



(12)发明专利申请

(10)申请公布号 CN 110692118 A

(43)申请公布日 2020.01.14

(21)申请号 201880035403.6

(74)专利代理机构 上海专利商标事务所有限公司 31100

(22)申请日 2018.05.07

代理人 陈洁 周全

(30)优先权数据

62/513,918 2017.06.01 US

(51)Int.Cl.

H01J 49/00(2006.01)

(85)PCT国际申请进入国家阶段日

2019.11.28

(86)PCT国际申请的申请数据

PCT/US2018/031382 2018.05.07

(87)PCT国际申请的公布数据

W02018/222345 EN 2018.12.06

(71)申请人 萨默费尼根有限公司

地址 美国加利福尼亚州

(72)发明人 P·F·叶 H·L·卡达西斯

小詹姆斯·L·斯蒂芬森

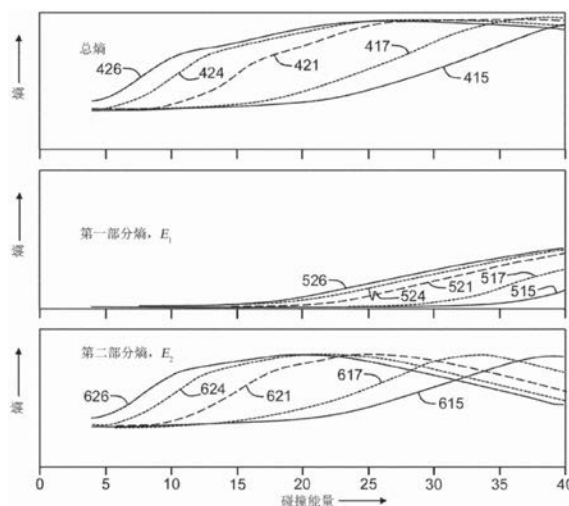
权利要求书1页 说明书34页 附图22页

(54)发明名称

质谱仪碰撞能量的自动测定

(57)摘要

本公开建立了新的解离参数,所述参数能用于测定使用碰撞池类型碰撞诱导解离来实现给定分析物前体离子的期望解离程度所需的碰撞能量(CE)。这种选择仅基于所述分析物前体离子的分子量MW和电荷态z。提出了能用作“解离程度”的参数的度量,并且针对实现每个度量的一系列值所需的CE建立了预测模型。每个模型都仅是前体离子的MW和z的简单平滑函数。通过结合实时质谱去卷积(m/z到质量)算法,根据本发明的方法能够通过以前体依赖性方式对碰撞能量进行自动实时选择来控制所述解离程度。



1. 一种用于使用质谱仪鉴定含有多个完整蛋白质的样品中的完整蛋白质的方法,所述方法包括:

(a) 将所述样品引入所述质谱仪的电离源;

(b) 使用所述电离源从所述多个完整蛋白质中产生多个离子种类,每个蛋白质由此产生所述多个离子种类的相应子集,其中每个子集中的每个离子种类是从所述完整蛋白质中相应的一个中产生的多质子化离子种类;

(c) 使用所述质谱仪的质量分析器对所述多个离子种类进行质量分析;

(d) 通过对由所述质量分析产生的数据进行数学分析,自动识别所述多个离子种类的每个子集,并将电荷态 z 分配给每个经过识别的离子种类,以及将分子量 MW 分配给每个完整蛋白质;

(e) 选择所述离子种类中的一个离子种类;

(f) 使用以下关系式自动计算用于使选定离子种类碎裂的碰撞能量 CE :

$$CE(D_P) = c + (1/k) [\ln(1/D_P) - 1],$$

其中 D_P 是所述选定离子种类中预期在所述碎裂之后保持未碎裂的一部分,而 c 和 k 仅是所述选定离子种类的电荷态 z 以及产生所述选定离子种类的所述完整蛋白质的分子量 MW 的函数;

(g) 使用自动计算的碰撞能量分离所述选定离子种类并使所述种类碎裂,从中形成碎片离子种类;以及

(h) 对所述碎片离子物质进行质量分析。

2. 一种用于使用质谱仪鉴定含有多个完整蛋白质的样品内的完整蛋白质的方法,所述方法包括:

(a) 将所述样品引入所述质谱仪的电离源;

(b) 使用所述电离源从所述多个完整蛋白质中产生多个离子种类,每个蛋白质由此产生所述多个离子种类的相应子集,其中每个子集中的每个离子种类是从所述完整蛋白质中相应的一个中产生的多质子化离子种类;

(c) 使用所述质谱仪的质量分析器对所述多个离子种类进行质量分析;

(d) 通过对由所述质量分析产生的数据进行数学分析,自动识别所述多个离子种类的每个子集,并将电荷态 z 分配给每个经过识别的离子种类,以及将分子量 MW 分配给每个完整蛋白质;

(e) 选择所述离子种类中的一个离子种类;

(f) 使用以下关系式自动计算用于使选定离子种类碎裂的碰撞能量 CE :

$$CE(D_E) = b_1 \times MW^{b_2} \times z^{b_3},$$

其中 D_E 是对应于通过所述碎裂产生的碎片离子种类的期望分布的参数, z 是所述选定离子种类的指定电荷态, MW 是产生所述选定离子种类的所述完整蛋白质的分子量,并且 b_1 、 b_2 和 b_3 是根据 DE 而变化的预定参数;

(g) 使用自动计算的碰撞能量分离所述选定离子种类并使所述种类碎裂,从中形成碎片离子种类;以及

(h) 对所述碎片离子物质进行质量分析。

质谱仪碰撞能量的自动测定

技术领域

[0001] 本发明涉及质谱法,更具体地说,涉及用于通过串联质谱法对蛋白质或多肽的复杂混合物进行质谱分析的方法和设备。更具体地说,本发明涉及利用碰撞诱导解离使前体离子碎裂的方法和设备,并且在所述方法和设备中,自动测定对要碎裂的前体离子的选择以及要赋予所选择的前体离子的碰撞能量的大小。

背景技术

[0002] 对活细胞和组织中的蛋白质的研究(蛋白质组学)是临床和基础科学研究的活跃领域,因为细胞和组织中的代谢控制是在蛋白质水平上进行的。例如,对健康组织与患病组织之间或病原微生物菌株与非病原微生物之间的蛋白质表达水平进行比较可以加快新药物化合物或农产品的发现和开发。此外,对患病组织或从接受治疗的生物体切除的组织中的蛋白质表达模式的分析还可以充当对疾病状态或治疗策略有效性的诊断,并提供有关个别患者的合适治疗方式和治疗选择的预后信息。更进一步地,对源自微生物(例如,细菌)的样品中的蛋白质组进行鉴定可以提供一种鉴定微生物的种类和/或菌株以及关于细菌鉴定此种类或菌株的可能的抗药性的手段。

[0003] 由于质谱法(MS)可用于提供详细的蛋白质和肽结构信息,因此其目前被认为是用于生化混合物分析和蛋白质鉴定的有价值的分析工具。因此,常规的蛋白质分析方法通常将用于分离和定量的二维(2D)凝胶电泳与蛋白质的质谱鉴定结合在一起。并且,毛细管液相色谱法以及各种其它“前端”分离或化学分馏技术已与电喷雾电离串联质谱法联用,以便在无凝胶电泳的情况下进行大规模蛋白质鉴定。通过使用质谱法可以鉴定质谱之间的质的差异,并且与仅在某些质谱中出现的峰相对应的蛋白质充当候选生物标志物。

[0004] 术语“自上而下的蛋白质组学”是指在无需事先进行酶促、化学或其它消化手段的情况下将蛋白质样品完整地引入质谱仪中的分析方法。自上而下的分析可以研究完整蛋白质,从而能够直接在蛋白质水平上进行鉴定、测定主要结构和对翻译后修饰(PTM)进行定位。自上而下的蛋白质组学分析通常由以下组成:将完整蛋白质引入质谱仪的电离源中;测定蛋白质的完整质量;将蛋白质离子碎裂;并测量这样产生的各个碎片的质荷比(m/z)和丰度。这种顺序的仪器步骤通常被称为串联质谱法,或者可替代地被称为“MS/MS”分析。这种技术可以有利地用于多肽研究。所得的碎片比简单肽的碎片复杂许多倍。对这种碎片质谱的解释通常包含将观察到的碎裂图型与蛋白质序列数据库进行比较,所述蛋白质序列数据库包含从已知样品生成的编译后的实验性碎裂结果,或者可替代地与理论上预测的碎裂图型进行比较。例如,Liu等人(“在四极杆/飞行时间串联质谱仪中通过离子阱碰撞诱导解离和离子/离子反应对先验未知蛋白质进行自上而下的蛋白质鉴定/表征(Top-Down Protein Identification/Characterization of a Priori Unknown Proteins via Ion Trap Collision-Induced Dissociation and Ion/Ion Reactions in a Quadrupole/Time-of-Flight Tandem Mass Spectrometer)”,《分析化学(Anal.Chem.)》2009,81,1433-1441)描述了对经过修饰和未经修饰的未知蛋白质进行的自上而下的蛋白质鉴定和表征,质量最高

可达 $\approx 28\text{kDa}$ 。

[0005] 自上而下的分析相对于自下而上的分析的一个优点在于,可以直接鉴定蛋白质,而不是像所谓的“自下而上”分析中的肽那样进行推测。另一个优点是可以鉴定蛋白质的替代形式,例如翻译后修饰和剪接变体。然而,自上而下的分析与自下而上的分析相比有一个缺点,因为许多蛋白质可能难以分离和纯化。因此,在质谱分析中,不完全分离的混合物中的每个蛋白质都可以产生多个离子种类,每个种类分别对应于不同的相应质子化程度和不同的相应电荷态,并且每个这样的离子种类都可以产生多个同位素变体。在自上而下的分析中测量的单个MS质谱可以容易地包含数百个甚至数千个属于不同分析物的峰——这些峰在给定的 m/z 范围内交织在一起,其中强度非常不同的离子信号重叠。

[0006] 当在MS分析之前进行如二维凝胶电泳或液相色谱等前端样品分级时,可以降低各个单独质谱的复杂性。尽管如此,这种样品部分的质谱仍可以包括多个蛋白质和/或多肽的特征。对通过液相色谱法(LC)分离的化合物所产生的离子进行质谱(MS)分析的一般技术可以被称为“LC-MS”。如果质谱分析作为串联质谱法(MS/MS)进行,则上述过程可以被称为“LC-MS/MS”。在常规的LC-MS/MS实验中,首先通过质谱法分析样品,以测定源自样品的离子的质荷比(m/z),并鉴定(即,选择)所关注的质谱峰。然后,通过对选定的一个或多个峰进行产物离子MS/MS扫描,进一步分析样品。更具体地说,在分析的第一阶段(通常被称为“MS1”)获得包括初始调查扫描的全扫描质谱。然后,在所述全扫描质谱之后选择一种或多种前体离子种类。使选定种类的前体离子发生碎裂,例如可以采用碰撞池或采用另一种形式的碎裂池(如表面诱导解离、电子转移解离或光解离)来完成。在第二阶段,使用同一质量分析器或第二质量分析器检测所得的碎片(产物)离子,以进行进一步分析(通常被称为“MS/MS”或“MS2”)。所得的产物谱图显示出一组碎裂峰(一个碎片组),在许多情况下,这些峰可以用作导出与前体离子种类有关的结构信息的手段。

[0007] 图1A示出了一种假设的实验情况,在这种情况下,归因于不同的分析物种类,不同的部分在被引入质谱仪后在色谱上得到很好的解析(及时)。曲线A10和A12代表每个相应分析物在各个时间的假设浓度,其中浓度表示为相对强度(R.I)标度的百分比,并且沿横坐标绘制时间作为保留时间。根据输入到质谱仪中的总离子电流的测量值可以容易地确定曲线A10和A12。将总离子电流的阈值强度水平A8设置为低于仅获取MS1数据的水平。当第一种分析物(检测为峰A10)洗脱时,总离子电流强度在时间 t_1 与阈值A8相交。发生这种情况时,质谱仪的机载处理器或其它控制器可能会启动要获取的一个或多个MS/MS质谱。随后,检测另一个洗脱峰A12的前沿。当总离子电流在时间 t_3 再次突破阈值强度A8时,将启动一个或多个另外的MS/MS扫描。通常,峰A10和A12将对应于不同分析物的洗脱,因此,在第一种分析物洗脱期间(时间 t_1 与时间 t_2 之间)选择不同的前体离子进行碎裂,而不是在第二种分析物的洗脱期间(时间 t_3 与时间 t_4 之间)进行选择。由于不同的前体离子通常将包括不同的 m/z 比和不同的电荷态,因此产生最佳碎裂所需的实验条件可能在两个不同的洗脱周期之间有所不同。

[0008] 在更复杂的分析物混合物中,可能存在洗脱峰完全重叠的组分,如图1B中的离子电流强度与保留时间关系图所示。在这个实例中,洗脱峰A11代表可归因于第一分析物产生的前体离子的离子电流,而洗脱峰A13代表可归因于第二分析物产生的不同前体离子的离子电流,其中这些不同的前体离子的质量和/或电荷态彼此不同。在图1B所示的假设情况

下,产生不同离子的化合物的洗脱几乎完全重叠,在共洗脱过程中,第一前体离子的质谱强度始终大于第二前体离子的质谱强度。如图1C中假设所示,在两种分析物共洗脱期间的任何时间——例如,时间 t_6 与时间 t_7 之间——所有前体离子的质谱都可能出现,其中,包络线78指示的线组由第一种分析物的电离引起,而包络线76指示的线组由第二种分析物的电离引起。在这些条件下,自动质谱分析不仅必须能够区分与不同的相应分析物相关的不同前体离子,而且还必须能够调整在质谱分析过程中赋予不同前体离子的碰撞能量,以使每个离子被最佳地碎裂。确实,如下所述,即使当分析物未共洗脱时,适当缩放所施加的碰撞能量也很重要。当多种分析物的特性(例如,MW和/或 z)显著不同时,无论相对洗脱时间如何,正确的缩放都特别重要。

[0009] 在MS/MS分析中引起离子碎裂的一种常见方法是碰撞诱导解离(CID),在所述方法中,使一群分析物前体离子加速进入目标中性气体分子,如氮气(N_2)或氩气(Ar),从而将内部振动能量赋予前体离子,这会导致键断裂和解离。分析碎片离子,以提供有关前体离子结构的有用信息。术语“碰撞诱导解离”包含通过共振激发过程将能量赋予前体离子的技术,其可以被称为RE-CID技术。除了主RF捕获电压之外,这种共振激发方法还包含将辅助交流电压(AC)施加到捕获电极。所述辅助电压通常具有相对较低的幅度(约为1伏特(V)),并且持续时间约为数十毫秒。选择所述辅助电压的频率以使其与离子的运动频率相匹配,离子的运动频率进而由主捕获场幅度、频率和离子的质荷比(m/z)决定。由于离子的运动与施加的电压共振,离子的能量增加,并且其运动幅度增大。

[0010] 图2示意性地示出了碰撞诱导解离的另一种方法,其有时被称为高能碰撞解离(HCD)。在HCD方法中,将选定的离子暂时储存在多极离子存储装置52中或使其通过所述多极离子存储装置,所述多极离子存储装置可以例如包括多极离子阱。在特定时间,改变栅电极组件54上的电势,以使选定的前体离子6加速离开离子存储装置并进入含有惰性目标气体分子8的碰撞室56中。将离子加速以使离子以动能与目标分子发生碰撞,所述动能由碰撞池与存储装置之间的电势差之差决定。

[0011] 当在MS/MS实验中使用HCD或RE-CID生成碎片离子时,令人高度期望的是,设置仪器以便将正确量的碰撞能量赋予选定的前体离子。对于HCD,通过设置电势差来设置碰撞能量(CE),离子通过所述电势差被加速进入HCD池。离子在那里与驻留气体发生一次或多次碰撞直到离子超出振动能阈值,以使键断裂,从而产生解离产物离子。产物离子可以保留足够的动能,使得进一步的碰撞导致连续解离事件。最佳碰撞能量根据选定的前体离子的性质而变化。将HCD碰撞能量设置得太高会导致此类连续解离事件,从而产生大量小型非特异性产物离子。相反,将这一电势设置得太低将导致提供有用信息的离子全部聚集在一起,因为至少某些碎片离子的质谱特征可能很弱或不存在。在任何一种情况下,都将无法从产物离子质谱中获得有关前体离子的充足的结构信息,从而无法提供鉴定或结构(或序列)说明。在任何给定的CE下,不同大小、结构和电荷容量的分析物以不同程度进行解离。因此,在自动质谱分析实验过程中,对所有前体离子仅使用单个碰撞能量设置,会带来某些离子的碎裂程度不理想或不可接受的风险。尽管如此,由于多种原因(例如,电离、色谱、碎裂等),经常对化学多样性降低的样品或样品部分执行质谱分析程序。减少化学多样性会增加通过调整相似分析物上的碰撞能量来设置适当的碰撞能量的可能性。

[0012] 尽管共振激发CID(RE-CID)和HCD根据相同蛋白质的相同电荷产生相似的质谱,但

是产生最大数量的结构信息所需的精确碰撞能量最佳值可能会有很大差异。在RE-CID的情况下,由于所施加的辅助频率与前体离子的运动处于相同的基本频率,因此前体离子的内部能量会增加,以至于达到最小的解离能量并且产生产物离子。随着所施加的能量的增加,碎裂程度达到最大值,并且随着前体离子的耗尽而达到平稳。如果进一步增加所施加的碎裂能量,则各个产物离子的相对丰度通常不会发生变化。相反,随着碎裂能量增加超过平稳区域的起始点,产物离子的相对丰度保持近似恒定,并且从这一过程中获得的附加相关结构信息很少甚至没有。

[0013] 相反,在HCD碎裂的情况下,碰撞活化过程仅是HCD池与相邻离子光学元件之间的电势差的函数。因此,在HCD池中形成的任何产物离子都可能根据其多余的内部能量进行进一步的碎裂。由于HCD工艺涉及使用氮气作为碰撞气体,而不是RE-CID实验中通常使用的氦气,因此,如果施加了近乎最佳的碰撞能量,则可以从HCD工艺获得更高的能量和更多的结构信息。在RE-CID工艺中,将所施加的碰撞能量增加至超过其最佳值会减少残留的前体离子的量,但不会显著改变相对的碎片离子的量。在HCD碎裂中,将所施加的碰撞能量增加至超过其最佳值通常会导致碎片离子进一步碎裂。

[0014] 图3A示出了增加能量对通过HCD碎裂产生的可识别的蛋白质碎片离子的数量的影响(曲线151)与增加能量对通过RE-CID碎裂产生的这种可识别离子的数量的影响(曲线152)之间的一般比较。曲线152示出了改变所施加的共振能量对源自蛋白肌红蛋白的前体离子的碎裂的影响。在这个实例中,当碰撞能量增加到超过25%RCE时,结构信息量将保持相对恒定。相反,当采用HCD工艺时(曲线151),对于大约28%RCE的HCD能量,获得的结构信息含量有一个明确定义的最大值。在碰撞能量小于或超出所述最佳RCE设置的情况下,从HCD实验获得的结构信息的质量可能会急剧下降。

[0015] 如图3B-3D的产物离子质谱所示,改变所施加的HCD碎裂能量的效果在蛋白质泛素的+8电荷态前体离子的碎裂中得到了很好的说明。图3B示出了使用25%的次优RCE设置时,离子碎裂产生的碎片离子的有限数量。在许多实验情况下,这种有限的碎裂将无法通过搜索标准串联质谱库或使用来自可用数据库的序列信息对蛋白质进行正确鉴定。但是,当将RCE设置更改为30%时,相同前体离子的HCD碎裂效果最佳,并且所得的产物离子质谱(图3C)显示了丰富的各种电荷态的碎片阵列,所述阵列使得能够使用几种方法中的任何一种对蛋白质进行鉴定。最后,如图3D所示,将RCE设置进一步增加到40%会导致过度碎裂的情况,在这一情况下,大多数生成的产物离子是单电荷的低质量碎片,其比实际的蛋白质序列本身更能反映蛋白质的氨基酸组成。因此,令人高度期望的是,实时调整用于未知蛋白质和复杂混合物的HCD碎裂的碰撞能量,以使可用信息含量最大化。

[0016] 在发明人Schwartz等人名下的美国专利第6,124,591号描述了一种通过RE-CID在四极离子阱中产生产物离子的方法,在所述方法中,所施加的共振激发电压的幅度与前体离子 m/z 比基本上线性相关。美国专利第6,124,591号中描述的技术试图将不同离子的最佳共振激发电压幅度的主要变化以及由于仪器差异而引起的变化归一化。Schwartz等人进一步发现,不同结构、电荷态和稳定性的贡献对测定所施加的碰撞能量的影响本质上是次要的,并且这些次要影响可以通过简单的校正系数进行建模。

[0017] 根据Schwartz等人的教导,可以逐个仪器地对施加的最佳CE与 m/z 之间的基本上线性关系进行简单且快速的校准。附图4A示意性地示出了校准曲线的产生和使用的原理。

最初,通过将线性关系拟合到校准数据来生成特定质谱仪的校准曲线,在所述校准曲线中观察到前体离子强度降低的特定百分比(如降低90%)。所述线性关系在图4A中以线22示出。Schwartz等人发现,两点校准足以表征线性关系,并且更简单地说,如果线的截距固定为某个值或零,则可以使用单点校准。在典型的校准中,校准线22的截距被假定为在原点,如图4A所示,并且单点校准包含以指定的参考质荷比(m/z)₀在参考点29处测定或计算所施加的碰撞能量。通常,参考点在 $m/z=500$ Da处,并且在校准过程中在500Da处测量的或外推到500Da的参考碰撞能量值可以表示为 CE_{500} 。

[0018] 一旦确定了仪器校准,则质谱仪的后续操作通常不采用线22表示的全CE值,而是采用相对碰撞能量(RCE)值,相对碰撞能量值表示为在任何给定的 m/z 下,线22表示的值中CE值的百分比。例如,图4A中所示的线24、26和28分别表示75%、50%和25%的RCE值。随后,用户可以简单地指定RCE的期望值。简单的标量电荷校正系数 $f(z)$ 解释了前体离子电荷态 z 对所施加的最佳CE的次要影响。已经发现,这些最初在RE-CID碎裂中测定的一般关系也对于HCD碎裂有效。通过这些简化,然后根据以下等式自动设置施加于每个前体的用于HCD碎裂的绝对碰撞能量 CE_{actual} (以电子伏特表示):

$$[0019] \quad CE_{actual} = RCE \times CE_{500} \times [(m/z)/500] \times f(z) \quad \text{等式 1}$$

