研究中多采用MALDI-TOF-MS。这是由于它具有快速、耗费低、谱图简单直观等特点。一般情况下,收集一个图谱用时不超过1min,有时甚至用时只有几秒。但是,有一个明显的缺点,就是寡核苷酸与基质形成的结晶不均匀。所以,为获得一个好的图谱,必须寻找所谓的“好”点(Sweet hot),这使得其自动化程度受到限制。此外,应用MALDI-TOF-MS的另一个缺陷是,随着寡核苷酸链长度的增加,质谱的准确度和分辨率都下降。所以,为准确区分碱基间的差异,特别是腺嘌呤(A)和胸腺嘧啶(T)(四个碱基中这两者之间的差值最小,为9Da),要求的寡核苷酸长度一般不超过45个碱基,这影响它分析多重SNP位点的能力。目前文献报道的应用多重分析SNP最多的达到12重⁷。

2.2.1 SNP分型检测的方法 已报道应用生物质谱作为SNP分型检测方法有很多,根据所选用确定SNP位点处等位基因基因型方法不同可分为等位基因特异杂交反应、等位基因特异连接反应、等位基因特异切割反应和等位基因特异延伸反应。截止到现在,大规模的应用主要是等位基因特异延伸反应⁸。

等位基因特异延伸反应的原理类似于微型测序方法。主要过程是利用PCR技术扩增出含有多态性碱基的产物,然后用核酸外切酶和磷酸酶处理PCR产物,以除去未参加反应的PCR引物和底物dNTP。随后加入用于SNP分型的引物,在四种ddNTP的存在下进行反应。分型引物直接在多态性碱基上游或下游发生退火;在ddNTP存在下,进行一个碱基的延伸;寡核苷酸引物的3'端的碱基紧挨于多态性碱基,引物开始延伸的第一个碱基是多态

性碱基,然后根据延伸产物的大小确定等位基因。杂合子个体在质谱图上有二个延伸产物,纯合子个体只有一个产物(见图3)。

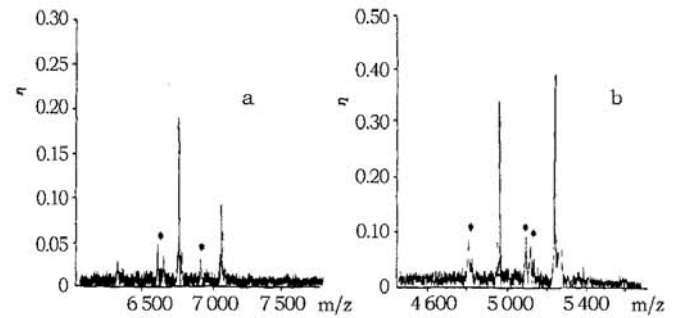


图3 SNP分型的MALDI-TOF MS谱图

* 峰是脱碱基造成;a 纯合子;b 杂合子

2.2.2 在同一体系进行多个SNP位点的分析 相对与单个SNP位点分析,在同一体系进行多个SNP位点的分析,有利于节省时间和降低质谱鉴定的费用。同时进行多位点分析,就要设计多种不同的延伸引物,要保证这些引物质量和长度上的不同及延伸后的序列不要有交叠。为更有效的使分离目的片段进行质谱分析,可以将用于单碱基扩增的ddNTP与生物素(biotin)相连。引物的生物素主要使延伸的产物能被固相结合,在进行质谱分析前得到纯化。应用此种方法,在同一体系进行多个SNP位点的分析,国内杨何义等人完成10重SNP_s的分型检测⁸(见图4)。

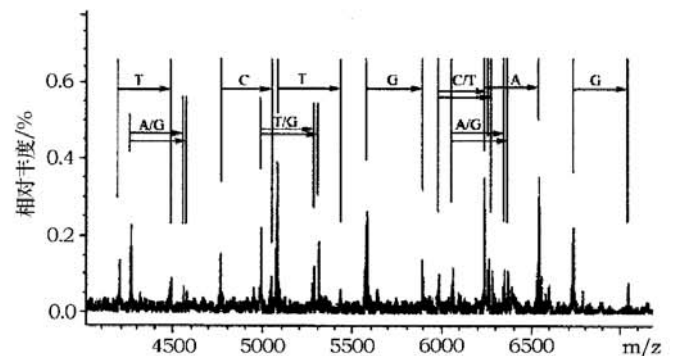


图4 10重SNP分型MALDI质谱图

箭头表示相应的延伸引物及其产物;没标出的峰都是由寡核苷酸脱碱基或寡核苷酸与碱金属形成的加合物造成。

2.3 生物质谱对短的串连重复序列的分析

短串联重复序列(short tandem repeats, STR)是由几个碱基对作为核心单位,串联重复形成的一类DNA序列,由于核心重复单位数目的变化构成STR基因座的遗传多态性。人类基因中,平均每15kb就有一个STR基因座,提供丰富的遗传标记来源。STR_s的研究在法医学,基因作图和疾病治疗上越来越受欢迎。