[0020] 其中 CE_{actual} 是所施加的碰撞能量,通常以电子伏特(eV)表示,RCE是相对碰撞能量——通常由用户为每个实验定义的百分比值,并且 $f(z)$ 是电荷校正系数。图4B中的表1列出了可接受的电荷校正系数。注意,括号中的分数的分子和分母均以道尔顿、Da(或更准确地说,汤姆森、Th)为单位表示。尽管这个等式通常足以在狭窄的前体离子特性范围内微调施加于样品的绝对CE,但应注意,由于当 $z \geq 5$ 时 $f(z)$ 产生固定值,碰撞能量对于具有较高电荷态的较重分子(如蛋白质和多肽)而言过高,从而导致这些种类的过度碎裂。

[0021] 近来,完整蛋白质和多肽的质谱分析已获得广泛普及。对于此类应用,样品内的分析物的尺寸、结构和电荷容量可能会发生巨大变化,因此需要非常不同的碰撞能量才能实现相同程度的解离。已经发现,即使电荷系数的范围被扩展并外推至+5以上的电荷态,上述等式也不能充分地将多肽或完整蛋白样品中所有前体的碰撞能量归一化。因此,这些特定的分析物需要修改后的模型。

发明内容

[0022] 本教导涉及建立新的解离参数,所述参数将用于测定实现给定分析物前体离子的期望解离程度所需的HCD(碰撞池类型CID)碰撞能量(CE)。所述选择仅基于分析物前体离子的分子量(MW)和电荷态(z)。为此,发明人设计了两个不同的指标,其可以用作“解离程度”D的测量,并代替先前使用的相对碰撞能量和归一化碰撞能量参数。这两个新指标是相对前体衰变(D_p)和谱熵(D_E),尽管未来可以想象描述解离程度的其它指标。发明人进一步针对实现每个这样的指标的一系列值所需的碰撞能量值开发了预测模型。每个模型都仅是前体离子的MW和 z 的简单平滑函数。通过结合能够测定分析物分子的分子量的实时光谱去卷积算法,这些新的技术将能够通过以前体依赖性方式对碰撞能量进行自动实时选择从而控制解离程度。通过这些新颖的碰撞能量测定方法,发明人消除了用户针对不同化合物或应用“调整”或“优化”碰撞能量的必要性,因为单个“解离程度”参数设置将适用于所有采样的MW和 z 。这种功能对于完整蛋白质分析是有利的,在这种情况下,前体可以在单个样品中涵盖广

泛的物理特性。现有方法是有限范围的分析物特性(如简单肽的特性)量身定制的,并且无法充分解决完整蛋白质和多肽分析的复杂性问题。

附图说明

[0023] 为了进一步阐明本公开的上述和其它优点及特征,将通过参考附图中示出的具体实施例来对本公开进行更具体的描述。应当理解,这些附图仅描绘了本公开的示出实施例,并且因此不应被视为限制其范围。将通过使用附图以另外的特征和细节来描述和解释本公开,在附图中:

[0024] 图1A是显示出良好分离的色谱洗脱峰的两个分析物部分的分析的示意图;

[0025] 图1B是具有高度重叠的洗脱峰的色谱图的一部分的示意图,所述洗脱峰均在分析阈值之上;

[0026] 图1C是两个同时洗脱的蛋白质或多肽分析物的假设的多个交错的质谱峰的示意图;

[0027] 图2是用于通过碰撞诱导解离使离子碎裂的常规设备和方法的示意图;

[0028] 图3A是增加能量对通过HCD碎裂产生的可识别的蛋白质碎片离子的数量的影响与增加能量对通过RE-CID碎裂产生的这种可识别离子的数量的影响之间的一般图形比较。

[0029] 图3B、3C和3D是使用分别为25、30和40的相对碰撞能量设置通过对蛋白质泛素的+8电荷态前体离子进行HCD碎裂而产生的碎片离子的质谱。

[0030] 图4A是示出了根据已知的“归一化碰撞能量”操作技术的所施加的碰撞能量与前体离子质荷比之间的关系图;

[0031] 图4B是示出了校正系数的表,所述校正系数应用于已知的归一化碰撞能量操作技术以补偿前体离子电荷态对通过碰撞诱导解离产生的碎裂程度的影响;

[0032] 图5A是根据本教导的用于产生并自动分析色谱/质谱的系统的示意图;

[0033] 图5B是适于结合根据本教导的方法使用的示例性质谱仪的示意图,所述质谱仪包括混合系统,所述混合系统包括四极质量过滤器、双压四极离子阱质量分析器和静电阱质量分析器;

[0034] 图6A是碎裂后剩余的各个前体离子种类随施加的碰撞能量而变化的百分比与通过逻辑回归图产生的数据拟合的一组图形集,其中前体离子种类是分子量约为29k道尔顿的碳酸酐酶的正22、+24、+26和+28电荷态;

[0035] 图6B是可以用于根据本教导的模型来计算碰撞能量的参数的表,所述碰撞能量应当通过实验来提供以产生各个期望的前体离子存活百分比 D_p ,按各个选定的 D_p 值制成表格。

[0036] 图7A是具有不同碰撞诱导解离程度的五个代表性的产物离子质谱的组合,其示出了根据本教导计算的“总质谱熵”值的变化;

[0037] 图7B是将两个产物离子质谱中的每个质谱划分成两个区域并测定与每个第一区域相关联的第一质谱熵 E_1 和与每个第二区域相关联的第二质谱熵 E_2 并在 E_1 、 E_2 和总质谱熵 E_{tot} 之间进行比较的实例;

[0038] 图8A是根据本教导的根据产物离子质谱计算的总质谱熵(上图)、 E_1 (中图)和 E_2 (下图)随赋予肌红蛋白(约17k道尔顿)的指示性前体离子电荷态的碰撞能量而变化的一组图;

[0039] 图8B是可以用于根据本教导的另一模型来计算碰撞能量的参数的表,所述碰撞能

量应当通过实验来提供以产生根据产物离子熵参数 D_E 分布的产物离子的集合,按各个选定的 D_E 值制成表格。

[0040] 图9A是随质荷比变化的常规计算的碰撞能量(实线)与根据本教导的熵模型计算的碰撞能量(虚线)之间的比较,并且针对+5的离子电荷态和常规相对碰撞能量的默认设置。

[0041] 图9B是按比例缩放的常规计算的碰撞能量(实线)与根据本教导的熵模型计算的碰撞能量(虚线)之间的比较,其中图9A中的常规计算的碰撞能量按0.79475的比例系数进行缩放。

[0042] 图10是可以应用于常规计算的碰撞能量以使那些常规计算的碰撞能量与根据本教导测定的某些计算结果一致的电荷态缩放系数的图;

[0043] 图11是图10中以图形方式描绘的电荷态缩放系数的表格形式;

[0044] 图12是根据本教导的用于使用自动碰撞能量测定对蛋白质或多肽进行串联质谱分析的方法的流程图;

[0045] 图13A是计算机屏幕信息显示器的图示,其示出了由采用根据本教导的方法的计算机软件所产生的,根据由细胞色素-c、溶菌酶、肌红蛋白、胰蛋白酶抑制剂和碳酸酐酶组成的五组分蛋白质混合物的质谱计算得出的峰簇分解结果;并且

[0046] 图13B是计算机屏幕信息显示器的图示,其示出了由采用根据本教导的方法的计算机软件所产生的峰簇分解结果,所述显示器示出了图13A中所示的分解结果的扩展部分。

[0047] 图A1示出了通过附录中教导的方法研究的质谱和一系列 m/z 值。

具体实施方式

[0048] 提供以下描述以使本领域的任何技术人员都能够制造和使用本发明,并且以下描述是在特定应用及其要求的背景下提供的。对所描述的实施例的各种修改对于本领域技术人员而言将是显而易见的,并且本文的一般原理可以应用于其它实施例。因此,本发明不旨在局限于所示出的实施例和实例,而是旨在符合根据权利要求书的尽可能宽的范围。参照附图1-13,结合以下讨论结合,本发明的特定特征和优点将变得更加明显。

[0049] 图5A是可结合本教导的方法采用的用于产生并自动分析色谱/质谱的通用系统30的示意性实例;根据众所周知的色谱原理,如液相色谱仪、高效液相色谱仪或超高效液相色谱仪等色谱仪33接收分析物混合物的样品32,并将分析物混合物至少部分地分离成单独的化学组分。在不同的相应时间将所得的至少部分分离的化学组分转移至质谱仪34以进行质量分析。当质谱仪接收每个化学组分时,所述化学组分被质谱仪的电离源112电离。电离源可以产生包括多个离子种类(即,多个前体离子种类)的多个离子,所述多个离子种类包括与每个化学组分不同的电荷或质量。因此,可以针对每个化学组分产生具有不同的相应质荷比的多个离子种类,每个这样的组分在其自身的特性时间从色谱仪中洗脱。通常通过空间或时间分离,通过质谱仪的质量分析器139对这些各个离子种类进行分析,并通过检测器35对其进行检测。作为此过程的结果,可以根据离子种类的各种质荷比(m/z)适当地鉴定所述离子种类。如图5A所示,质谱仪包括反应池23,所述反应池用于碎裂前体离子或引起前体离子的其它反应,从而产生包括多个产物离子种类的多法产物离子。

[0050] 仍参考图5A中,可编程处理器37电连接到质谱仪的检测器,并在一个或多个样品

的色谱/质谱分析期间接收由检测器产生的数据。可编程处理器可以包括单独的独立计算机,或者可以仅包括电路板或由固件或软件操作的任何其它可编程逻辑装置。任选地,可编程处理器也可以电连接到色谱仪和/或质谱仪,以便将电子控制信号传输到这些仪器中的一个或另一个,从而控制其操作。此类控制信号的性质可能响应于从检测器传输到可编程处理器的数据或通过根据本教导的方法执行的对所述数据的分析来确定。可编程处理器也可以电连接到显示器或其它输出38,以将数据或数据分析结果直接输出给用户或电子数据存储装置36。图5A所示的可编程处理器通常可操作为:从色谱/质谱仪设备接收前体离子色谱/质谱和产物离子色谱/质谱,并根据下面讨论的各种方法自动执行各种仪器控制、数据分析、数据检索和数据存储操作。

[0051] 图5B是可用于执行根据本教导的方法的特定示例性质谱仪200的示意图。图5B所示的质谱仪一种混合质谱仪,其包括一种以上类型的质量分析器。具体地说,质谱仪200包含离子阱质量分析器216以及Orbitrap™分析仪212,所述Orbitrap™分析仪是静电阱质量分析器的一种。Orbitrap™质量分析器212采用图像电荷检测,其中通过检测离子阱内离子的运动在电极上感应的图像电流来间接检测离子。根据本教导的各种分析方法采用多个质量分析数据采集。因此,可以通过同时使用两个或多个分析仪有利地采用混合质谱仪系统来改善占空比。然而,图5B所示类型的混合动力系统是不需要的,并且可以在能够进行串联质谱法并且采用碰撞诱导解离的任何质量分析器系统上采用根据本教导的方法。合适类型的质量分析器和质谱仪包含但不限于三重四极杆质谱仪、四极杆飞行时间(q-TOF)质谱仪和四极杆Orbitrap™质谱仪。

[0052] 在质谱仪200的操作中,电喷雾离子源201将待分析样品的离子提供到撇渣器202的孔中,在所述孔处离子进入第一真空室。进入后,离子被堆叠环式离子导向器204捕获并聚焦为紧密束。第一离子光学传递组件203a将束传递到质谱仪的下游高真空区域中。大多数剩余的中性分子和不期望的高速离子簇(如溶剂化离子)通过弯曲的束导向器206与离子束分离。中性分子和离子簇遵循直线路径,而所关注的离子则在阻力场的作用下弯曲90度,从而产生分离。

[0053] 质谱仪200的四极质量过滤器208在其常规意义上用作可调质量过滤器,以便仅使离子在选定的狭窄 m/z 范围内通过。随后的离子光学传递组件203b将过滤后的离子传递到弯曲的四极离子阱(“C-阱”)组件210。C-阱210能够沿着四极质量过滤器208与离子阱质量分析器216之间的路径传递离子。C-阱210还具有临时收集并储存大量离子,然后将这些离子作为脉冲或小包传递到Orbitrap™质量分析器212中的能力。通过在C-阱210与设置在C-阱210和Orbitrap™质量分析器212之间的一组注入电极211之间施加电势差来控制离子小包的传递。C-阱的曲率被设计成使离子群在空间上聚焦,以匹配Orbitrap™质量分析器212的入射孔的角谱宽度。

[0054] 多极离子导向器214和光学传递组件203b用于引导在C-阱210与离子阱质量分析器216之间的离子。多极离子导向器214提供临时的离子存储能力,使得稍后可以在后续步骤中检索在分析方法的第一处理步骤中产生的离子以进行处理。多极离子导向器214也可以充当碎裂池。沿着C-阱210与离子阱质量分析器216之间的路径的各个栅电极是可控制的,使得离子可以在任一方向上转移,这取决于任何特定分析方法中所需的离子处理步骤的顺序。

[0055] 离子阱质量分析器216是包括高压线性阱池217a和低压线性阱池217b的双压四极线性离子阱(即二维阱),这两个池彼此相邻,通过具有小孔的平板透镜隔开,所述小孔允许两个池之间的离子转移(这产生了泵送限制)并允许在两个阱中保持不同的压力。高压池217a的环境有利于离子冷却和通过碰撞诱导解离或电子转移解离或离子-离子反应(如质子转移反应)引起的离子碎裂。低压池217b的环境有利于具有高分辨力和质量精度的分析扫描。低压池包含双倍增极离子检测器215。

[0056] 如图5B所示,质谱仪200进一步包含控制单元37,所述控制单元可以通过电子链连接到系统200的各个组件。如先前所讨论的图5A所示,控制单元37可以连接到一个或多个另外的“前端”设备,所述设备将样品供应给质谱仪200并且可以在将样品材料供应给质谱仪之前执行各种样品制备和/或分离步骤。例如,作为控制液相色谱仪的操作的一部分,控制器37可以控制液相色谱仪内的流体的总体流量,包含将各种试剂或流动相施加到各个样品上。控制单元37还可以充当数据处理单元,以例如处理(例如,根据本教导)来自质谱仪200的数据或将数据转发到一个或多个外部服务器以进行处理和存储(未示出外部服务器)。

[0057] 用于模型开发的数据采集

[0058] 在以下11种蛋白质标准品上收集了解离质谱数据(MS/MS串联质谱数据):泛素(~8kDa)、细胞色素c(~12kDa)、溶菌酶(~14kDa)、RNase A(~14kDa)、肌红蛋白(~17kDa)、胰蛋白酶抑制剂(~19kDa)、利妥昔单抗LC(~25kDa)、碳酸酐酶(~29kDa)、GAPDH(~35kDa)、烯醇酶(~46kDa)和牛血清白蛋白(~66kDa)。通过直接注入引入样品,并通过电喷雾电离使样品电离。由于这些蛋白质众所周知的碎裂图型和作为典型的自上而下蛋白质标准品的性能,选择这些蛋白质来构建模型。通过HCD解离,选择每种蛋白质的大约10个电荷态进行MS/MS分析。在这些实验中,每个前体离子的绝对碰撞能量CE均按照1电子伏特(eV)步长从5到50eV的绝对碰撞能量变化。从这些衰变曲线中,获得了每种分析的电荷态的逻辑回归图。计算每个质谱的度量值 D_p 和 D_E ,然后使用这些值针对实现一系列D值所需的CE(即前体MW和z的函数)建立预测模型。

[0059] 前体衰变模型

[0060] 方法1

[0061] 对于每种蛋白质标准品,在每个前体离子电荷态z下,在每个绝对碰撞能量(CE)下,计算相对于测得的总离子电流 D_p 的剩余前体离子强度。 D_p 随CE的变化遵循如图6A所示的标准衰变曲线,其中衰变曲线302、304、306和308分别表示碳酸酐酶的正22、+24、+26和+28电荷态的前体离子衰变曲线。发明人通过逻辑回归对变化进行建模

[0062] $CE = c + (1/k) [\ln(1/D_p) - 1]$ 等式2

[0063] 其中参数c表示相对前体剩余为50%时的CE,而参数k为c处的-斜率。对应于z=+24的图6A的曲线304包含另外的标记,以进一步描绘针对所述特定电荷态的参数c和k的计算。具体地说,点311是曲线304与50%阈值相交的点,因此,参数c定位在约17.6eV处。此外,线313是曲线304在点311处的切线。因此,将参数k确定所述该切线的斜率。在计算上,通过最小二乘法拟合计算出的相对剩余强度,得到c和k的值。最佳拟合参数取决于蛋白质标准品的分子量MW以及蛋白质碎裂时的电荷态z。可以将参数c和k建模为MW和z的幂的简单乘积。再次使用最小二乘拟合获得c和k的最佳拟合幂,如下所示。

[0064] $c = 0.0018 \times MW^{1.6} \times z^{-2.2}$ 等式3

[0065] $k = 0.00025 \times MW^{1.7} \times z^{1.9}$ 等式4

[0066] 使用方法1,一旦确定了分子量MW和电荷z(如下所述),就可以根据等式3和4确定c和k参数的值。然后,对于任何期望的残留前体离子百分比 D_p ,可以通过等式2使用计算出的c和k值来计算必须施加的所需碰撞能量CE。

[0067] 方法2

[0068] 在通过等式2的逻辑回归对每个衰变曲线建模的步骤之后,第二种方法不同于上述的“方法1”。第二种方法不将参数c表示为两个变量MW和z的独立函数并同样将参数k表示为相同的两个自变量的另一个独立函数,而是采用了更逐步的策略。在这种方法中,首先指定了剩余相对前体强度的目标百分比 D_p 。然后,采用等式1(使用根据各个衰变曲线确定的c和k值)来将所有CE、MW和z值汇总成表,这些值组合起来会增加目标前体离子百分比 D_p 。然后,使用最小二乘拟合获得所述目标处CE的函数形式,即MW和z的幂的乘积。以这种方式,对于每个所关注的 D_p ,可以获得适当的CE的更定制化的模型。在这种定制化模型中,可以根据以下形式的一组等式计算实现一定百分比 D_p 的前体离子存活率所需的碰撞能量(CE):

[0069] $CE(D_p) = a_1 \times MW^{a_2} \times z^{a_3}$ 等式5

[0070] 其中 a_1 、 a_2 和 a_3 是可以针对所关注的各个 D_p 值中的每一个进行预先计算并制成表格的参数。提供这些参数针对 D_p 的各个选定值的取值的表,如附图6B中提供的表2所示。

[0071] 熵模型

[0072] 对于质心产物离子质谱,定义了解离程度的另一个度量,即总谱熵,如下:

[0073] $E_{total} = \sum_i P_i \ln(P_i)$ 等式6

[0074] 其中 p_i 是指数i的质谱峰(以m/z为单位)的质心强度(或面积),其通过所有这些峰的总强度(或面积)或总离子电流TIC进行归一化。合计是质谱中所有质心的总和(所有i)。如上所述,发现HCD产物离子谱的总谱熵的计算值紧密反映了数据中观察到的解离程度,直至 E_{total} 的值约为0.7,此时离子电流的位置变为重要考虑因素(图7A)。为了增强区分(或解析)“理想解离”到过度碎片化范围(高总谱熵)的能力,将总熵分为第一部分熵(E_1)和第二部分熵(E_2),其中 E_1 代表从最小值m/z到前体离子m/z的一半的MS/MS谱图区域的熵,而 E_2 代表从前体离子m/z的一半到最终m/z的谱图区域的熵(图7B)。因此,使用等式6计算 E_1 ,仅使用了 E_1 区域内的m/z峰质心的 p_i 值,并同样使用等式6计算 E_2 ,仅求和了 E_2 区域内的m/z峰质心的 p_i 值。 E_1 和 E_2 二者的计算中 p_i 的计算中的分母再次是质谱的总离子流(E_1 和 E_2 区域)。

[0075] 图8A中示出了针对选定的肌红蛋白的前体离子电荷态计算的 E_{total} 、 E_1 和 E_2 ,所述肌红蛋白是来自模型数据集的大约17kDa的蛋白质。曲线426、526和626分别表示针对肌红蛋白的+26电荷态计算的 E_{total} 、 E_1 和 E_2 ,即所施加的碰撞能量的函数。同样,曲线424、524和624分别表示针对肌红蛋白的+24电荷态计算的 E_{total} 、 E_1 和 E_2 ,即所施加的碰撞能量的函数。同样,曲线421、521和621分别表示针对肌红蛋白的+21电荷态计算的 E_{total} 、 E_1 和 E_2 ,即所施加的碰撞能量的函数。同样,曲线417、517和617分别表示针对肌红蛋白的+17电荷态计算的 E_{total} 、 E_1 和 E_2 ,即所施加的碰撞能量的函数。最后,曲线415、515和615分别表示针对肌红蛋白的+15电荷态计算的 E_{total} 、 E_1 和 E_2 ,即所施加的碰撞能量的函数。

[0076] 考虑到所有蛋白质图,观察到:(a) E_1 值在所关注的CE范围内单调增加;(b) E_1 曲线比 E_2 曲线平滑得多,并且(c)所有 E_1 曲线都可以通过逻辑回归很好地建模。仅使用 E_1 数据的

缺点是曲线相对无特征,因此很难标准化不同的 E_1 值。但是,利用了以下事实:每个 E_2 曲线几乎总是包含一个明确定义的最大值,所述最大值用于定义每种蛋白质标准品的每种电荷态的参考CE。同样地,本发明人已经对MW、前体z和 E_2 曲线中的最大值处的CE值之间的关系进行了建模,这导致了以下等式7:

$$[0077] \quad CE^{E2_{max}} = 0.1 \times MW^{0.93} \times z^{-1.5} \quad \text{等式7}$$

[0078] 现在将这组参考CE值应用于 E_1 曲线,就可以确定与每种蛋白质标准品的每种电荷态的 E_2 最大值相对应的 E_1 值。此外,通过对每个 E_1 曲线进行逻辑拟合,可以针对每个标准品的每个z定义产生任何期望的参考熵的分数值的CE。所述分数参考熵变成新参数 D_E 。具体地说,参数 D_E 是针对任何特定的z定义的,如

$$[0079] \quad D_E = E_1 / E_1^{E2_{max}} \quad \text{等式8}$$

[0080] 其中, $E_1^{E2_{max}}$ 是第一部分熵 E_1 的值,即碰撞能量 $CE^{E2_{max}}$ 的值,所述值与第二部分熵 E_2 中的最大值相关。可以将CE值的任何特定分数熵值的集合拟合成类似于等式7的幂函数形式,所述等式以一般形式写成:

$$[0081] \quad CE(D_E) = b_1 \times MW^{b_2} \times z^{b_3} \quad \text{等式9}$$

[0082] 其中 b_1 、 b_2 和 b_3 是可以针对 D_E 的各个值进行预先计算并制成表格的参数,如附图8B中出现的表3所示。正如预期的那样,在 $D_E = 1$ 处,我们重新获得了等式6。还可以容易地将谱熵的概念扩展以捕获解离。例如,首先对产物离子谱执行m/z到质量的去卷积步骤,以获得产物离子的电荷和分子量,而不是仅仅根据m/z分布计算熵。可以分别基于产物离子分子量和电荷的分布容易地定义分子量熵和电荷态熵。

[0083] 可以采用上述等式9来确定在HCD碎裂过程中通过实验施加的碰撞能量的值,以便产生产物离子m/z值的展度,所述值对应于根据上述讨论计算的熵参数 D_E 的给定值。据发明人所知,这是首次提出基于产物离子集合的期望性质施加碰撞能量的模型的实例。本发明不限于使用特定度量(D_E)来表示产物离子的分布或展度,因为在某些特定情况下,产物离子m/z展度的其它替代性度量可能是有利的。

[0084] 在表3的每一行中列出的 b_1 、 b_2 和 b_3 值与某个产物离子展度(“熵分数”) D_E 相关,所述展度由等式8给出,其中 D_E 的范围是 $\{0.1, 0.2, \dots, 2.0\}$ 。默认级别1.0对应于碎裂谱图的最大熵 E_{max} ,并且通过对MW、z和观察到 E_{max} 的碰撞能量之间的关系进行建模获得相应的参数集。低于和高于1.0的水平与 E_{max} 的分数相关,并且可以分别进行建模以分别为较低和较高的破碎度提供最佳的碰撞能量。通常,在对含有未知化合物的样品进行实验或分析之前,如上所述,可能有必要通过获取已知标准的初始测试数据来确定任何特定仪器的参数 p_1 、 p_2 、 p_3 (即执行校准)。

[0085] 实时精细校准

[0086] 仪器与仪器之间的微小可变性和任何特定仪器的时间漂移都是可以预料的。考虑到这一点,提供了一种自动校正可变性的机制,所述机制导致任何给定模型的固定偏移。例如,在给定熵模型的情况下,如果 D_E 设置为0.68,并且最新质谱(如100个最新质谱)的滚动平均值 D_E 相差大于此值的 $\pm 15\%$,则系统应自动调整以使实际测得的 D_E 更接近要求的“目标” D_E 。我们预期一个简单的乘法校正系数就足够了,而无需更改基本方程的系数。

[0087] 使常规电荷态校正系数适应新方法

[0088] 图9A示出了使用如美国专利第6,124,591号中所述的归一化碰撞能量(NCE)方法按 $z=5$ 常规计算的碰撞能量(曲线703)与根据熵模型使用1.0的熵分数 D_E 计算的碰撞能量的35%相对碰撞能量(RCE)(曲线704)之间的比较。为了进行熵模型计算,将分子量计算为 $(m/z-1.007) \times z$ 。与NCE曲线(按照定义是一条直线)一样,根据熵模型计算的曲线在相关的500..2000的 m/z 范围内似乎是线性的。因此,应该有可能对NCE曲线应用缩放系数,以获得与根据熵模型计算的碰撞能量值的趋势匹配的拟合曲线。实际上,拟合曲线705与熵模型曲线很好地匹配(图9B)。可以使用曲线拟合以基本相同的拟合优度(数据未示出)对1..100范围内的所有电荷态执行这种缩放。

[0089] 前5个电荷态的所得缩放系数显著低于1,这意味着与使用默认RCE值35%的标准NCE方法相比,熵模型往往会分配更低的碰撞能量。因此,由拟合得出的 $z = \{1..5\}$ 的缩放系数与归一化碰撞能量模型中使用的常规校正系数显著不同,并且预计“中间”电荷态在6..10左右的范围内会有类似的偏差(当将RCE校正系数外推到 >5 的较高电荷态时)。然而,出于兼容性原因,应避免将已建立的校正系数(表1)更改为低电荷态。

[0090] 为了解决这个问题,将两种方法组合如下:将常规校正系数的曲线以-0.05的步长外推,直到与本文通过曲线拟合确定的缩放系数的曲线相交为止。在 $z \approx 10$ 处观察到所述相交,这标志着传统方法向本文所述的新型熵方法的转变。所得的缩放系数在图10中被示出为曲线708a和708b。因此,所得的扩展NCE曲线(图10,曲线708a和708b)定义如下:

[0091] • 对于 $z = \{1..5\}$,使用表1中给出的常规校正系数。

[0092] • 对于 $z = \{6..10\}$,通过以0.05的步长降低最后一个值 $f(5) = 0.75$ 来外推校正系数,即 $f(z = \{6..10\}) = \{0.70, 0.65, 0.60, 0.55, 0.50\}$ 。

[0093] • 对于 $z > 10$,校正系数由上述拟合得出的缩放系数给出,并归一化为所应用的NCE校正系数0.75(以避免使用双重缩放)。

[0094] 扩展的NCE系数在图11中示出的表4中给出。

[0095] 分子量计算方法的实例的总结

[0096] 上述模型需要对分析物的分子量(MW)有所了解,以便估计用于使所述分析物的选定离子碎裂的最佳碰撞能量。在通过电喷雾电离而将蛋白质和多肽分子的离子电离的情况下,离子主要包括具有多个加合质子的完整分子。在这种情况下,每个主要分析物离子种类上的电荷仅等于加合质子的数量。在这种情况下,至少在理论上可以容易地确定分子量,条件是可以鉴定质谱中表示的各个多重质子化的分子离子种类并根据这些离子种类的分子起源将其分配给基团(即,电荷态系列)。令人遗憾的是,这种鉴定和分配的过程通常由于典型的质谱通常包含代表多个重叠的电荷态序列的线而变得复杂,并且由于给定电荷态的每个离子种类的特征可以通过同位素变化来划分而变得更加复杂。

[0097] 由于生物来源的样品通常非常复杂,因此单个MS质谱可以容易地包含数百个甚至数千个属于不同分析物的峰——这些峰在给定的 m/z 范围内交织在一起,其中强度非常不同的离子信号重叠并互相压制。由此带来的计算挑战是将每个峰追溯到某个或某些分析物。消除“噪声”和确定正确的电荷分配是应对这一挑战的第一步。一旦确定了峰的电荷,则可以进一步使用电荷态序列中的电荷态之间的已知关系来对与分析物相关的电荷态进行分组。这一信息可以进一步用于在被最佳描述为数学分解(在本领域中也称为数学去卷积)的过程中确定一个或多个分析物的分子量。

[0098] 此外,必须“实时”(即,在获取质谱数据时)执行鉴定各个重叠的电荷态序列所需的数学去卷积,因为前体离子质谱的去卷积结果被立即用于选择要解离的离子种类并确定在解离过程中要施加的适当的碰撞能量,其中所施加的碰撞能量可能因不同的种类而不同。为了获得成功,需要一种预测每个离子种类的多条质谱线的数据采集策略和一种优化的实时数据分析策略。通常,去卷积过程应在不到一秒钟的时间内完成。在美国授权前公开第2016/0268112A1号(其公开内容通过引用整体并入本文)中描述了一种算法,所述算法在这种时间限制内实现了复杂样品的所需分析,并作为应用软件运行。可替代地,于2016年9月9日提交的共同未决的欧洲专利申请第16188157号教导了用于另一种合适的数学去卷积算法的方法。前述欧洲专利申请的文本作为本文件的附录被包括在内,并且其附图作为图A1被包括随附图纸集中。所述算法可以被编码到与质谱仪相连的硬件处理器中,从而运行得更快。以下段落简要总结了上述专利申请公开第2016/0268112A1号中描述的计算去卷积算法的一些主要特征。

[0099] 仅使用质心。

[0100] 标准质谱电荷分配算法使用质谱中的线的完整轮廓数据。相反,美国授权前公开第2016/0268112A1号中描述的计算方法使用质心。使用质心而不是线轮廓的主要优点是数据减少。通常,轮廓数据点的数量比质心的数量大约一个数量级。与所述标准分配方法相比,任何使用质心的算法都将在计算效率上获得显著的优势。对于需要实时电荷分配的应用,优选地设计仅需要质心数据的算法。使用质心的主要缺点是 m/z 值不精确。如质量准确性、分辨率和峰提取效率等因素往往都会损害质心数据的质量。但是,通过将 m/z 不精确性考虑进采用质心数据的算法中,可以大大缓解这些担忧。

[0101] 强度是二进制的。

[0102] 如美国授权前公开第2016/0268112A1号所述,质谱线强度被编码为二进制(或布尔(Boolean))变量(真/假或存在/不存在)。布尔方法仅考虑质心强度是否高于阈值。如果强度值满足基于信号强度或信噪比或两者的用户可设置标准,则所述强度值将采用布尔“True”值,否则将分配“False”值,不论强度的实际数值是多少。使用布尔值的一个众所周知的缺点是信息丢失。然而,如果可以使用大量的数据点——例如,典型的高分辨率质谱中的数千个质心,那么布尔变量的数量远足以补偿强度信息的损失。因此,所引用的去卷积算法利用这种数据丰度来实现效率和准确性。

[0103] 在替代性实施例中,可以通过使用近似强度值而不仅是布尔真/假变量,在无显著计算速度损失的情况下实现另外的精确性。例如,可以设想仅将相似高度的峰相互比较的情况。通过将强度值离散化为少量的低分辨率箱(例如“低”、“中”、“高”和“非常高”),可以轻松容纳添加的信息。这种分箱可以在不牺牲非常简化表示的强度的计算简单性的情况下,实现具有“高度信息”的良好平衡。

[0104] 为了达到与单独使用布尔变量时相当的计算效率同时仍并入强度信息,一种方法是将强度编码为字节,字节大小与布尔变量相同。通过在计算中使用强度的对数(而不是原始强度)以及合适的对数基数,可以轻松实现这一目标。可以将强度的对数进一步转换为整数。如果适当选择对数底数,则对数(强度)值都将舒适地落在值0-255的范围内,所述范围可以表示为一个字节。另外,可以通过仔细选择对数底数来最小化将双精度变量转换为整数时的舍入误差。

[0105] 为了进一步使可能由字节算法(而不是布尔算法)引起的任何性能下降最小化,用于分离或分组质心的计算可能仅需要计算强度比,而不是字节值强度本身。比率的计算非常有效,因为:1)比率的 \log_2 不使用浮点除法,而是简单的对数之差,在这种情况下,仅转换为两个字节的减法;以及2)要从对数值的差异中重新获得准确的比率,只需对对数值的差异进行取幂即可。由于此类计算将仅遇到有限的预定义数字集的指数(即2个字节(-255到+255)之间的所有可能的整数差),因此可以预先计算指数并将其储存为查找数组。因此,使用对数强度的字节表示形式和预先计算的指数查找数组,不会影响计算效率。

[0106] 质荷比值的分箱

[0107] 如美国授权前公开第2016/0268112A1号所述,将质荷比值转换并组装到低分辨率的箱中,并且相对电荷态间隔进行了一次预先计算,并进行了缓存以提高效率。此外,质谱线的 m/z 值已从其正常的线性标度以道尔顿转换为更自然的无因次对数表示形式。这种转换极大地简化了对例如属于同一蛋白质但潜在地代表不同电荷态的任何峰的 m/z 值的计算。转换不影响精度。当使用转换后的变量进行计算时,可以利用缓存的相对 m/z 值来提高计算效率。

[0108] 基于简单计数的电荷态评分和统计选择标准。

[0109] 如美国授权前公开2016/0268112A1号所述,所有相关质谱的全部内容都被编码为一个布尔值数组。电荷态对质心的评分在适合查询的电荷态的已转换 m/z 位置处减少为简单计数,即布尔变量的是或否(真或假)。这种方法绕开了涉及双精度变量的计算量大的运算。一旦为一系列潜在的荷电状态编制了得分,就可以通过简单的统计程序轻松选择最佳值。使用统计标准比使用任意得分截止值或仅选择最高得分的电荷态更为严格和可靠。

[0110] 电荷态分配的迭代优化

[0111] 前述美国授权前公开第2016/0268112A1号的教导使用了一种迭代过程,所述过程由电荷分配的完全自洽性定义。所述方法的最终关键特征是使用适当的最优性条件,将电荷分配引向解决方案。简单地将最佳条件定义为质谱中所有质心的电荷的最一致分配。所述条件的基础在于,分配给每个质心的电荷态应与分配给质谱中其它质心的电荷态一致。所述出版物中描述的算法实现了一个迭代过程,以按照上述最优性条件指导生成电荷态分配。这一过程符合优化过程的公认规范。即,首先定义适当的最优性条件,然后设计一种算法来满足这一条件,最后,可以通过其满足最优性条件的程度来判断算法的有效性。

[0112] 质谱去卷积结果的实例

[0113] 图13A示出了由细胞色素c、溶菌酶、肌红蛋白、胰蛋白酶抑制剂和碳酸酐酶组成的五种蛋白质混合物的去卷积结果,其中去卷积是根据美国授权前公开第2016/0268112A1号的教导进行的。图形用户界面显示器的顶部显示面板1203示出了从质谱仪获取的以质心表示的数据。定位在中央的主显示面板1201将每个峰图示为相应的符号。顶部面板1203和中央面板1201的水平放置的质荷比(m/z)标度1207示出在中央面板下方。显示器左侧的面板1205示出了计算出的蛋白质分子的一个或多个分子量,以道尔顿为单位。侧面板1205的分子量(MW)标度在显示器上竖直定向,其垂直于与检测到的离子有关的水平定向的 m/z 标度1207。在这个实例中,中央面板1201中的每条水平线指示对蛋白质的检测,其中虚线轮廓线对应于算法分配的离子电荷态,所述电荷态被示出为先前讨论的转化计算的直接结果。在图13B示出了属于同一数据集的显示器,其中分子量(MW)标度相对于图13A所示的视图大大

扩展。图13B的放大视图示出了单个蛋白质电荷态(左面板1205的最下部)以及潜在的加合物或杂质峰(显示器中存在两个)的良好分辨的同位素。这三个分子中最强烈的是胰蛋白酶抑制剂蛋白。

[0114] 图12是根据本教导的用于使用自动碰撞能量测定对蛋白质或多肽进行串联质谱分析的方法(方法800)的流程图。在方法800(图12)的步骤802中,将包括多种蛋白质和/或多肽的样品或样品部分输入质谱仪中并进行电离。优选地,通过电离技术或电离源进行电离,所述电离技术或电离源产生能够根据离子的质荷比(m/z)的测量值来计算各种蛋白质或多肽化合物的分子量的类型的离子种类。具体地说,优选地,电离技术或电离源从每种分析物化合物产生包括一系列电荷态的离子种类,其中每个这样的离子种类包括分析物化合物的原本完整的分子,但包括一个或多个加合物。电喷雾和热喷雾电离是合适的电离技术的两个例子,因为通过这些特定电离技术从蛋白质和/或多肽产生的主要离子种类是具有不同质子化程度的多质子化分子。由电离源产生并从离子源引入质谱仪中的离子可以被称为“第一代离子”。

[0115] 在将第一代离子引入质谱仪之后,在步骤804中对第一代离子进行质量分析以生成质谱,在此将所述质谱称为“MS1”质谱,以表明其与第一代离子相关。质谱是在多个 m/z 值中的每个值下测量的离子电流(强度与检测到的离子数量成正比)的简单列表或表格,通常维持在计算机可读存储器中。然后,在步骤806中,以能够根据在质谱中检测到其存在的离子的 m/z 比计算各种蛋白质或多肽化合物的分子量的方式自动检查MS1光谱。如果需要,执行所述步骤可能需要事先将质谱数据数学分解(去卷积)为单独的已识别的电荷态序列,其中每个电荷态对应于不同的相应蛋白质或多肽化合物。可以根据上述美国授权前公开第2016/0268112A1号中描述的方法执行数学去卷积和对电荷态序列的识别。可替代地,可以通过任何等效算法来执行数学去卷积。例如,于2016年9月9日提交的共同未决的欧洲专利申请第16188157教导了这种替代性数学算法。前述欧洲专利申请的文本作为本文件的附录被包括在内,并且其附图作为图A1被包括随附图纸集中。在某些情况下,算法应该是经过优化的算法,以便可以在包括方法800作为一部分的质谱实验所规定的时间限制内执行所需的去卷积。

[0116] 在方法800(图12)的步骤808中,从在先前步骤中已识别的一个或多个电荷态序列的每一个中选择具有相应 m/z 的至少一个前体离子种类。优选地,如果选择了一个以上的前体离子,则不同的前体离子选自不同的电荷态序列。然后,在步骤810中,计算每个选定的前体离子种类的最佳碰撞能量(CE),其中稍后将每个计算出的最佳碰撞能量在离子碎裂步骤中赋予相应的选定前体离子种类中的离子,并且其中在计算与所述离子种类相关的最佳碰撞能量时,使用从中产生相应的选定离子的分子种类的计算分子量。任选地,可在与所述离子种类相关的最佳碰撞能量的计算中包括每个相应的选定离子种类相应的所识别 z 值。

[0117] 可以根据本文教导的方法在步骤810中计算最佳碰撞能量。例如,如果选择最佳碰撞能量以使在碎裂之后留下残留的剩余前体离子强度百分比 D_p ,则可以使用等式2来计算碰撞能量,其中参数 c 和 k 是根据等式3和等式4确定的,或者由这两个等式形式的等式来计算,但是具有根据特定质谱仪设备的先前校准所确定的不同数值。可替代地,可以选择最佳碰撞能量以便在使用等式5结合表2中列出的参数值进行碎裂之后留下残留的剩余前体离子强度百分比 D_p 。作为另一个替代,可以选择最佳碰撞能量以使选定的前体离子种类碎裂

后存在的产物离子分布与使用等式9结合表3中列出的参数值获得的某个期望的熵参数 D_E 一致。

[0118] 在方法800的步骤812中,通过已知的隔离手段在质谱仪内隔离选定的前体离子种类。例如,如果将MS1离子种类暂时储存在多极离子阱设备中,则可以向阱的电极施加辅助振荡电压(辅助AC电压),以使除特定选定种类之外的所有其它种类均从离子阱中排出,从而仅将选定的种类隔离在阱内。随后,在步骤814中,通过HCD技术使选定并隔离的前体离子种类的离子碎裂以产生碎片离子,其中将先前计算的最佳碰撞能量赋予选定的离子以引发碎裂。在步骤815,获取碎片离子的质谱(即,MS2谱)并将其储存在计算机可读存储器中。

[0119] 如果在执行步骤815之后,存在任何剩余的未被碎裂的选定的前体离子种类,则执行返回到步骤814,然后进行步骤815,在此步骤中将另一个选定的前体离子种类隔离并碎裂。否则,执行前进到步骤818或步骤820。在步骤818中,将从MS1光谱获得的选定前体离子的 m/z 或分子量与来自MS2光谱的信息相结合,以识别或确定有关被分析的样品或样品部分中的多肽或蛋白质的结构信息。任选步骤818不必在步骤816之后立即执行,可以延迟到方法800终止之前执行,或者实际上可以在以后的时间执行,条件是储存来自相关MS1和MS2质谱的信息以备后用和分析。最后,如果在步骤820处确定存在另外的样品或样品部分待分析,则执行返回到步骤802,在此步骤中分析下一个样品或样品部分。可以通过对初始均质的样品进行分级(如通过毛细管电泳、液相色谱法等)来生成各个样品部分,使得在步骤802的每次执行中输入到质谱仪的材料在化学上比原始的未分级样品要简单。分级的某些测量的方面(如观察到的保留时间)可以与相应的MS1和MS2信息结合,以便在后续执行步骤818期间鉴定一种或多种分析物。

[0120] 结论:模型检验

[0121] 通过将相关参数 D_p 和 D_E 以及上述美国授权前公开第2016/0268112A1号的质谱去卷积算法纳入现有数据采集控制软件,测试了前体衰变模型和熵模型。通过使用前体离子衰变和产物离子熵模型对液相色谱部分的MS/MS分析,以及通过各种优化的固定归一化碰撞能量,分析了大肠杆菌细胞裂解物的蛋白质部分。在这些实验中,观察到,相对于优化的固定常规归一化碰撞能量方案,使用任一模型来计算最佳碰撞能量均会改善解离程度的控制。使用本教导的方法,这种改进的碎裂已在各种数据集中导致蛋白质鉴定的改进。

[0122] 附录:用于鉴定分子种类的单同位素质量的方法

技术领域

[0123]

[0124] 本发明属于用于鉴定至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的方法。所述方法使用质谱仪测量样品的质谱。通过所述方法,可以鉴定分子种类的单同位素质量或与同位素分布的同位素质量相关的参数,所述分子种类包含在质谱仪研究的样品中,或者至少通过电离过程从质谱仪研究的样品中产生。优选地,电离过程产生由质谱仪分析的离子。

背景技术

[0125]

[0126] 鉴定一个分子种类(大多数情况下为各个分子种类)的至少单同位素质量或与同位素分布的同位素质量相关的参数的方法通常是可用的。优选地,这些方法通常用于鉴定

质量通常在200u和5,000,000u之间,优选在500u和100,000u之间并且特别优选在5,000u和50,000u之间的大分子(如肽、蛋白质、核酸、脂质和碳水化合物)的单同位素质量。