STR精确分析所要求的质量分辨率和精确度,取决于被分析的STR类型。只包含一种重复序列和非一致性的等位基因STR,可以方便地通过MAL-

DI-MS 或 ESI-MS 进行分析。因为两个非一致性等位基因的 STR 之间至少差一个核苷酸的分子量 (~300Da), 分辨率不是很高的质量分析器都可以将他们分开。若含有复杂的 STR_s 序列, MALDI-TOF-MS 就无法用来分析。它这就要求有更高分辨率和分析质量精确度的质谱。

Muddiman⁵ 等人首先运用 ESI 和傅立叶转换离子回旋共振质谱(FTICR)对混和的或复杂的 STR_s 进行分析。他们的实验结果证明, FTICR 为 STR 的基因分型提供高分辨率和高精确度的分析手段。

2.4 生物质谱对寡聚核苷酸片断的序列分析

尽管用于 DNA 片段的序列分析的凝胶技术已经非常完美, 但是还存在下列缺点: 第一手续繁杂, 较为费时; 第二容易受单链 DNA 结构的影响, 尤其对于富含 GC(G: 鸟嘌呤核苷酸, C: 胞嘧啶核苷酸)的片段, 往往不能得到明确的结果。ESI 和 MALDI 质谱技术的出现为寡核苷酸及其类似物的结构和序列分析提供强有力的方法。

质谱寡核苷酸的序列分析通常有 3 种方法: (1) 用质谱代替凝胶电泳, 分析双脱氧法合成的混合寡核苷酸段。国内邓慧敏等用 DE-MALDI-MS(延迟提取-基质辅助激光解析-质谱)法测定混合碱基 DNA, 获得高分辨率的 DNA 质谱图。(2) 分别用外切酶 3' 或 5' 端进行部分降解切割寡核苷酸片段, 在不同时间内分别取样进行质谱分析, 获得寡核苷酸部分降解的分子离子峰信号, 通过对相邻两个碎片分子质量进行比较, 可以计算出被切割的核苷酸单体分子质量, 将其与四个脱氧核苷酸的标准分子量进行对照, 就可以读出寡核苷酸的序列。(3) 运用串联质谱直接分析寡核苷酸的序列。第一级质量分析器获得所有组分的分子离子峰(母离子), 经质量过滤器挑出需要进一步分析母离子, 此母离子在碰撞室内经高流速性气体碰撞诱导离解产生碎片分子(子离子), 子离子进入第二级质量分析器获得母离

子的碎片峰。通过研究母离子和子离子的裂解关系, 可以获得寡核苷酸的序列信息。

3 结论

与其他分析生物大分子方法相比, 生物质谱有准确性和灵敏度高、快速、易于大规模和高通量操作等优点, 在基因组学和蛋白质组学研究中得到很大的应用和发展。但生物质谱在分析核糖核酸中, 由于碱基易断裂, 基体或痕量碱金属离子导致加合物的生成以及随着寡核苷酸链长度的增加, 质谱的准确度和分辨率都下降等缺点, 使生物质谱在分析核糖核酸方面的应用要远落后于用于蛋白质和多肽的分析。

尽管如此, 由于生物质谱本身所拥有的优点及技术的不断完善, 生物质谱已经在核酸分析基础研究领域得到广泛的应用, 在法医遗传学领域也有很好的应用前景。我们相信随着质谱技术的不断完善, 它用于核酸分析所存在的问题今后能够得到解决, 从而能更好地发挥其准确、快速、灵敏度高的特点, 为生物学家解开生命的奥秘和改善人们的生活提供更加强有力的武器。

参考文献

- 1 黄凌云. 生物质谱在蛋白质组学研究中的应用, 现代仪器, 2004(1):7~12
- 2 史国放. 液相色谱-质谱联用技术, 现代仪器, 2001(1):9~13
- 3 夏其昌主编. 蛋白质化学与蛋白质组学, 北京: 北京科学出版社, 2004:312~314
- 4 张健. 生物质谱技术在法医遗传学研究中的应用展望. 刑事技术, 2004(3):22~26
- 5 Zhaojing Meng, et al. The use of mass spectrometry in genomics, *Biomolecular Engineering*2004(21):1~13
- 6 赵善楷. 基体辅助激光解吸电离飞行时间质谱仪的离子延时引出新技术研究, 现代仪器, 2001(2):1~3
- 7 Brayms, Boerwink, Dorispa. High throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise (J. Hum Mutat, 2001, 17(4):296~304
- 8 Heyi Yang, Multiplex single-nucleotide polymorphism genotyping by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Analytical Biochemistry* 2003, 314:54~62

The application of mass spectrometry in genomic analysis

Wu Jing Ji Jianguo

(Department of Biochemistry and Molecular Biology, Peking university Beijing 100871)

Abstract The application progress on biological mass spectrometric analysis of nucleotide has been reviewed. The principle of electrospray ionization-mass spectrometry(ESI-MS) and matrix assisted laser desorption ionization-mass spectrometry(MALDI-MS) is described, as well as their applications in the genomic analysis. The study result on mass spectrometric analysis of single nucleotide polymorphisms genotyping, short tandem repeat, and sequencing of oligonucleotide. Simultaneously, the issue is put forward and the development direction is prospected.

Key words Genomic analysis ESI-MS MALDI-MS