[0127] 这些方法用于研究样品。这些样品可能含有分子种类,所述分子种类可以通过其单同位素质量或与其同位素分布的同位素质量相关的参数来进行鉴定。

[0128] 分子种类被定义为具有相同分子式的一类分子(例如,水具有分子式H₂O,而甲烷具有分子式CH₄)。

[0129] 或者可以通过至少通过电离过程从样品中产生的离子更好地理解所研究的样品。离子可以优选地通过电喷雾电离(ESI)、基质辅助激光解吸电离(MALDI)、等离子体电离、电子电离(EI)、化学电离(CI)和大气压化学电离(APCI)产生。所产生的离子是带电粒子,大部分具有分子几何形状和相应的分子式。在本专利申请的上下文中,术语“至少通过电离过程从样品产生的分子种类”应理解为是指至少通过电离过程从样品产生的离子的分子式。因此,可以通过在离子的电荷减少到零之后寻找离子的分子式并根据如下所述的电离过程相应地更改分子式,根据至少通过电离过程从样品产生的离子推导至少通过电离过程从样品产生的分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0130] 在分子种类中,所有分子根据分子式具有相同的原子组成。但是分子中的大多数原子可以不同的同位素形式出现。例如,有机化学的基本元素——碳原子——以两个稳定的同位素形式存在,即自然发生概率为98.9%的¹²C同位素和自然发生概率为1.1%的¹³C同位素(原子核中多一个中子)。由于同位素存在的这种可能性,由较高数量的原子组成的质量较高的复杂分子尤其具有许多同位素异构体,其中分子的原子以不同的同位素形式存在。在专利申请的整个上下文中,分子种类的这些同位素异构体被称为“分子种类的同位素”。这些同位素具有不同的质量,导致分子种类的同位素的质量分布,这在本专利申请的内容中被称为分子种类的同位素分布(短期:ID)。因此,每个分子种类可以具有不同的质量,但是为了更好地理解和鉴定每个分子的分子种类,将其分配为单同位素质量。当分子的每个原子以最低质量的同位素形式存在时,所述单同位素质量是分子的质量。例如,甲烷分子的分子式为CH₄,氢的同位素¹H在原子核中具有一个质子而同位素²H(氘)在原子核中具有另外的中子。所以碳的最低质量的同位素是¹²C,氢的最低质量的同位素是¹H。因此,甲烷的单同位素质量为16u。但质量为17u、18u、19u、20u和21u的其它甲烷同位素也有可能存在。所有这些其它同位素属于甲烷同位素分布,并且可以在质谱仪的质谱中可见。

[0131] 通过用质谱仪测量所研究样品的质谱来鉴定至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。通常可以使用本领域技术人员已知的每一种质量分析器来测量所述样品的质谱。具体地说,优选使用高分辨率的质谱仪(例如具有Orbitrap作为质量分析器的质谱仪)、FT质谱仪、ICR质谱仪或MR-TOF质谱仪。可以应用本发明方法的其它质谱仪尤其是TOF质谱仪和具有HR四极杆质量分析器的质谱仪。但是,如果用具有低分辨率的质谱仪测量质谱,则很难用已知的鉴定方法鉴定分子种类的单同位素质量或与同位素分布的同位素质量相关的参数,具体地说,因为无法区分质量差为1u的同位素的相邻峰。

[0132] 一方面,已经存在于样品中的分子被释放并且仅通过电离过程(例如通过接收和/或发射电子)来带电。由于在质谱仪的质谱中检测到分子种类的离子,因此本发明的方法能够为样品中包含的这些分子种类分配其单同位素质量。

[0133] 另一方面,电离过程可以通过将样品中包含的分子碎裂成较小的带电粒子或将原子或分子添加到样品中包含的分子中而改变样品中包含的分子,从而导致较大的分子因这一过程而带电。同样,通过电离过程,样品的基质可以分裂成带电的分子。因此,所有这些离子均通过上述电离过程从样品中产生。因此,对于这些离子,必须通过用于鉴定至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的方法来研究源自样品的分子的相应种类。

[0134] 迄今为止,已经公开了许多鉴定质谱中的同位素峰的单同位素质量的方法,包含帕特森(Patterson)函数,傅立叶变换或其组合(M.W.Senko等人,《美国质谱学会杂志(J.Am.Soc.Mass Spectrom.》1995,6,52;D.M.Horn等人,《美国质谱学会杂志》2000,11,320;L.Chen和Y.L.Yap,《美国质谱学会杂志》2008,19,46)、 m/z 准确性得分(Z.Zhang和A.G.Marshall,《美国质谱学会杂志》1998,9,225)、实验观察到的峰型与理论模型的拟合(P.Kaur和P.B.O'Connor,《美国质谱学会杂志》2006,17,459;X.Liu等人,《分子与细胞蛋白组学》2010,9,2772)以及基于熵的去卷积算法(B.B.Reinhold和V.N.Reinhold,《美国质谱学会杂志》1992,3,207)。这些方法通常针对特定的应用(如肽和/或完整蛋白),并且报告的执行时间在2.2-GHz CPU上以秒为时间范围(Liu等人,2010),所述时间不足以在线检测并随后选择种类进行进一步的MS分析,如MS蛋白质组学的标准方法。P.Yip等人的未公开方法已经优化为使用潜在相关峰的大量相关性分析完整蛋白,所述潜在相关峰先前已经从原始数据转换为具有二进制强度信息的对数 m/z 轴。然而,对于傅里叶变换质谱仪而言,速度还不够快。显然,对于获取速度(即每单位时间可通过实验分析的数据量)至关重要的应用领域,需要不仅适用于更广泛的应用(包含肽、有机小分子和完整蛋白质)而且还适用于在数据采集后立即进行快速在线分析(不延迟后续扫描的采集)的整体方法。

发明内容

[0135]

[0136] 上述目的通过用于鉴定至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的新方法得到解决,所述分子包含在样品中和/或至少通过根据权利要求1所述的电离过程从样品中产生。

[0137] 本发明的方法包括以下步骤:

[0138] (i) 用质谱仪测量样品的质谱

[0139] (ii) 将所测得的样品质谱的 m/z 值的至少一个范围划分成多个部分

[0140] (iii) 将所测得的 m/z 值的至少一个范围中的至少一些部分分配给几个提供的处理器中的一个处理器

[0141] (iv) 在所测得的 m/z 值的至少一个范围的至少一个部分中,根据所测得的质谱推导包含在样品中和/或源自样品的至少一个分子种类中的每一个分子的离子的同位素分布,所述离子具有特定电荷 z ,以及

[0142] (v) 根据包含在样品中和/或源自样品的至少一个分子种类中的每一个分子的离子的至少一种推导出的同位素分布,推导所述分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0143] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施

例中,其中在所测得的 m/z 值的至少一个范围的每个部分中,检测一个分子种类的离子的至少一种同位素分布,所述离子具有特定电荷 z 。

[0144] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施例中,在所测得的 m/z 值的至少一个范围的至少一个部分中,推导除所述至少一个分子种类之外的至少另一个分子种类的离子的同位素分布,所述离子具有特定电荷 z 。

[0145] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施例中,根据包含在样品中和/或至少通过电离过程从样品中产生的所述分子种类中的一些分子的离子(所述离子具有特定电荷 z)的两种或更多种推导出的同位素分布,推导单同位素质量或与同位素分布的同位素质量相关的参数。

[0146] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施例中,根据包含在样品中和/或至少通过电离过程从样品中产生的所述分子种类中的一些分子的离子(所述离子具有特定电荷 z)的两种或更多种推导出的同位素分布,推导单同位素质量或与同位素分布的同位素质量相关的参数,所述离子的两种或更多种推导出的同位素分布是从所测得的 m/z 值的至少一个范围的不同部分中推导出的。

[0147] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施例中,通过评估从所测得的 m/z 值的至少一个范围的不同部分中推导出的具有特定电荷 z 的离子的同位素分布,在所测得的 m/z 值的至少一个范围的至少一个部分中,根据包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的离子(所述离子具有特定电荷 z)的至少一种推导出的同位素分布,推导所述分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0148] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的优选实施例中,通过评估从分配给处理器的所有部分中推导出的具有特定电荷 z 的离子的同位素分布,在所测得的 m/z 值的至少一个范围的至少一个部分中,根据包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的离子(所述离子具有特定电荷 z)的至少一种推导出的同位素分布,推导所述分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0149] 在用于鉴定包含在样品中的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施例中,通过将四个子电荷得分 $C_{SP_PX}(z)$ 、 $C_{SAS_PX}(z)$ 、 $C_{SAC_PX}(z)$ 和 $C_{SIS_PX}(z)$ 中的至少三个相乘推导出所测质谱峰PX的电荷得分 $C_{SPX}(z)$,从而根据所测得的质谱推导包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类中的每一个分子的离子的至少一个同位素分布,所述所述离子具有特定电荷 z 。

[0150] 在用于鉴定包含在样品中的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的优选实施例中,通过将四个子电荷得分 C_{SP_PX}

(z)、 $C_{SAS_PX}(z)$ 、 $C_{SAC_PX}(z)$ 和 $C_{SIS_PX}(z)$ 中的至少三个相乘推导出所测质谱峰PX的电荷得分 $C_{SPX}(z)$ 。

[0151] 在用于鉴定包含在样品中的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的一个实施例中,通过针对电荷1与最大电荷态 Z_{max} 之间的每个电荷态z推导出所测质谱峰PX的电荷得分 $C_{SPX}(z)$,从而根据所测得的质谱推导包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类中的每一个分子的离子的至少一个同位素分布,所述所述离子具有特定电荷z。

[0152] 上述目的通过用于鉴定至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的新方法得到进一步解决,所述分子包含在样品中和/或至少通过根据权利要求11所述的电离过程从样品中产生。

[0153] 本发明的方法包括以下步骤:

[0154] (i) 用质谱仪测量样品的质谱

[0155] (ii) 通过将四个子电荷得分 $C_{SP_PX}(z)$ 、 $C_{SAS_PX}(z)$ 、 $C_{SAC_PX}(z)$ 和 $C_{SIS_PX}(z)$ 中的至少三个相乘推导出所测质谱峰PX的电荷得分 $C_{SPX}(z)$,从而根据所测得的质谱推导包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类中的每一个分子的离子的至少一个同位素分布,所述所述离子具有特定电荷z,以及

[0156] (iii) 根据包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类中的每一个分子的离子(所述离子具有特定电荷z)的至少一种推导出的同位素分布,推导所述分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0157] 在用于鉴定包含在样品中和/或至少通过电离过程从样品中产生的至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的本发明方法的优选实施例中,其中通过将四个子电荷得分 $C_{SP_PX}(z)$ 、 $C_{SAS_PX}(z)$ 、 $C_{SAC_PX}(z)$ 和 $C_{SIS_PX}(z)$ 中的至少三个相乘推导出所测质谱峰的电荷得分 $C_{SPX}(z)$ 。

[0158] 上述目的通过用于鉴定至少一个分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的新方法得到进一步解决,所述分子包含在样品中和/或至少通过根据权利要求13所述的电离过程从样品中产生。

[0159] 本发明的方法包括以下步骤:

[0160] (i) 用质谱仪测量样品的质谱

[0161] (ii) 根据所测得的质谱推导包含在样品中和/或源自样品的至少一个分子种类中的每一个分子的离子的至少两种同位素分布,所述离子具有特定电荷z,以及

[0162] (iii) 根据包含在样品中和/或源自样品的至少一个分子种类中的每一个分子的离子的所述至少两种推导出的同位素分布,推导所述分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0163] 本发明的方法利用来自分子种类的相关同位素分布的信息,这大大提高了鉴定所述分子种类的单同位素质量或与同位素分布的同位素质量相关的参数的准确性。这对于完整蛋白质尤其有利,完整蛋白质往往由于电离而形成具有较高电荷态的分子种类的离子的广泛的同位素分布。通过确定最大可分辨同位素分布,可以动态处理解析度差或完全无法解析的ID(即,同位素峰没有或仅部分解析的ID)。由于灵活的m/z窗口,防止了单个ID的分离。已经针对广泛的应用,包含肽、有机小分子(包含同位素峰型不常见的有机小分子)和完

整蛋白质,优化了已实施的电荷得分。通常,检测和注释不限于肽/蛋白质的平均模型。与现有技术的方法相反,本发明的方法允许为每个分子种类分配多种同位素分布。为了提高新方法性能,避免了如傅立叶变换等耗时程序,并在尽可能采用多处理以及速度优化的过程。本发明的方法使用峰的原始强度来更好地区分相邻的和重叠的ID,这对于肽数据以及肽和蛋白质的混合物特别重要。新方法只需不到20毫秒即可处理信噪比阈值为10(这意味着在第二种算法中,仅聚焦高于这个阈值的峰以进行电荷态分析)的复杂蛋白质样品的质谱(包含单同位素质量的测定)。任选的动态S/N阈值允许在含有多个相邻/重叠ID的峰密集区域中增加阈值,以限制运行时间。

[0164] 本发明代表用于确定峰的单同位素质量或与质谱中至少一个分子种类的同位素分布的同位素质量相关的参数的整体方法,所述方法适用于广泛的应用/化学种类,但重点是完整蛋白质和带有高电荷态的多重带电种类。对所述方法的速度优化是必不可少的要素,这确保其适用于在约20-30毫秒内在线检测复杂蛋白质样品的质谱中所包含的大多数种类。

[0165] 所述方法能够处理未解析的同位素分布,因此即使是复杂蛋白质样品的低分辨率质谱也可以用于本发明方法。

具体实施方式

[0166]

[0167] 本发明的方法用于鉴定一个分子种类(大多数情况下为各个分子种类)的至少单同位素质量。优选地,所述方法通常用于鉴定质量通常在200u和5,000,000u之间,优选在500u和100,000u之间并且特别优选在5,000u和50,000u之间的大分子(如肽、蛋白质、核酸、脂质和碳水化合物)的单同位素质量。

[0168] 本发明的方法用于研究样品。这些样品可能含有分子种类,所述分子种类可以通过其单同位素质量或与其同位素分布的同位素质量相关的参数来进行鉴定。

[0169] 在下文中,仅描述本发明方法的实施例以鉴定分子种类的单同位素质量。然而,所有描述的方法也可以用于鉴定与分子种类的同位素分布的同位素质量相关的参数。具体地说,这一参数表示一个分子种类的同位素分布的同位素平均质量,一个分子种类的同位素分布中出现频率最高的同位素的质量以及一个分子种类的同位素分布的质心质量。

[0170] 分子种类被定义为具有相同分子式的一类分子(例如,水具有分子式 H_2O ,而甲烷具有分子式 CH_4)。

[0171] 或者可以通过至少通过电离过程从样品中产生的离子更好地理解所研究的样品。离子可以优选地通过电喷雾电离(ESI)、基质辅助激光解吸电离(MALDI)、等离子体电离、电子电离(EI)、化学电离(CI)和大气压化学电离(APCI)产生。所产生的离子是带电粒子,大部分具有分子几何形状和相应的分子式。在本专利申请的上下文中,术语“至少通过电离过程从样品产生的分子种类”应理解为是指至少通过电离过程从样品产生的离子的分子式。

[0172] 因此,可以通过在离子的电荷减少到零之后寻找离子的分子式并根据如下所述的电离过程相应地更改分子式,根据至少通过电离过程从样品产生的离子推导至少通过电离过程从样品产生的分子种类的单同位素质量或与同位素分布的同位素质量相关的参数。

[0173] 在分子种类中,所有分子根据分子式具有相同的原子组成。但是分子中的每个原子可以不同的同位素形式出现。因此,有机化学的基本元素——碳原子——以两个稳定的

同位素形式存在,即自然发生概率为98.9%的 ^{12}C 同位素和自然发生概率为1.1%的 ^{13}C 同位素(原子核中多一个中子)。由于同位素存在的这种可能性,由较高数量的原子组成的质量较高的复杂分子尤其具有许多同位素。这些同位素具有不同的质量,导致同位素的质量分布,这在本专利申请的内容中被称为分子种类的同位素分布(短期:ID)。因此,每个分子种类可以具有不同的质量,但是为了更好地理解和鉴定每个分子的分子种类,将其分配为单同位素质量。当分子的每个原子以最低质量同位素形式存在时,所述单同位素质量是分子的质量。例如,甲烷分子的分子式为 CH_4 ,氢的同位素 ^1H 在原子核中具有一个质子而同位素 ^2H (氘)在原子核中具有另外的中子。所以碳的最低质量同位素是 ^{12}C ,氢的最低质量同位素是 ^1H 。因此,甲烷的单同位素质量为16u。但质量为17u、18u、19u、20u和21u的其它甲烷同位素也有可能存在。所有这些其它同位素属于甲烷同位素分布,并且可以在质谱仪的质谱中可见。

[0174] 在本发明方法的第一步中,用质谱计测量样品的质谱。通常可以使用本领域技术人员已知的每一种质量分析器来测量样品的质谱。具体地说,优选使用高分辨率的质谱仪(例如具有Orbitrap作为质量分析器的质谱仪)、FT质谱仪、ICR质谱仪或MR-TOF质谱仪。可以应用本发明方法的其它质谱仪尤其是TOF质谱仪和具有HR四极杆质量分析器的质谱仪。但是本发明的方法还具有以下优点:如果用具有低分辨率的质谱仪测量质谱,使得例如无法区分质量差为1u的同位素的相邻峰时,本发明的方法能够鉴定分子种类的单同位素质量。

[0175] 一方面,已经存在于样品中的分子被释放并且仅通过电离过程(例如通过接收和/或发射电子、质子(H^+)和带电粒子)来带电。由于在质谱仪的质谱中检测到分子种类的离子,因此本发明的方法能够为样品中包含的这些分子种类分配其单同位素质量。

[0176] 另一方面,电离过程可以通过将样品中包含的分子碎裂成较小的带电粒子或将原子或分子添加到样品中包含的分子中而改变样品中包含的分子,从而导致较大的分子因这一过程而带电。同样,通过电离过程,样品的基质可以分裂成带电的分子,或者可以形成分子簇。因此,所有这些离子均通过上述电离过程从样品中产生。因此对于这些离子,可以通过本发明的方法研究源自样品的分子的相应种类,并且所述方法可能能够鉴定其单同位素质量。

[0177] 在本发明方法的下一个可能步骤中,至少将所测得的质谱的质量范围划分成多个部分。例如可以通过作为质谱仪一部分的处理器执行这一步骤,所述处理器可以具有如控制质谱仪等另外的其它功能。划分质量范围的目的是,可以将每个部分分配给由具有多个中央处理器单元(CPU)的多处理器提供的几个处理器中的一个处理器,然后所述处理器可以以单个线程在质量范围内的已分配部分中推导具有特定电荷 z 的分子种类的离子的同位素分布。通常,多处理器具有2个或4个CPU,用于在分配给特定CPU的部分中推导具有特定电荷 z 的分子种类的离子的同位素分布。但是可以使用更多CPU(例如6个、8个或12个)推导同位素分布。如果相应地将更多CPU用于更多部分,则可以平行推导具有特定电荷 z 的分子种类的离子的同位素分布。

[0178] 在通过质谱仪测量样品的质谱后,必须定义通过所述测量检测到的 m/z 值的哪个范围将用于鉴定包含在样品中的和/或至少通过电离过程从样品中产生的分子种类在质谱仪中分离期间的单同位素质量。用户可以定义检测到的 m/z 值的使用范围。可以在开始测量

质谱之前或在质谱显示在图形输出系统(如显示器)上之后定义所述范围。可以基于样品研究的意图和/或基于所得到的质谱来定义所述范围。因此,如果在一系列 m/z 值中未观察到峰,则可以中止对所述一系列 m/z 值的进行进一步评估,并且其不属于划分成多个部分的一系列 M/Z 值。

[0179] 还可以通过控制鉴定方法的控制器来定义使用的一系列检测到的 m/z 值。例如,如果在一系列 m/z 值的所测得的质谱中未观察到峰或强度高于阈值的峰,则可以通过控制器来中止对所述一系列 m/z 值的进行进一步评估,所述控制器限制用于鉴定单同位素质量的一系列 m/z 值。

[0180] 在本发明方法的一个实施例中,将由质谱仪检测到的并因此在所测得的质谱中示出的 m/z 值的整个范围划分成用于推导同位素分布的多个部分。

[0181] 这在图1中示出,该图显示了通过质谱仪测量的质谱。质谱仪检测 m/z 值(离子质量 m 与离子电荷 z 的比率)介于最小值 m/z_{\min} 与最大值 m/z_{\max} 之间的离子。然后,可以将介于最小值 m/z_{\min} 与最大值 m/z_{\max} 之间的 m/z 值的整个范围划分成多个部分,然后将这些部分分配给离散处理器(CPU),以推导包含在样品中和/或至少通过电离过程从样品中产生的分子种类的离子的同位素分布,所述离子具有特定电荷 z 。

[0182] 在本发明方法的另一个实施例中,将由质谱仪检测到的并因此在所测得的质谱中示出的 m/z 值的部分范围划分成用于推导同位素分布的多个部分。在这个实施例中,仅将由质谱仪检测到的质谱的 m/z 值的一个或多个特定范围划分成用于推导同位素分布的多个部分。

[0183] 这也在图1中示出,该图显示了通过质谱仪测量的质谱。质谱仪检测 m/z 值(离子质量 m 与离子电荷 z 的比率)介于最小值 m/z_{\min} 与最大值 m/z_{\max} 之间的离子。但是,也可以将介于最小值 m/z_{\min} 与最大值 m/z_{\max} 之间的 m/z 值的部分范围划分成多个部分,然后将这些部分分配给离散处理器(CPU),以推导包含在样品中和/或至少通过电离过程从样品中产生的分子种类的离子的同位素分布,所述离子具有特定电荷 z 。也有可能将所测得的 m/z 值的特定范围划分成多个部分,然后将这些部分分配给离散处理器(CPU)以推导同位素分布。在图1中,示出了 m/z 值的范围A和范围B。在一个实施例中,仅将所测得的 m/z 值的范围A划分成多个部分,然后将这些部分分配给离散处理器(CPU)以推导同位素分布。在另一个实施例中,仅将所测得的 m/z 值的范围B划分成多个部分,然后将这些部分分配给离散处理器(CPU)以推导同位素分布。在另外的实施例中,将两个范围(所测得的 m/z 值的范围A和所测得的 m/z 值的范围B)划分成多个部分,然后将这些部分分配给离散处理器(CPU)以推导同位素分布。根据图1,在这个实施例中,仅将所述范围(范围A和B)划分成多个部分,并用于推导同位素分布,在所述范围中已经测量到相对丰度大于5%的峰。

[0184] 首先,将所测得的 m/z 值的至少一个范围划分成特定窗口宽度 $\Delta m/z_{\text{start}}$ 的一部分。通常,窗口宽度 $\Delta m/z_{\text{start}}$ 略大于 1Th (汤普森; $1\text{Th}=1\text{u}/e$; u :原子质量单位; e :基本电荷; $1\text{u}=1.660539 \times 10^{-27}\text{Kg}$; $1e=1,602176 \times 10^{-19}\text{C}$)。在优选的实施例中,窗口宽度 $\Delta m/z_{\text{start}}$ 介于 1.000Th 与 1.100Th 之间,在更优选的实施例中,窗口宽度 $\Delta m/z_{\text{start}}$ 介于 1.005Th 与 1.050Th 之间,并且在特别优选的实施例中,窗口宽度 $\Delta m/z_{\text{start}}$ 介于 1.010Th 与 1.020Th 之间。将窗口宽度 $\Delta m/z_{\text{start}}$ 选择为在 1Th 的范围内,因为在离子的最低电荷态下,电荷为 $z=1$,因此相邻同位素的 m/z 值之间的最小距离为 1Th 。这周到地考虑了一些技术公差,必须将窗口宽度 Δ

m/z_{start} 选择为略大于 $1Th$ 。技术公差源自例如由于化学元素、峰宽、 m/z 峰的质心化而产生的偏差。

[0185] 所有具有起始窗口宽度 $\Delta m/z_{start}$ 的这些部分如果具有显著的峰,则都将被研究。仅将具有这种峰的部分分配给处理器,所述处理器随后将在所测得的 m/z 值的至少一个范围的部分内,根据所测得的质谱推导同位素分布。大多数情况下,在所测得的 m/z 值的至少一个应划分的范围的一个边界(最高 m/z 值或最低 m/z)处开始研究具有起始窗口宽度 $\Delta m/z_{start}$ 的部分是否具有显著的峰。如果所述部分中强度最高的峰的信噪比 S/N 高于阈值 T ,则所述部分具有显著的峰。

[0186] 在研究了具有起始窗口宽度 $\Delta m/z_{start}$ 的部分是否具有显著的峰之后,将研究先前未研究的具有起始窗口宽度 $\Delta m/z_{start}$ 的相邻部分是否具有显著的峰。如果两个部分都包括具有特定电荷的一个分子种类的离子的相同同位素分布的同位素或连续同位素分布或重叠同位素分布的同位素,则将相邻的部分串接起来构成具有较大窗口宽度 $\Delta m/z$ 的部分。因此,如果两个相邻的部分之一没有显著的峰,则不会将其串接在一起。

[0187] 如果在所测得的 m/z 值的至少一个应划分的范围的一个边界(最高 m/z 值或最低 m/z)处开始研究具有起始窗口宽度 $\Delta m/z_{start}$ 的部分是否具有显著的峰,则所述研究在先前未研究的相邻部分中结束,所述相邻部分包括所测得的 m/z 值的至少一个应划分的范围的第二边界。如果仅将所测得的 m/z 值的一个范围划分成多个部分,则完成对这些部分的整个研究。如果将所测得的 m/z 值的不止一个范围划分成多个部分,则将尚未划分成多个部分的所测得的 m/z 值的下一个应划分范围以相同的方式或使用不同的参数划分成多个部分。在将所测得的 m/z 值的所有被定义为要划分的范围都划分成多个部分之后,完成所述划分。

[0188] 可以将具有起始窗口宽度 $\Delta m/z_{start}$ 的部分的串接限制于这些部分的特定数量。由于单个处理器的这种太长的操作时间,以致无法推导出分配的串接部分中的同位素分布,这将增加执行本发明方法的整个时间。在本发明方法的优选实施例中,应将不超过20个具有起始窗口宽度 $\Delta m/z_{start}$ 的部分串接起来,在本发明方法的更优选的实施例中,应将不超过12个具有起始窗口宽度 $\Delta m/z_{start}$ 的部分串接起来,并且在本发明方法的特别优选的实施例中,应将不超过8个具有起始窗口宽度 $\Delta m/z_{start}$ 的部分串接起来。

[0189] 在本发明方法的一个实施例中,定义部分是否具有显著的峰的阈值 T 对于所有研究的部分都是相同的。通常,使用的阈值 T 在2.0到5.0的范围内,优选在2.5到4.0的范围内,并且特别优选在2.8到3.5的范围内。

[0190] 在另一个实施例中,阈值 T 是动态调整的。在一个优选的实施例中,其根据部分的峰密度而改变。如果部分中显著峰的数量 N 很高,则阈值 T 会增加,以限制峰的数量 N ,处理器根据所述峰推导出同位素分布。因此,在每个部分中限制信噪比 S/N 大于阈值 T 的峰的数量 N 。可以将这样的部分与具有起始窗口宽度 $\Delta m/z_{start}$ 的部分进行串接。部分中的显著峰的数量 N 受极值 N_{max} 限制。这可以由用户、控制器或控制器的生产者通过硬件或软件进行设置。通常, N_{max} 在100到500的范围内,优选在180到400的范围内,并且特别优选在230到300的范围内。开始时设置初始阈值 T_i 。通常,将初始阈值 T_i 设置在2.0到5.0的范围内,优选在2.5到4.0的范围内,并且特别优选在2.8到3.5的范围内。如果部分中信噪比 S/N 高于阈值 T 的显著峰的数量 N 大于极值 N_{max} ,则阈值 T 增加一个系数,然后再次研究所述部分中信噪比 S/N 高于阈值 T 的显著峰的数量 N 。重复增加阈值,直到信噪比 S/N 高于阈值 T 的峰的数量低于极值 N_{max} 。

通常,阈值T以1.10到2.50之间的系数增加。优选地,阈值T以1.25到1.80之间的系数增加。特别优选地,阈值T以1.35到1.6之间的系数增加。阈值T的增加受到阈值的最大值 T_{\max} 的限制。通过此限制,将避免忽略样品的显著峰。阈值的最大值 T_{\max} 可以由用户、控制器或控制器的生产者通过硬件或软件进行设置。通常,将阈值的最大值 T_{\max} 设置在6到40之间。优选地,将阈值的最大值 T_{\max} 设置在10到30之间。特别优选地,将阈值的最大值 T_{\max} 设置在12到20之间。

[0191] 如果对于多个部分(可能是具有起始窗口宽度 $\Delta m/z_{\text{start}}$ 的部分或由具有起始窗口宽度 $\Delta m/z_{\text{start}}$ 的部分串接而成的具有较大窗口宽度 $\Delta m/z$ 的部分)进行逐个研究,则这些部分的阈值T尚未增加并且部分的阈值高于初始阈值 T_i ,然后随后的相邻部分的阈值T将降低,优选地连续降低至初始阈值 T_i 。阈值T的降低可以通过减去特定值或通过将阈值T减小一个系数来完成。通常,减去的特定值在0.10到0.70之间,优选在0.15到0.40之间,特别优选在0.20到0.30之间。降低阈值T的系数通常在0.85到0.99之间,优选在0.92到0.97之间,特别优选在0.95到0.96之间。还可以使用两种方法来同时减小阈值T,并使用随后的相邻部分的阈值T的较高或较低减小值。不应将阈值降低到初始阈值 T_i 以下。如果发生这种情况,应使用初始阈值 T_i 研究随后的相邻部分。

[0192] 如果已经用高于初始阈值 T_i 的阈值T研究了具有起始窗口宽度 $\Delta m/z_{\text{start}}$ 的部分,并且所述部分没有显著的峰,则在本发明方法的一个实施例中,用初始阈值 T_i 再执行一次研究。然后,如果观察到所述部分的显著峰,则将所述部分标记为信噪比S/N低的部分。

[0193] 在本发明方法的另一个可能步骤中,将所测得的m/z值的至少一个范围的部分中的至少一些部分分配给处理器。处理器是由具有几个中央处理器单元(CPU)的多处理器提供的几个处理器中的一个处理器。处理器可以以单个线程在质量范围内的已分配部分中推导具有特定电荷z的分子种类的离子的同位素分布。通常,多处理器具有2个或4个CPU,用于在分配给特定CPU的部分中推导具有特定电荷z的分子种类的离子的同位素分布。但是可以使用更多CPU(例如6个、8个或12个)推导同位素分布。如果相应地将更多CPU用于更多部分,则可以平行推导具有特定电荷z的分子种类的离子的同位素分布。多处理器的处理器可以物理上定位在一个位置。然后,多处理器可以成为质谱仪的一部分。所述多处理器还可以用于质谱仪的其它功能,例如本领域技术人员已知的控制质谱仪的功能。物理上定位在一个位置处的多处理器可以与质谱仪分开,并且例如仅接收质谱仪所测得的质谱的文件。而且,各个多处理器可以定位在不同的位置,并且可以与质谱仪连接,例如与质谱仪的控制单元连接。

[0194] 例如可以通过作为质谱仪一部分的处理器执行将所测得的m/z值的至少一个范围的部分中的至少一些部分分配给处理器的这一步骤,所述处理器可以具有如控制质谱仪等另外的其它功能。

[0195] 在本发明方法的优选实施例中,仅将具有显著峰的部分分配给处理器。一方面,这些部分可以具有起始窗口宽度 $\Delta m/z_{\text{start}}$ 。另一方面,这些部分可以具有较大的窗口宽度 $\Delta m/z$,因为这些部分是由串接的相邻部分构建而成的。

[0196] 在本发明方法的另一个优选实施例中,仅将具有显著峰的部分和被标记为信噪比S/N低的部分的部分分配给处理器。

[0197] 在本发明的优选实施例中,用于在所测得的m/z值的至少一个范围内的已分配部

分中,根据所测得的质谱推导具有特定电荷 z 的分子种类的离子的同位素分布的多处理器中的每个处理器 P_i 均被分配了峰值计数 C_i 和列表,所述列表中储存了与已分配部分有关的信息。通过将所有已分配部分的显著峰的数量 N 相加来计算峰值计数 C_i ,即分配给处理器 P_i 的每个部分的显著峰的数量 N 。当将所测得的 m/z 值的至少一个范围划分成多个部分时,研究每个部分中显著峰的数量 N ,以评估显著峰的数量 N 是否超出的显著峰的有限数量 N_{max} 。

[0198] 将具有显著峰的部分或具有显著峰的部分以及被标记为低信噪比 S/N 低的部分的部分依次分配给处理器 P_i 。待分配给处理器的下一个部分始终被分配给所述处理器,直到所述时刻为止,已分配部分的显著峰的总数最少。这意味着待分配给处理器的下一个部分始终被分配给峰值计数 C_i 最低的处理器 P_i 。所述已分配部分的显著峰的数量被添加到峰值计数 C_i 。因此,具有显著峰的下一个部分始终被分配给分配有最低数量的显著峰的处理器。通过这种分配,确保了已分配的部分中显著峰的数量甚至分布在整个处理器上。这确保了每个处理器几乎同时根据分配给处理器的部分推导同位素分布。通过这种分配,实现了通过几个提供的处理器快速推导出同位素分布。

[0199] 可以连续或并行完成将样品质谱的所测得的 m/z 值的至少一个范围划分成多个部分的步骤和将所测得的 m/z 值的至少一个范围的部分中的至少一些部分分配给几个提供的处理器中的一个处理器的步骤。如果这些步骤是并行执行的,则在将样品质谱的所测得的 m/z 值的至少一个范围划分成多个部分的步骤中定义每个部分,并在定义后将其立即分配给处理器,所述处理器将推导所述部分的同位素分布。

[0200] 在本发明方法的下一步骤中,在 m/z 值的至少一个范围的部分中的至少一个部分中,根据所测得的质谱推导一个分子种类的离子的同位素分布,所述离子具有特定电荷 z 。所述推导出的

[0201] 具有特定电荷 z 的离子的同位素分布是针对包含在样品中和/或至少通过电离过程从样品中产生的一个分子种类的离子进行推导的。优选地,可以针对包含在样品中和/或至少通过电离过程从样品中产生的一个分子种类的几种离子,推导具有比电荷 z 的离子的同位素分布。

[0202] 在本发明方法的一个实施例中,在所测得的 m/z 值的至少一个范围的部分中的每一个部分中,检测具有特定电荷 z 的一个分子种类的离子的至少一种同位素分布。

[0203] 并非所有分子种类都可以通过本发明的方法推导出单同位素质量,但其具有特定电荷 z 的离子的同位素分布已被推导出。

[0204] 根据本发明方法的优选实施例,以下描述了在被分配给一个处理器的所测得的 m/z 值的至少一个范围的一个部分中,如何根据所测得的质谱推导一个分子种类的离子的同位素分布,所述离子具有特定电荷 z 。优选地,仅使用先前已被鉴定为如上所述的显著峰的峰。

[0205] 首先,定义在所测得的 m/z 值的所研究部分中具有最高强度的峰。然后,必须定义可以被分配给所述最高强度峰的最大电荷态 z_{max} 。因此,必须鉴定与所述最高强度峰相邻的最近峰。所述最近峰的强度应不低于与所述最高强度峰相比的相对强度值(通常为最高强度峰的强度的2%到6%,优选地3%到5%,特别优选地4%)。同样优选地,这些峰的距离不应大于起始窗口宽度 $\Delta m/z_{start}$ 。根据最高强度峰到与所述最高强度峰相邻的最近峰之间的距离 d ,可以根据平均分布考虑平均同位素质量差距离 Δm_{ave} 来假设可能的最大电荷态 z_{max}

(描述于以下文献中:例如Senko等人,《美国质谱学杂志(J.J.Am.Mass Spectrom.)》1995, 6, 229-233和Valkenborg等人,《美国质谱学杂志》2008, 19, 703-712)

$$[0206] \quad z_{\max} = \frac{\Delta m_{ave}}{d}$$

[0207] 平均同位素质量差距离 Δm_{ave} 的值通常在1.0020u到1.0030的范围内,优选介于1.0023到1.0025u之间。特别优选地,将值1.00235用作平均同位素质量差距离 Δm_{ave} 。

[0208] 优选地,如此评估的最大电荷态 z_{\max} 可以进一步增加大于1的系数。因此,应确保研究至少一个较高的电荷态。通常,与所评估的最大电荷态相乘的系数在1.10到1.30的范围内,优选在1.125到1.20的范围内。优选地,如此获得的结果向上舍入到下一个自然数,即正整数。

[0209] 优选地,可以最大电荷态 z_{\max} 限制为最大值。这可以取决于通过本发明的方法研究的样品的类型。因此,如果研究完整蛋白质,则优选地将最大电荷态 z_{\max} 限制为50到60之间的值,如果研究肽,则优选地将最大电荷态 z_{\max} 限制为低于20的值。合理选择最大电荷态 z_{\max} 的极值避免了对不现实的电荷态的研究,并因此减少了推导同位素分布的时间。最大电荷态 z_{\max} 的极值可以由用户、控制器或控制器的生产者通过硬件或软件进行设置。优选地,如果最大电荷态 z_{\max} 的极值由控制器或控制器的生产者通过硬件或软件进行设置,则其是根据用户的信息(应当研究哪种样本)来设置的。

[0210] 在针对电荷1与最大电荷态 z_{\max} 之间的每个电荷态 z 测量的 m/z 值的所研究部分中,定义了具有最高强度 P_1 的所研究峰的最大电荷态 z_{\max} 的值,然后根据所测 m/z 值的所研究部分中的质谱评估得分值,即电荷得分 $c_{SP_1}(z)$ 。通常,所测得的峰 PX ($X=1, \dots, N$) 的电荷得分 $c_{SPX}(z)$ 反映了所测得的峰 PX 属于具有电荷 z 的同位素分布的可能性。

[0211] 在本发明方法的优选实施例中,在以下模式下,所测得的峰 PX 的电荷得分 $c_{SPX}(z)$ 被假定为具有最高强度的同位素分布:

[0212] 首先基于平均模型,定义了对于具有较小 m/z 值的峰 PX 可以预期有多少同位素分布的峰,即 $N_{left_PX}(z)$, 以及对于具有较高 m/z 值的峰 PX 可以预期有多少同位素分布的峰,即 $N_{right_PX}(z)$ 。优选地,仅考虑强度不小于所研究同位素分布的最高峰 PX 的强度百分比(即截止强度)的同位素分布的峰。通常,所述截止强度在最高峰 PX 的强度的0.5%到6%的范围内,优选在最高峰 PX 的强度的0.8%到4%的范围内。特殊地,截止强度是最高峰 PX 强度的1%。

[0213] 例如,具有较小 m/z 值的峰的数量 $N_{left_PX}(z)$ 和具有较大 m/z 值的峰的数量 $N_{right_PX}(z)$ 可以通过以下公式进行计算:

$$[0214] \quad V_{left_PX}(z) = A * \sqrt{\frac{m}{z}(PX) * z} - B$$

$$[0215] \quad V_{right_PX}(z) = C * \sqrt{\frac{m}{z}(PX) * z} + D$$

[0216] 值 $m/z(PX)$ 是所测得的峰 PX 的 m/z 值。常数 A 、 B 、 C 和 D 由使用的平均模型给出。典型值为: $0.075 < A < 0.080$ 、 $2.35 < B < 2.40$ 、 $0.075 < C < 0.080$ 、 $0.80 < D < 0.85$ 。

[0217] 因此, $N_{left_PX}(z)$ 是小于值 $V_{left_PX}(z)$ 的第一个正整数,否则为0,而 $N_{right_PX}(z)$ 是最接近值 $V_{right_PX}(z)$ 的整数。

[0218] 然后,针对分配给峰PX和电荷z的同位素分布的所有峰,定义了相应的理论m/z值。

[0219] 如果假设同位素分布的平均同位素质量差为 Δm ,则同位素分布的峰具有以下理论m/z值:

$$[0220] \quad m/z(z)_k = m/z(PX) + k * \Delta m/z$$

$$[0221] \quad \text{其中 } k = (-N_{\text{left_PX}}(z), \dots, N_{\text{right_PX}}(z) - 2, N_{\text{right_PX}}(z) - 1, N_{\text{right_PX}}(z))$$

[0222] 因此,例如,如果 $N_{\text{left_PX}}(z) = 1$ (意味着在峰PX的左侧,电荷z的同位素分布中有一个峰)而 $N_{\text{right_PX}}(z) = 6$ (意味着在峰PX的右侧,电荷z的同位素分布中有六个峰),则同位素分布的峰具有以下理论m/z值:

$$[0223] \quad m/z(z)_k = m/z(PX) + k * \Delta m/z$$

$$[0224] \quad \text{其中 } k = (-1, 0, 1, \dots, 4, 5, 6)$$

[0225] 细节如下:

$$[0226] \quad m/z(z)_{-1} = m/z(PX) - \Delta m/z$$

$$[0227] \quad m/z(z)_0 = m/z(PX)$$

$$[0228] \quad m/z(z)_1 = m/z(PX) + \Delta m/z$$

$$[0229] \quad m/z(z)_2 = m/z(PX) + 2 * \Delta m/z$$

$$[0230] \quad m/z(z)_3 = m/z(PX) + 3 * \Delta m/z$$

$$[0231] \quad m/z(z)_4 = m/z(PX) + 4 * \Delta m/z$$

$$[0232] \quad m/z(z)_5 = m/z(PX) + 5 * \Delta m/z$$

$$[0233] \quad m/z(z)_6 = m/z(PX) + 6 * \Delta m/z$$

[0234] 然后,在分配给所测m/z值的所研究部分的所测得的质谱中,鉴定出分配给峰PX和电荷z的同位素分布的所有峰。

[0235] 因此,对于每个峰,围绕其先前定义的理论m/z值定义搜索窗口。

[0236] 在本发明方法的优选实施例中,对于正k值,具有理论m/z值 $m/z(z)_k$ 的同位素分布的峰的搜索窗口被定义为:

$$[0237] \quad m/z(z)_{k-k*\delta\Delta m_{\text{low}}/z} \leq m/z \leq m/z(z)_{k+k*\delta\Delta m_{\text{high}}/z}$$

[0238] 值 $\delta\Delta m_{\text{low}}$ 和 $\delta\Delta m_{\text{high}}$ 与峰的平均同位素质量差 Δm 的可能偏差(同位素分布到较低质量和较高质量)相关。

[0239] $\delta\Delta m_{\text{low}}$ 的典型值介于0.004到0.007之间,优选地介于0.005到0.006之间。 $\delta\Delta m_{\text{high}}$ 的典型值介于0.003到0.006之间,优选地介于0.0035到0.0045之间。

[0240] 对于理论m/z值 m/z_k 周围的m/z值的搜索窗口中每个定义的同位素分布的峰,鉴定具有最高强度的峰并将其分配给所述峰。对于这个峰,确定强度 $I_k(z)$ 和实际观察到的m/z值 $m/z(z)_{k_{\text{obs}}}$ 。

[0241] 为了进一步评估电荷得分 $C_{\text{SPX}}(z)$,仅考虑强度不小于所研究同位素分布的最高峰PX的强度百分比的峰。通常,最高峰PX的强度百分比(所考虑的峰应具有)应介于2%到10%之间,具体地说,介于3%到6%之间。

[0242] 在本发明的一个实施例中,还考虑了定位在m/z值的搜索窗口的边界处且不能被识别为与周围相比具有最大值的真实峰的峰。在这种情况下,未将边界处的峰分配给同位素分布的搜索峰。然后,将m/z值的搜索窗口的边界之外的下一个峰鉴定为同位素分布的搜索峰,因为在这种情况下,所述峰的侧面定位在m/z值的搜索窗口的边界处。同样对于这个

峰,确定强度 $I_k(z)$ 和实际观察到的 m/z 值 $m/z(z)_{k_obs}$ 。

[0243] 在本发明方法的优选实施例中,可以根据至少三个子电荷得分 $c_{Si_PX}(z)$ 推导出所测得的峰PX的电荷得分 $c_{SPX}(z)$ 。

[0244] 在一个实施例中,可以通过将至少三个子电荷得分 $c_{Si_PX}(z)$ 相乘来推导出所测得的峰PX的电荷得分 $c_{SPX}(z)$ 。

[0245] 在优选的实施例中,可以通过将四个子电荷得分 $c_{Si_PX}(z)$ 与 $i=1,2,3,4$ 相乘来推导出所测得的峰PX的电荷得分 $c_{SPX}(z)$ 。

[0246] $c_{SPX}(z) = c_{S1_PX}(z) * c_{S2_PX}(z) * c_{S3_PX}(z) * c_{S4_PX}(z)$

[0247] 可以在本发明的方法中使用的评估子电荷得分 $c_{SP_PX}(z)$ 的一种可能性是使用帕特森函数。所述方法描述于以下文献中:M.W.Senko等人,《美国质谱学会杂志》1995,6,52-56。

[0248] 通常,这种子电荷得分通过下式计算:

$$[0249] \quad c_{SP_PX}(z) = \sum_{j=-N_{left_PX}(z)+1}^{N_{right_PX}(z)} I_{j-1}(z) * I_j(z)$$

[0250] 在优选的实施例中,在子电荷得分 $c_{SP_PX}(z)$ 的计算中,通过定义同位素分布的每个峰的校正强度 $I_{corr_k}(z)$ 来考虑同位素分布的每个峰的观察到的 m/z 值 $m/z(z)_{k_obs}$ 与理论 m/z 值 $m/z(z)_k$ 的偏差:

$$[0251] \quad I_{corr_k}(z) = I_k(z) * (1 - 2 * ((m/z(z)_{k_obs} - m/z(z)_k) / W_k)^2)$$

[0252] W_k 是具有理论 m/z 值 $m/z(z)_k$ 的同位素分布峰的半峰全宽(FWHM)。

[0253] 仅使用在观察到的 m/z 值 $m/z(z)_{k_obs}$ 的 m/z 范围内高于噪声水平的校正强度 $I_{corr_k}(z)$ 。否则,将校正后的强度 $I_{corr_k}(z)$ 设置为观察到的 m/z 值 $m/z(z)_{k_obs}$ 的 m/z 范围内的噪声水平。

[0254] 然后,通过下式计算子电荷得分:

$$[0255] \quad c_{SP_PX}(z) = \sum_{j=-N_{left_PX}(z)+1}^{N_{right_PX}(z)} I_{corr_j-1}(z) * I_{corr_j}(z)$$

[0256] 可以在本发明的方法中使用的评估子电荷得分 $c_{SAS_PX}(z)$ 的第二种可能性是使用准确性得分。所述方法描述于以下文献中:Z.Zhang和A.G.Marshall,《美国质谱学会杂志》1998,9,225-233。

[0257] 首先,定义同位素分布的每个峰的Z得分。所述值描述了同位素分布峰可能出现的最大偏差与实际观察到的 m/z 值 $m/z(z)_{k_obs}$ 和理论值 $m/z(z)_k$ 的实际偏差之间的比率。Z得分 $Z_k(z)$ 由下式给出:

$$[0258] \quad Z_k(z) = \delta_{m/z_max} * m/z_{PX} / |m/z(z)_{k_obs} - m/z(z)_k|$$

[0259] δ_{m/z_max} 是用于测量样品质谱的质谱仪的 m/z 的最大相对偏差。

[0260] 优选地,将Z得分 $Z_k(z)$ 限制于特定的一系列值。这可能是例如1到5之间的一系列值。

[0261] 然后,通过汇总所研究同位素分布的所有峰的Z得分值来评估子电荷得分 $c_{SAS_PX}(z)$

$$[0262] \quad CS_{AS_PX}(z) = \sum_{j=-N_{left_PX}(z)}^{N_{right_PX}(z)} Z_k(z).$$

[0263] 可以在本发明的方法中使用的评估子电荷得分 $CS_{AC_PX}(z)$ 的第三种可能性是使用自相关函数,所述自相关函数对同位素分布的峰中的波动进行评估。

[0264] 为了计算所述子电荷得分,再次使用上述针对同位素分布的每个峰的校正强度 $I_{corr_k}(z)$ 。

[0265] 通过下式计算子电荷得分 $CS_{AC_PX}(z)$ ：

$$[0266] \quad CS_{AC_PX}(z) = \sum_{j=-N_{left_PX}(z)+1}^{N_{right_PX}(z)} I_{corr_j-1}(z) * I_{corr_j}(z) / \sum_{j=-N_{left_PX}(z)}^{N_{right_PX}(z)} I_{corr_j}(z)^2$$

[0267] 所述电荷得分优选地仅用于具有至少3个峰(优选4个峰)的同位素分布。否则,电荷得分将设置为值1。

[0268] 可以在本发明的方法中使用的评估子电荷得分 $CS_{IS_PX}(z)$ 的第四种可能性是使用同位素得分。所述得分将观察到的同位素分布的峰数 $N_{obs_PX}(z)$ 与理论上预期的峰数 $N_{theo_PX}(z) = N_{left_PX}(z) + N_{left_PX}(z) + 1$ 相对应。

[0269] 可以通过下式计算子电荷得分 $CS_{IS_PX}(z)$ ：

$$[0270] \quad CS_{IS_PX}(z) = (N_{obs_PX}(z) + 0.5) / (N_{theo_PX}(z) - 1).$$

[0271] 在本发明方法的优选实施例中,通过将四个子电荷得分 $CS_{P_PX}(z)$ 、 $CS_{AS_PX}(z)$ 、 $CS_{AC_PX}(z)$ 和 $CS_{IS_PX}(z)$ 中的至少三个相乘来推导出所测得的峰PX的电荷得分 $CS_{PX}(z)$ 。

[0272] 在本发明方法的特别优选的实施例中,通过将四个子电荷得分 $CS_{P_PX}(z)$ 、 $CS_{AS_PX}(z)$ 、 $CS_{AC_PX}(z)$ 和 $CS_{IS_PX}(z)$ 相乘来推导出所测得的峰PX的电荷得分 $CS_{PX}(z)$ 。

$$[0273] \quad CS_{PX}(z) = CS_{P_PX}(z) * CS_{AS_PX}(z) * CS_{AC_PX}(z) * CS_{IS_PX}(z)$$

[0274] 在电荷1与最大电荷态 z_{max} 之间的每个电荷态 z 都有一个得分值之后,根据所测得的 m/z 值的所研究部分中的质谱评估出峰P1(最高强度峰)的电荷得分 $CS_{P1}(z)$,对峰P1的电荷得分 $CS_{P1}(z)$ 进行排名。然后,将电荷态 z_1 的最高值电荷得分 $CS_{P1}(z_1)$ 与电荷态 z_2 的第二最高值电荷得分 $CS_{P1}(z_2)$ 进行比较。如果这些值的比率高于阈值 T_{cs} ,则将电荷态 z_1 接受为峰P1及其相关同位素分布的正确电荷态。

$$[0275] \quad CS_{P1}(z_1) / CS_{P1}(z_2) > T_{cs}$$

[0276] 因此,如果电荷态 z_1 被接受,则根据所测得的质谱的峰P1及其周围的质谱推导出其相关的同位素分布,所述同位素分布具有峰强度 $I_k(z_1)$ 和实际观察到的 m/z 值 $m/z(z_1)_{k_obs}$ ($k = (-N_{left_PX}(z_1), \dots, N_{right_PX}(z_1))$)和特定电荷 z_1 。所述同位素分布是一个分子种类的离子的同位素分布。所述分子种类包含在被研究的样品中,所述样品已通过电离过程带电而质量不变,或者一个分子种类的离子至少通过电离过程从样品中产生。

[0277] 通过阈值 T_{cs} 的值,可以定义具有最高值的两个最佳评估电荷得分 $CS_{P1}(z_1)$ 和 $CS_{P1}(z_2)$ 与电荷态 z_1 相关的同位素分布必须有多大差异,所述同位素分布可以被明确推导为包括峰P1的同位素分布。通常,阈值 T_{cs} 的值在1.10到3的范围内,优选在1.15到2的范围内,并且优选在1.20到1.50的范围内。阈值 T_{cs} 的值可以由用户、控制器或控制器的生产者通过硬件或软件进行设置。

[0278] 根据推导出的具有特定电荷 z_1 的一个分子种类的离子的同位素分布,可以通过本领域技术人员已知的方法推导所述分子种类的单同位素质量和/或所述分子种类的单同位

素峰,例如通过对同位素分布的峰图进行平均拟合或直接在同位素分布的同位素图型中寻找单同位素峰。

[0279] 在可以推导出包括峰P1的同位素分布之后,从部分的显著峰中除去所述同位素分布的峰。然后定义所述部分的其余显著峰中的最高强度峰。然后,必须以与峰1相同的方式定义所述峰P2的最大电荷态 z_{\max} ,针对电荷1与最大电荷态 z_{\max} 之间的每个电荷态 z ,必须根据所测得的 m/z 值的所研究部分中的质谱评估电荷得分 $C_{SP2}(z)$,并且必须检查最高值电荷得分 $C_{SP2}(z_1)$ 是否被接受为峰P2的正确电荷态。通过尽可能多地重复此过程,可以根据由单个处理器提供的质谱中所测得的 m/z 值的至少一个范围内的部分推导出具有特定电荷 Z 的分子种类的离子的同位素分布以及所述分子种类的单同位素质量。

[0280] 优选地,通过指定处理器对具有显著峰的质谱的所测得的 m/z 值的至少一个范围内的所有部分进行此操作。

[0281] 因此,根据所测得的 m/z 值的至少一个范围内的整个 m/z 范围,可以通过与多个处理器的多个处理器并行推导,逐步推导具有特定电荷的分子种类的离子的同位素分布。通过将所测得的 m/z 值的至少一个应研究的范围划分成多个部分并将这些部分分配给几个处理器,可以快速推导出所测得的 m/z 值的至少一个范围内的整个 m/z 范围的同位素分布,而且还可以根据所述推导出的同位素分布来推导单同位素质量。具体地说,推导出的单同位素质量可以用于定义特定分子种类,将使用第二质量分析器进一步研究所述分子。特别是对于这个实验,本发明的方法非常有用,因为现在可以在更短的时间内获得有关特定分子的单同位素质量的信息。在将要用第二质量分析器进一步研究的特定分子种类提供给质量分析器之前,可以在碰撞池或反应池中通过 MS^2 或 MS^N 质谱分析中使用的典型过程(例如碎裂、解离)将其转化为另一个分子。

[0282] 在本发明方法的另一个可能步骤中,根据包含在样品中的和/或源自样品的至少一个分子种类中的每一个分子的至少一种推导出的同位素分布,推导所述分子种类的单同位素质量。在本发明方法的一个实施例中,在推断出同位素分布之后,立即根据所述分子种类的同位素分布推断出包含在样品中的和/或源自所研究样品的所述分子种类的单同位素质量。在这个实施例中,可以规定,在推导一个分子种类的同位素分布之前推导另一分子种类的单同位素质量。在本发明方法的一个实施例中,规定某些分子种类的单同位素质量推导发生在其它分子种类的同位素分布推导之前。

[0283] 通常,在本发明方法的一些实施例中,同位素分布的推导(本发明方法的步骤(iv))和单同位素质量的推导(步骤(v))可以并行发生。

[0284] 在本发明方法的优选实施例中,针对包含在样品中的和/或至少通过电离过程从样品中产生的某些分子种类,根据其离子的两种或更多种推导出的同位素分布来推导单同位素质量,所述离子具有不同的特定电荷 z 。

[0285] 在通过用多个处理器的几个处理器进行并行推导,按部分根据所测得的 m/z 值的至少一个范围内的整个 m/z 范围,推导出具有特定电荷 z 的分子种类的离子的同位素分布之后,推导出的同位素分布中的两种或更多种同位素分布可能是一个分子种类的离子的同位素分布,所述离子具有不同的特定电荷 z 。大多数情况下,在所测得的 m/z 值的至少一个范围内的不同部分中推导出这些同位素分布。但是,也可以在所测得的 m/z 值的至少一个范围内的一个部分中推导出这些同位素分布。当根据所测得的 m/z 值的至少一个范围内的部分推

导出同位素分布时,也可能已经鉴定了具有特定电荷 z 的一个分子种类的离子的一种同位素分布,而尚未根据所测得的 m/z 值的至少一个范围内的部分推导出具有另一特定电荷 z' 的同一分子种类的离子的另一种同位素分布。

[0286] 通常,质谱仪可检测到的一个分子种类的不同离子可以按以下方式变化:

[0287] (i) 仅不同离子的电荷偏离而质量相同。这种离子可能是由于通过电离过程添加或删除电子而产生的。

[0288] 实例:添加电子(电荷 $z=-1$)

[0289] 第一离子:质量 m 电荷 z

[0290] 第二离子:质量 m 电荷 $z-1$

[0291] (ii) 添加质量为 m_a 且电荷为 z_a 的离子

[0292] 实例:添加质量为 m_a 且电荷为 z_a 的离子

[0293] 第一离子:质量 m 电荷 z

[0294] 第二离子:质量 $m+m_a$ 电荷 $z+z_a$

[0295] 以离子形式添加的典型加合物为 H^+ 、 Na^+ 、 K^+ 以及乙酸和甲酸的离子。

[0296] 在电喷雾电离期间,添加质量 $m=1$ 且电荷 $z=1$ 的质子(H^+):带有或不带有添加的质子的两个所得离子为:

[0297] 第一离子:质量 m 电荷 z

[0298] 第二离子:质量 $m+1$ 电荷 $z+1$

[0299] 具有不同特定电荷的同一分子的离子的同位素分布的可能发生率可用于本发明方法的另一步骤中,以改进对所述分子种类的单同位素质量的确定。

[0300] 首先,根据从所测得的 m/z 值的至少一个范围内的整个 m/z 范围中按部分推导出的具有特定电荷 z 的分子种类的离子的全部同位素分布,定义了分子 $M1$ 种类的同位素分布,当根据所测得的 m/z 值的至少一个范围内的部分推导出所述分子的同位素分布时,发现其电荷得分 $c_{SM1}(z)$ 的最高值。对于所述分子 $M1$,研究了具有最高 S 值的 S 电荷得分为 $c_{SM1}(z_1) \cdots c_{SM1}(z_s)$ 的离子的同位素分布。通常,所研究的电荷得分的数量在2到8之间,优选在4到6之间。对于具有特定电荷 z 的特定分子的离子的这种同位素分布中的每一个,考虑了特定分子种类的离子的相邻同位素分布,所述离子具有介于 $z-\Delta z$ 与 $z+\Delta z$ 之间的电荷。 Δz 的典型值介于1到5之间,优选为2或3。因此,对于 $\Delta z=2$,考虑了具有电荷 $z-2$ 、 $z-1$ 、 z 、 $z+1$ 、 $z+2$ 的离子。还必须考虑到,取决于分子种类的离子的电离过程,离子的质量也可以如上所述改变。

[0301] 通过将所考虑的相邻同位素分布的电荷得分加到电荷得分上,可以计算出 S 电荷得分为 $c_{SM1}(z_1) \cdots c_{SM1}(z_s)$ 的离子的同位素分布的新电荷得分 $c_{SM1_A}(z_X)$ 。

[0302] 例如:

[0303] $c_{SM1_A}(z_1) = c_{SM1}(z_1 - \Delta z) + \cdots + c_{SM1}(z_1) + \cdots + c_{SM1}(z_1 + \Delta z)$

[0304] 如果已经根据所测得的 m/z 值的至少一个范围内的部分推断出了特定分子种类的离子的相邻同位素分布,则可以使用推导出的同位素分布的评估电荷得分。否则,根据所研究的同位素分布的最高峰的 m/z 值 m_h/z_h ,可以得出所考虑的相邻同位素分布的最高峰的 m/z 值,一个分子种类的不同离子可以根据其如上所述的电离而变化。例如,对于电喷雾电离,电荷 $z+\Delta z$ 的相邻峰的 m/z 值为 $(m_h + \Delta z) / (z_h + \Delta z)$ 。

[0305] 具有理论 m/z 值 m/z_n 的相邻同位素分布的最高峰的搜索窗口由以下定义:

[0306] $m/z_n - \delta m/z_{iso} \leq m/z \leq m/z_n + \delta m/z_{iso}$

[0307] 可以根据相邻同位素分布的电荷和/或相邻同位素分布的观察到的和预期的最高峰的质量的最大偏差来选择窗口宽度 $2 * \delta m/z_{iso}$ 。

[0308] 对于在搜索窗口中观察到的相邻同位素分布的这一最高峰PN,必须鉴定同位素分布的其它峰,并且必须根据上述方法评估根据其电荷 z_n 的电荷得分 $c_{SPN}(z_n)$,以便在所测得的 m/z 值的至少一个范围内的部分中推导同位素分布。然后使用这些电荷得分 $c_{SPN}(z_n)$ 计算新的电荷得分 $c_{SM1_A}(z_x)$ 。可以在多处理器的不同处理器上并行完成缺失的相邻同位素分布的鉴定和电荷得分 $c_{SPN}(z_n)$ 的评估,以加快处理速度。

[0309] 如果已计算出S电荷得分为 $c_{SM1}(z_1) \cdots c_{SM1}(z_s)$ 的离子的同位素分布的新电荷得分 $c_{SM1_A}(z_x)$,则对新电荷得分 $c_{SM1_A}(z_x)$ 进行排名。然后,将电荷态 z_{H1} 的最高值电荷得分 $c_{SM1_A}(z_{H1})$ 与电荷态 z_{H2} 的第二最高值电荷得分 $c_{SM1_A}(z_{H2})$ 进行比较。如果这些值的比率高于阈值 T_{cs2} ,则将电荷态 z_{H1} 接受为分子M1种类的正确起始电荷态,以定义分子M1种类的相关同位素分布的正确集合。

[0310] $c_{SM1_A}(z_{H1}) / c_{SM1_A}(z_{H2}) > T_{cs2}$

[0311] 通过阈值 T_{cs2} 的值,可以定义具有最高值的两个最佳评估电荷得分 $c_{SM1_A}(z_{H1})$ 和 $c_{SM1_A}(z_{H1})$ 与起始电荷态 z_{H1} 相关的同位素分布集必须有多大差异,所述同位素分布可以被明确推导为分子M1种类的同位素分布集。通常,阈值 T_{cs2} 的值在1.10到3的范围内,优选在1.15到2的范围内,并且优选在1.20到1.50的范围内。阈值 T_{cs2} 的值可以由用户、控制器或控制器的生产者通过硬件或软件进行设置。

[0312] 根据推导出的分子M1种类的离子的同位素分布集,可以通过本领域技术人员已知的方法推导所述分子M1种类的单同位素质量和/或所述分子M1种类的单同位素峰,例如通过对同位素分布的峰图进行平均拟合或直接在同位素分布的同位素图型中寻找单同位素峰。

[0313] 在可以推导出分子M1种类的同位素分布集之后,从所测得的 m/z 值的至少一个范围内的整个 m/z 范围内的所有显著峰中除去所述同位素分布集的峰。

[0314] 然后,根据从所测得的 m/z 值的至少一个范围内的整个 m/z 范围(其显著峰未被去除)中按部分推导出的具有特定电荷 z 的分子种类的离子的全部剩余同位素分布,定义了分子M2种类的同位素分布,当根据所测得的 m/z 值的至少一个范围内的部分推导出所述分子的同位素分布时,发现其电荷得分 $c_{SM2}(z)$ 的最高值。对于所述分子M2,研究了具有最高S值的S电荷得分为 $c_{SM2}(z_1) \cdots c_{SM2}(z_s)$ 的离子的同位素分布。

[0315] 然后,必须以与分子M1种类相同的方式推导出所述分子M2种类的同位素分布集。

[0316] 根据推导出的分子M2种类的离子的同位素分布集,可以通过本领域技术人员已知的方法推导所述分子M2种类的单同位素质量和/或所述分子M2种类的单同位素峰,例如通过对同位素分布的峰图进行平均拟合或直接在同位素分布的同位素图型中寻找单同位素峰。

[0317] 通过尽可能频繁地重复此过程,可以推导出分子种类的离子的尽可能多的同位素分布集以及尽所述分子种类的可能多的单同位素质量。

[0318] 所有实施例也属于本发明的说明书的内容,它们是本发明前述实施例的组合。因此,涵盖了所有实施例,其包括之前仅针对单个实施例描述的特征的组合。

[0319] 在所有描述的实施例中,将Avergin模型用作预期的同位素分布的模型。对于本领域技术人员显而易见的是,还可以根据本发明方法中所研究的分子使用预期的同位素分布的其它模型。

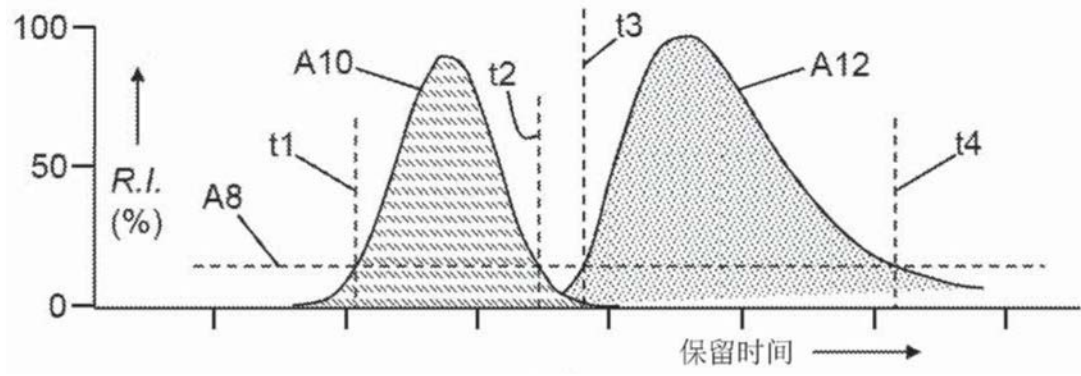


图1A

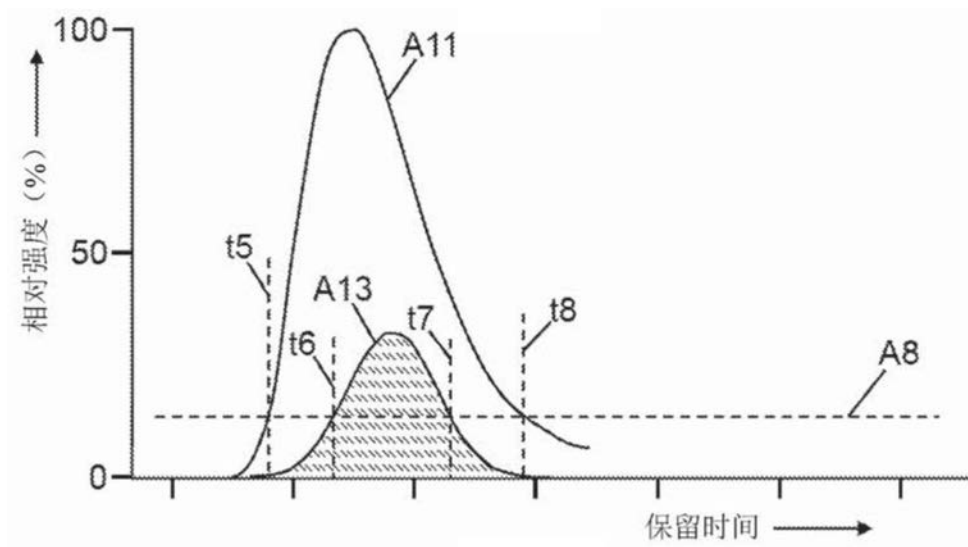


图1B

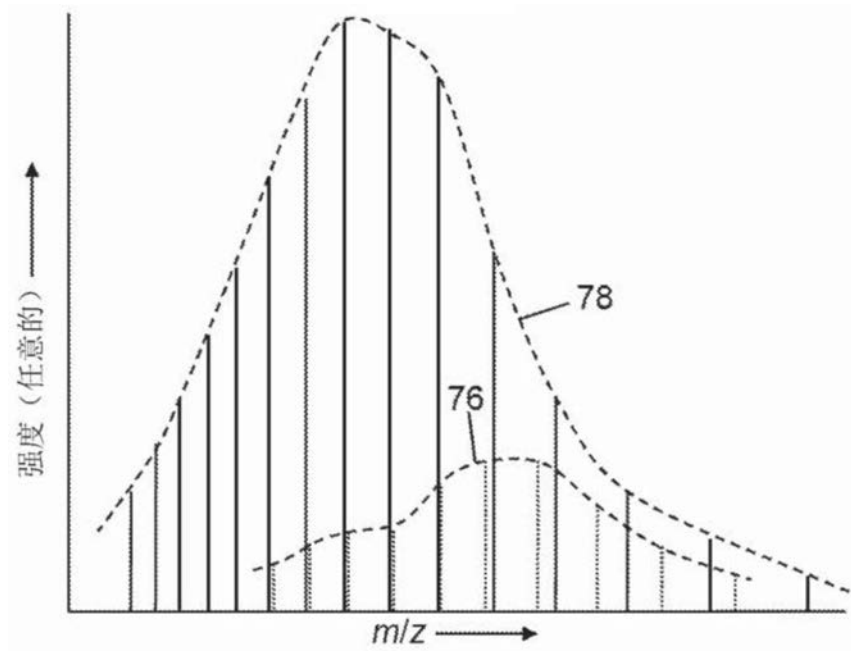


图1C

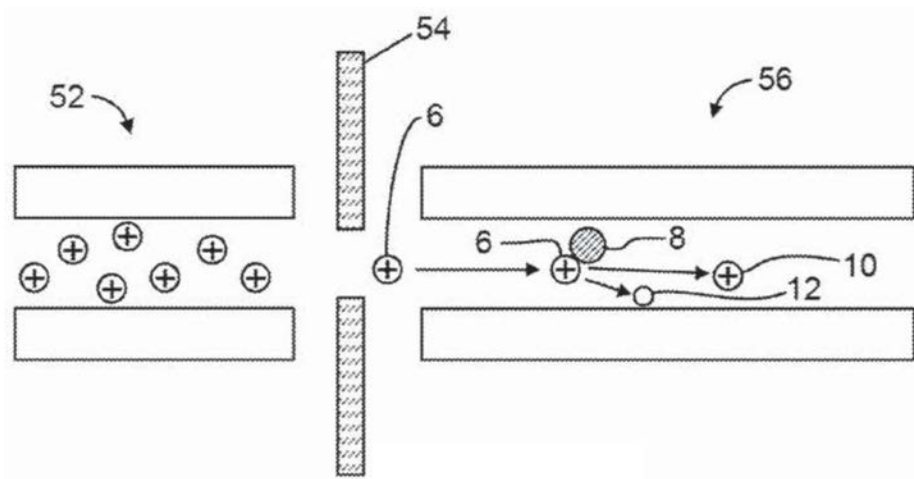


图2 (现有技术)

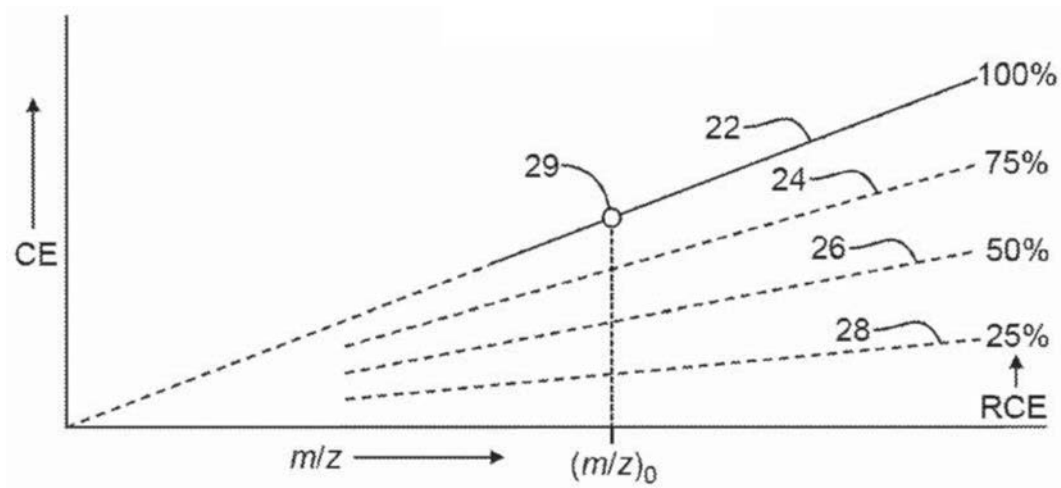


图4A(现有技术)

表1

前体电荷 (z)	系数
1	1
2	0.9
3	0.85
4	0.8
≥5	0.75

图4B(现有技术)

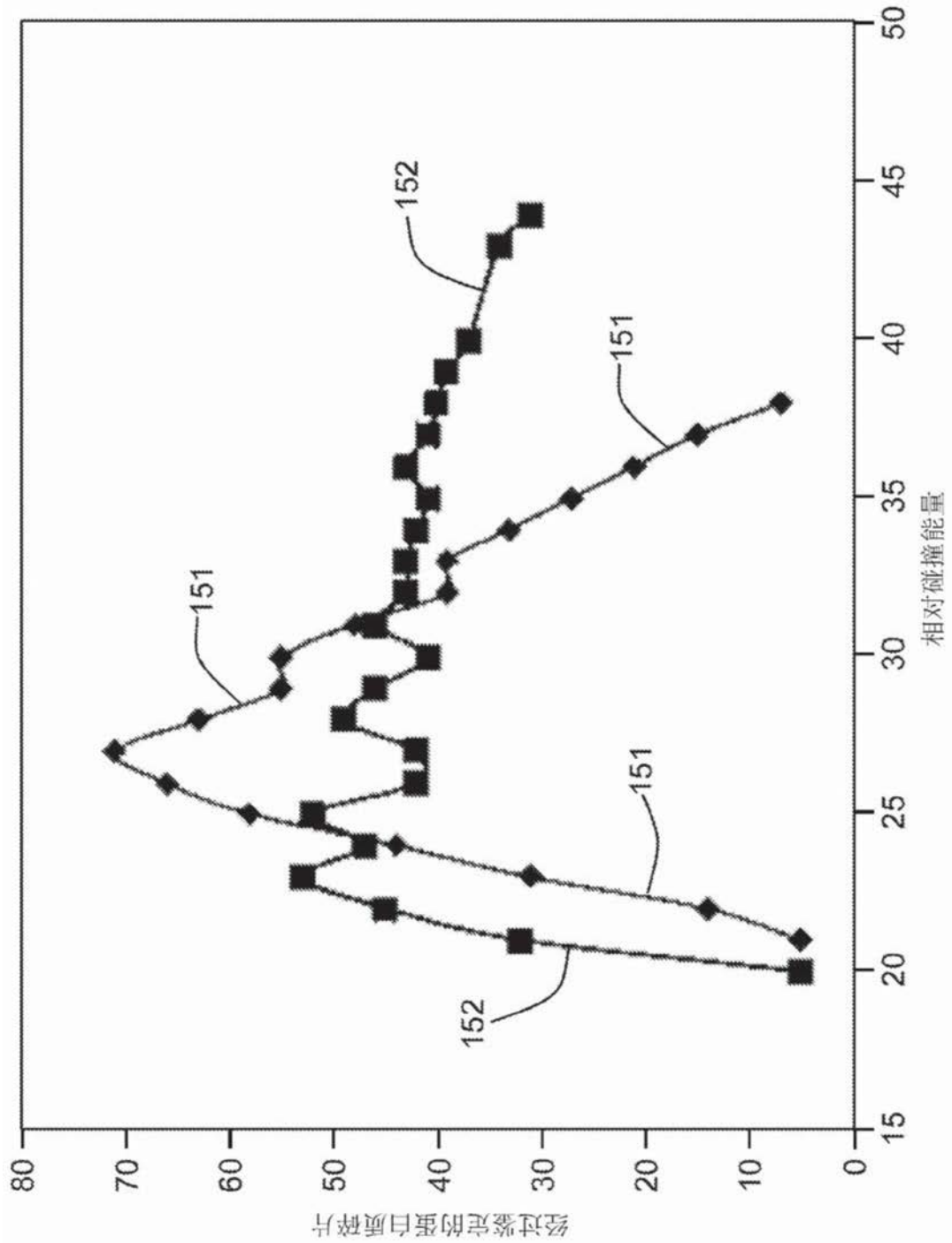


图3A

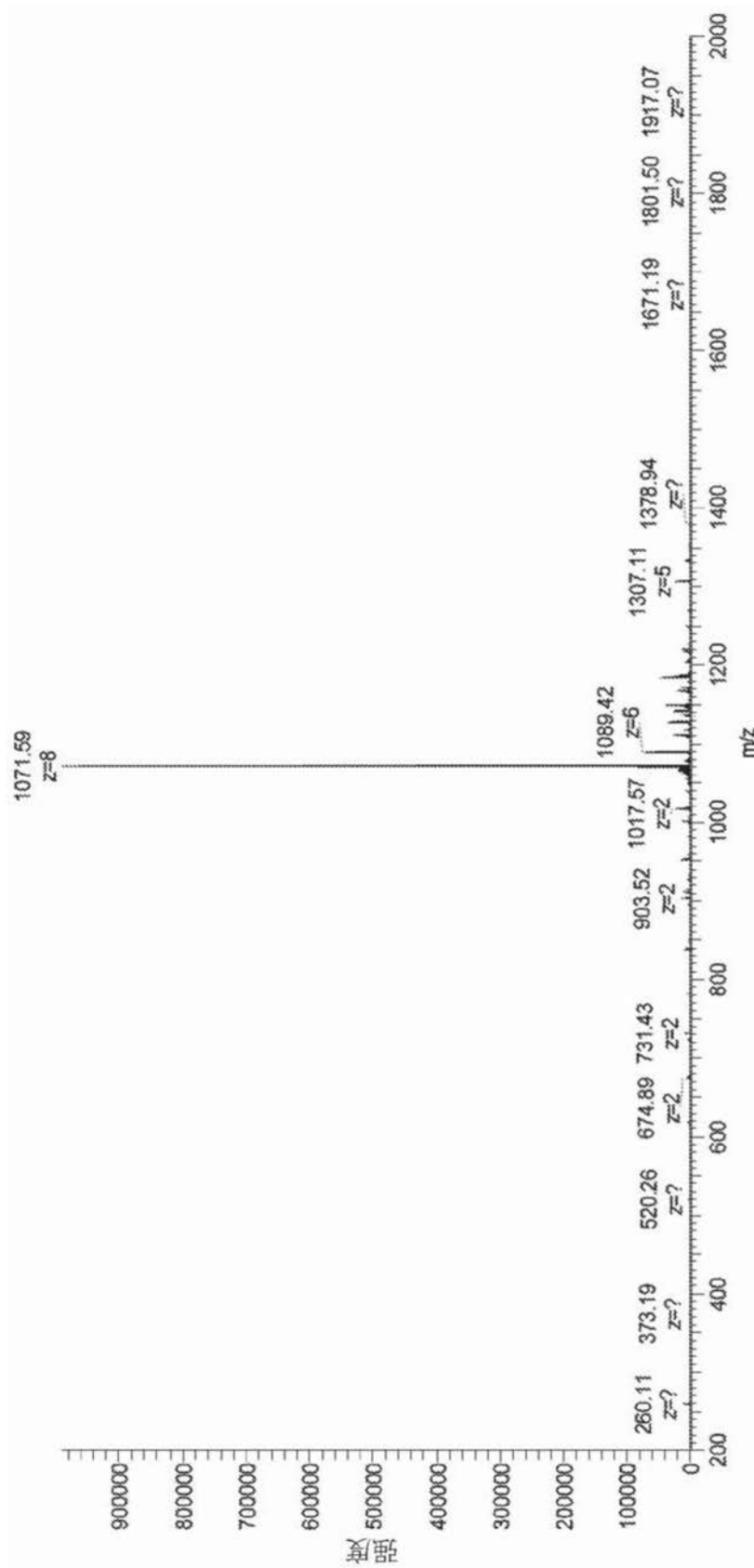


图3B

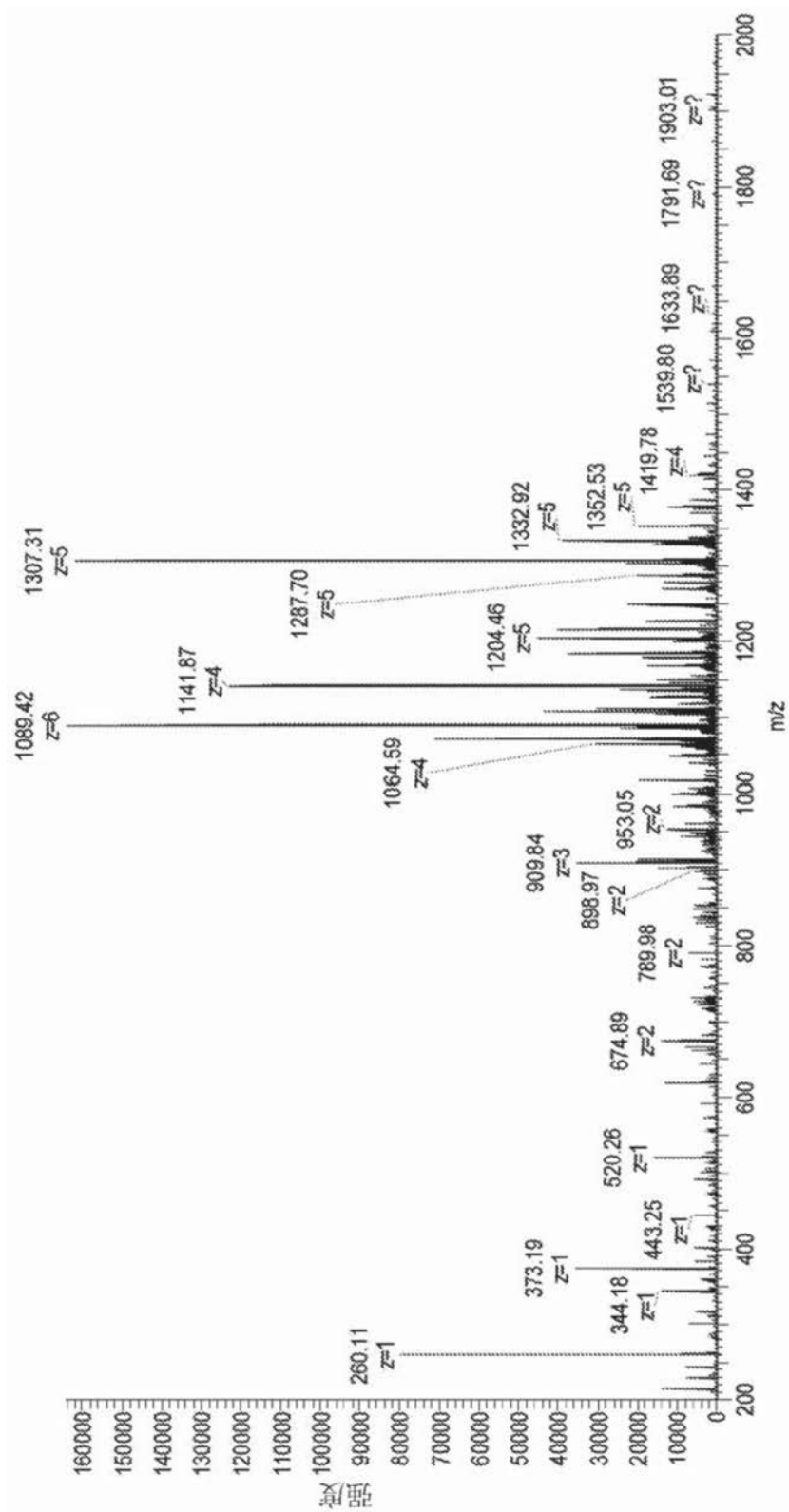


图3C

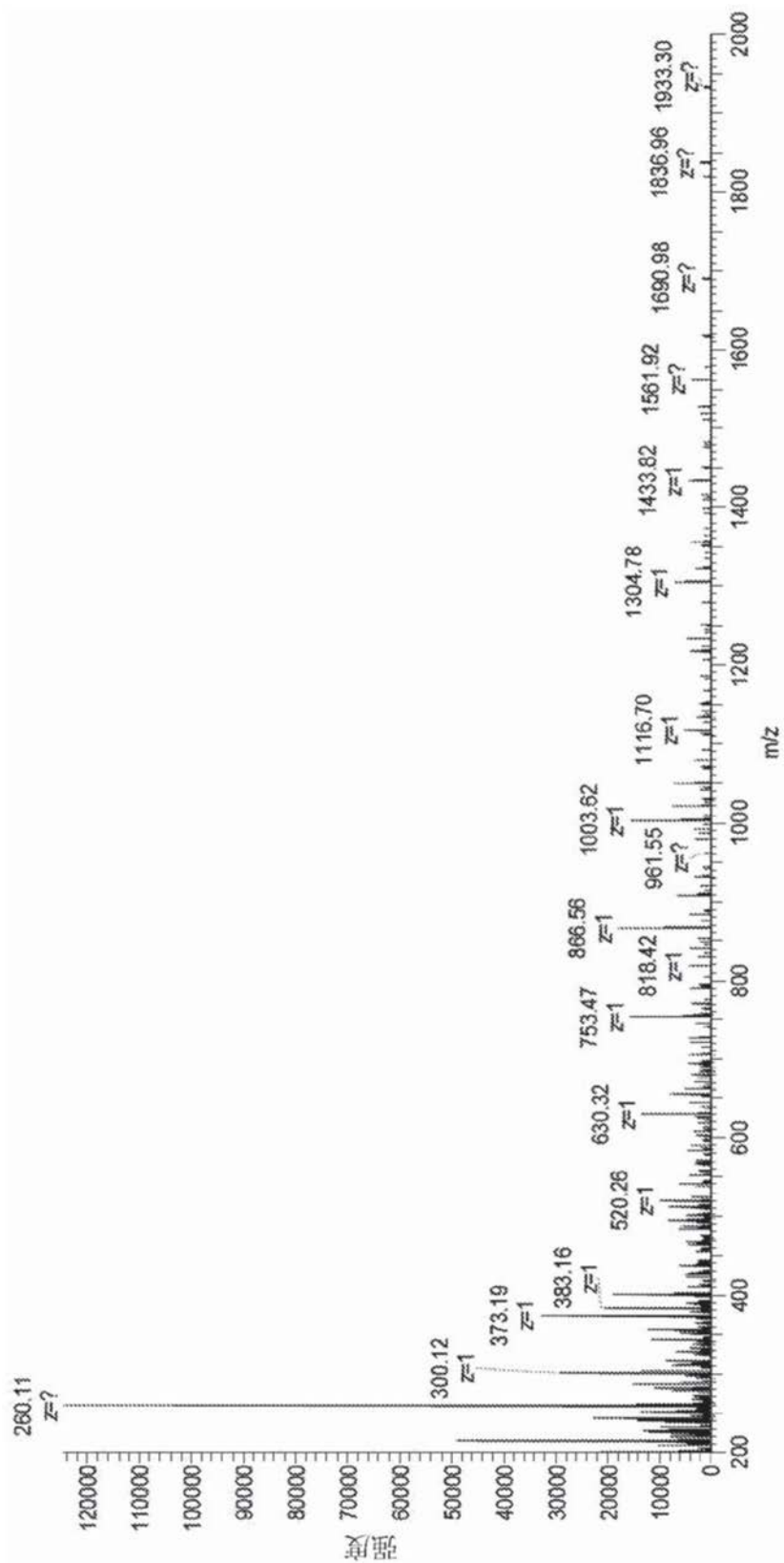


图3D

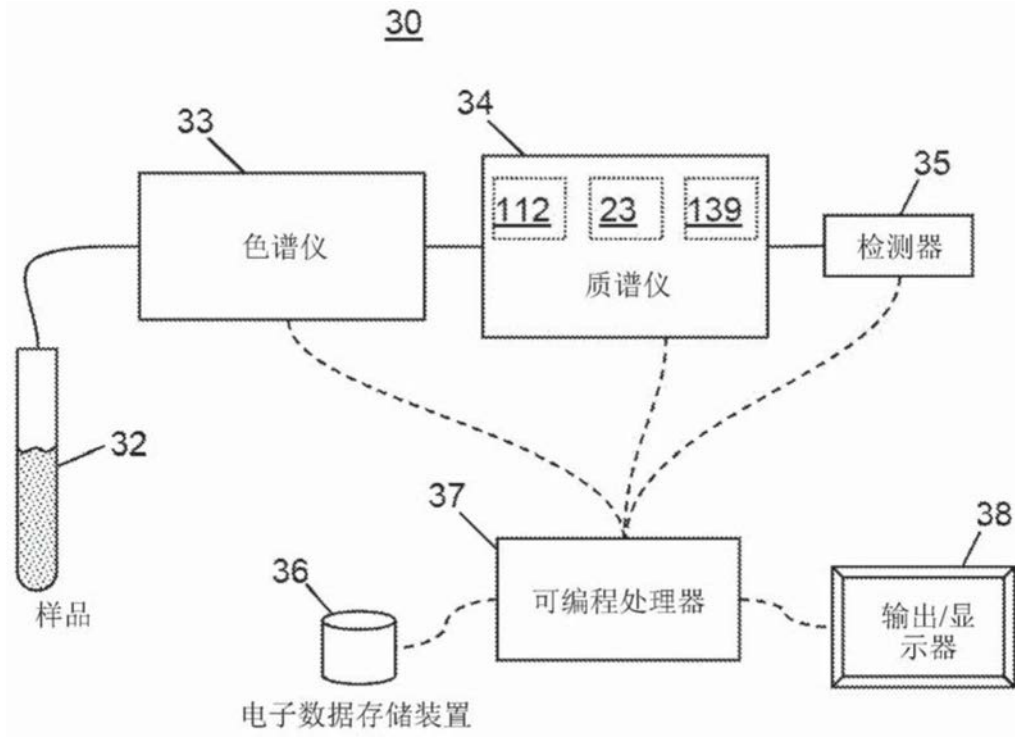


图5A(现有技术)

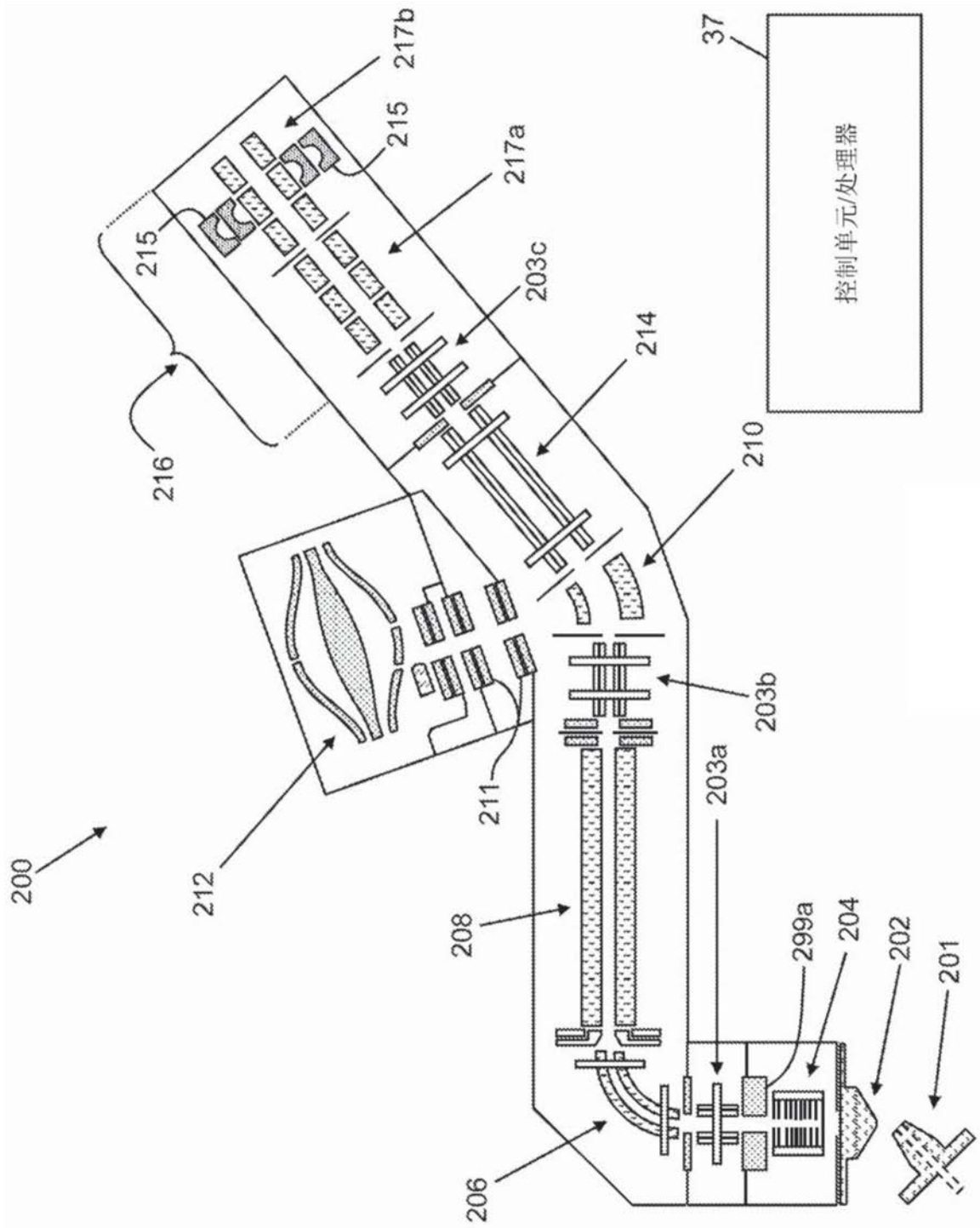


图5B(现有技术)

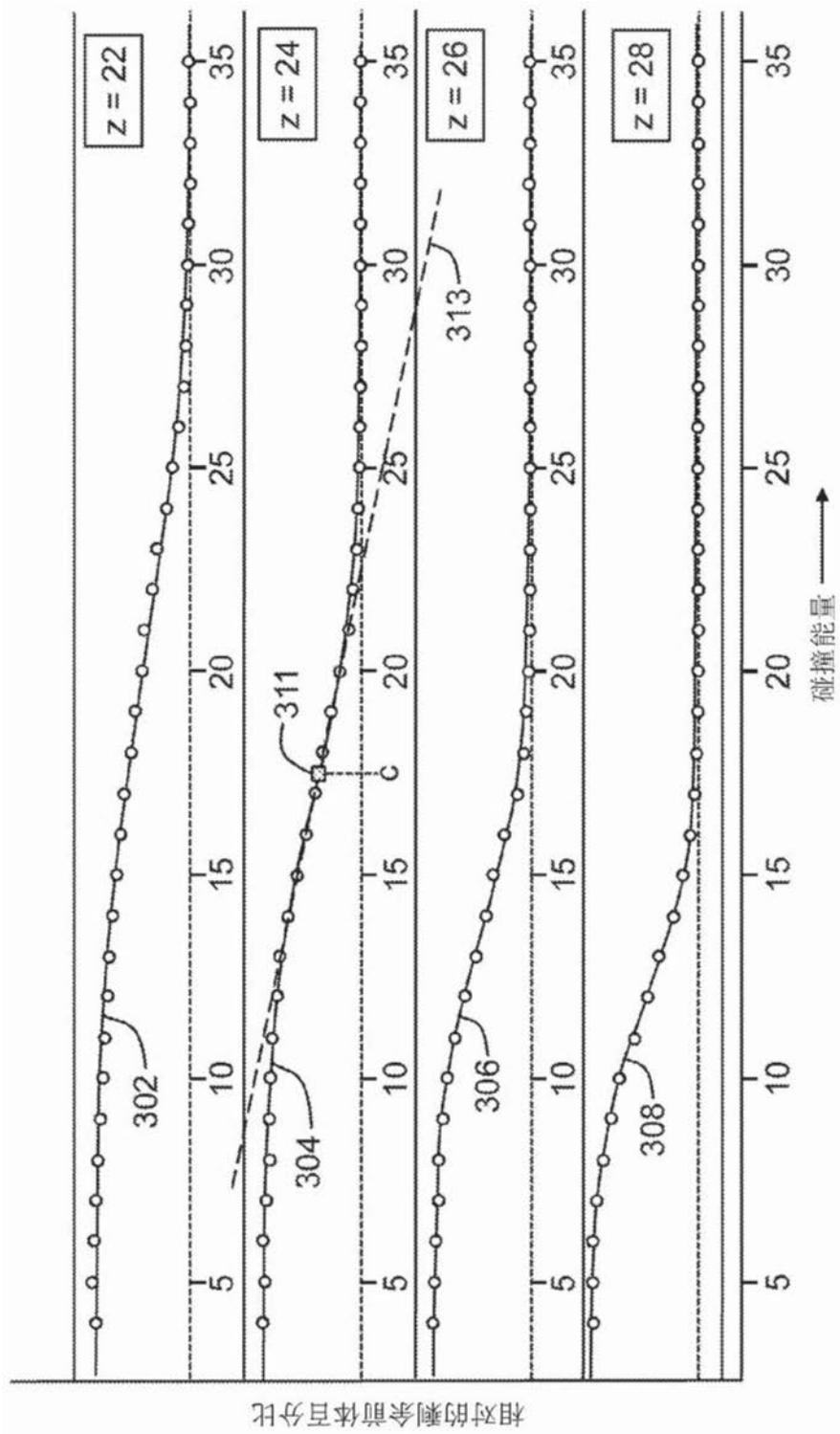


图6A

表2

前体存活率 D_p (百分比)	a1	a2	a3
1	0.00470078	-1.8820703	1.44351911
2	0.00464301	-1.8902764	1.44249578
3	0.00459776	-1.8971044	1.44162661
4	0.00456329	-1.9027399	1.44089112
5	0.00453784	-1.9073683	1.44026882
6	0.0045197	-1.9111701	1.43973978
7	0.00450736	-1.9143058	1.43928624
8	0.00449936	-1.9169309	1.43889099
9	0.00449424	-1.9192008	1.43853683
10	0.00449053	-1.9212711	1.43820653
15	0.00447334	-1.9308513	1.43666196
20	0.00446814	-1.9386297	1.43532955
50	0.00454622	-1.9755979	1.42792781

图6B

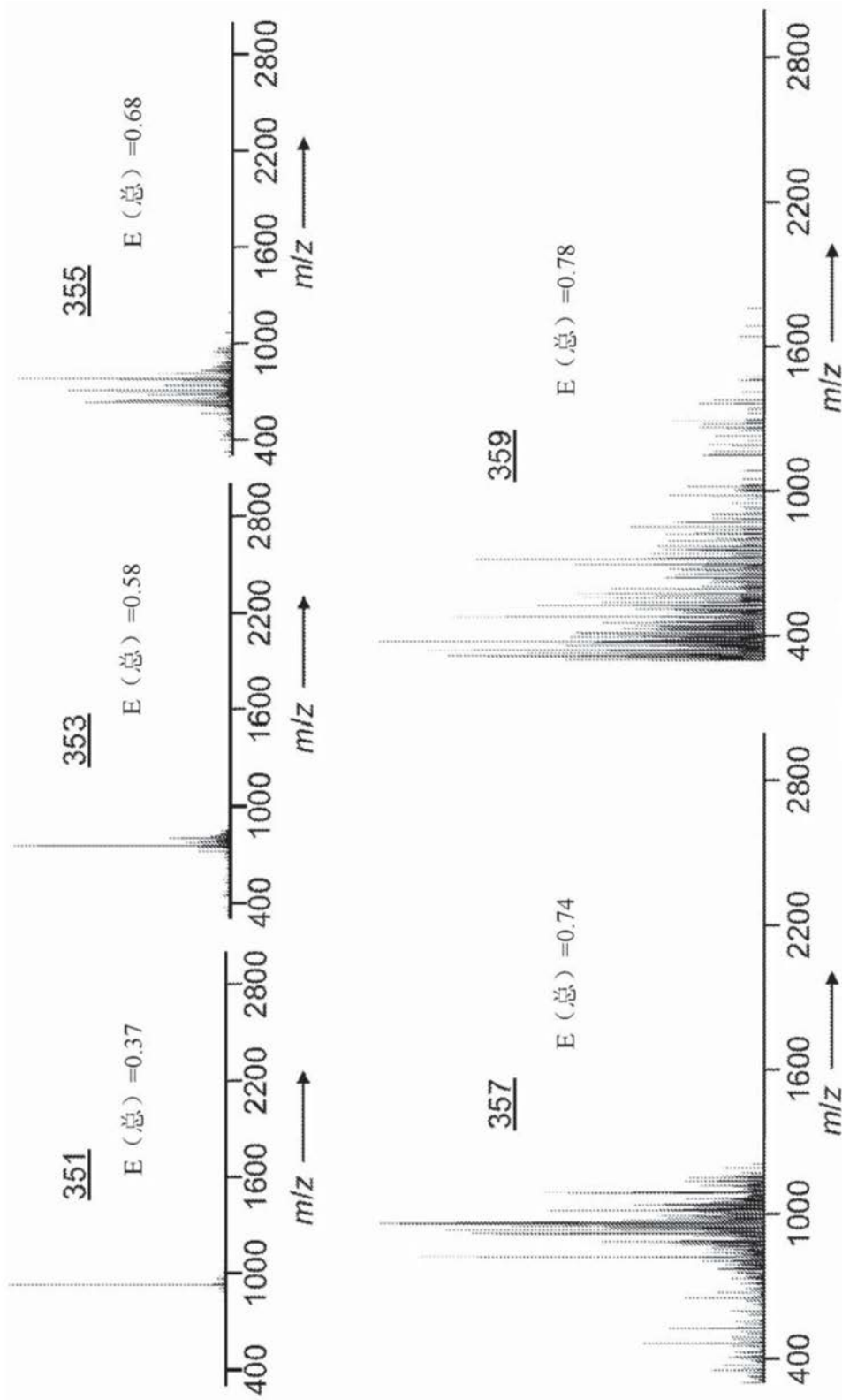


图7A

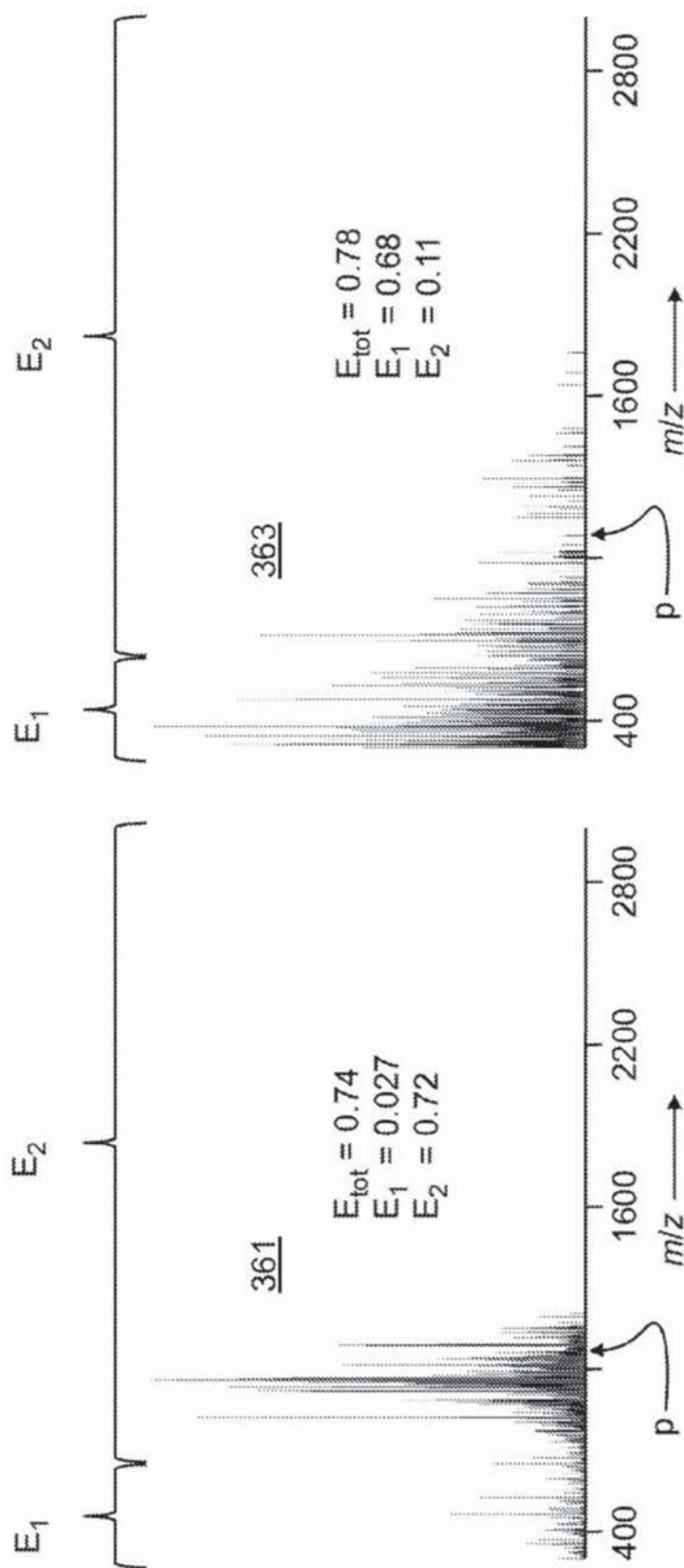


图7B

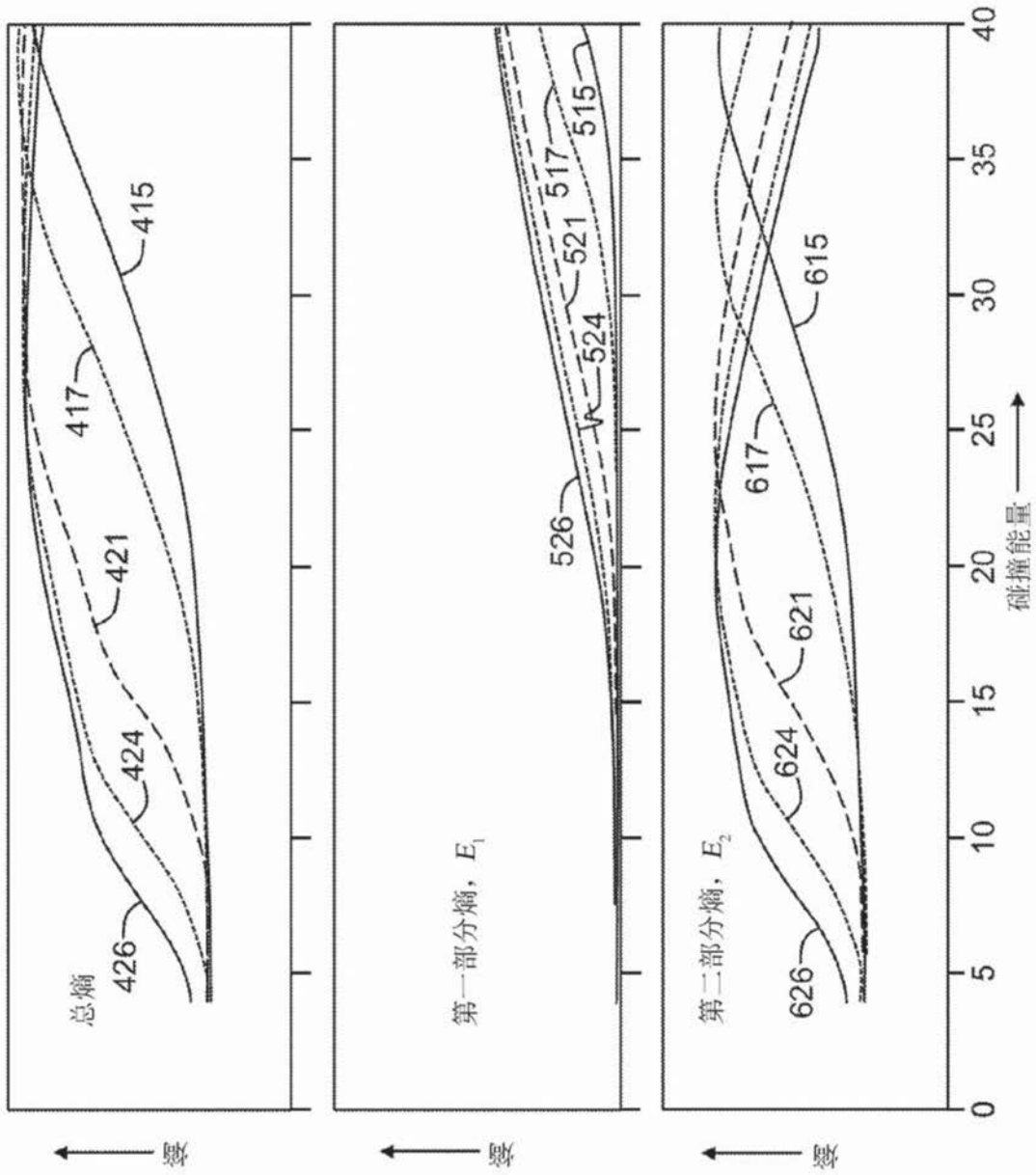


图8A

表3

D_e	$b1$	$b2$	$b3$
0.1	0.03929987	-1.5314701	1.08690053
0.2	0.05394166	-1.4162222	1.03482284
0.3	0.06411948	-1.357906	1.00753659
0.4	0.07229396	-1.3197634	0.98913378
0.5	0.07933121	-1.2917429	0.97521669
0.6	0.08564912	-1.2697276	0.96396574
0.7	0.09148209	-1.2516553	0.95446493
0.8	0.09697853	-1.2363531	0.94618879
0.9	0.10223907	-1.2230954	0.93881014
1	0.10733599	-1.2114043	0.93211223
1.1	0.11232501	-1.2009492	0.92594327
1.2	0.11725044	-1.1914937	0.92019345
1.3	0.12214901	-1.1828638	0.91478086
1.4	0.1270532	-1.1749287	0.90964218
1.5	0.13199199	-1.1675894	0.90472816
1.6	0.13699306	-1.1607707	0.89999942
1.7	0.14208352	-1.1544175	0.89542448
1.8	0.14729118	-1.1484927	0.89097846
1.9	0.15264588	-1.1429787	0.88664267
2	0.15818156	-1.1378844	0.88240572

图8B

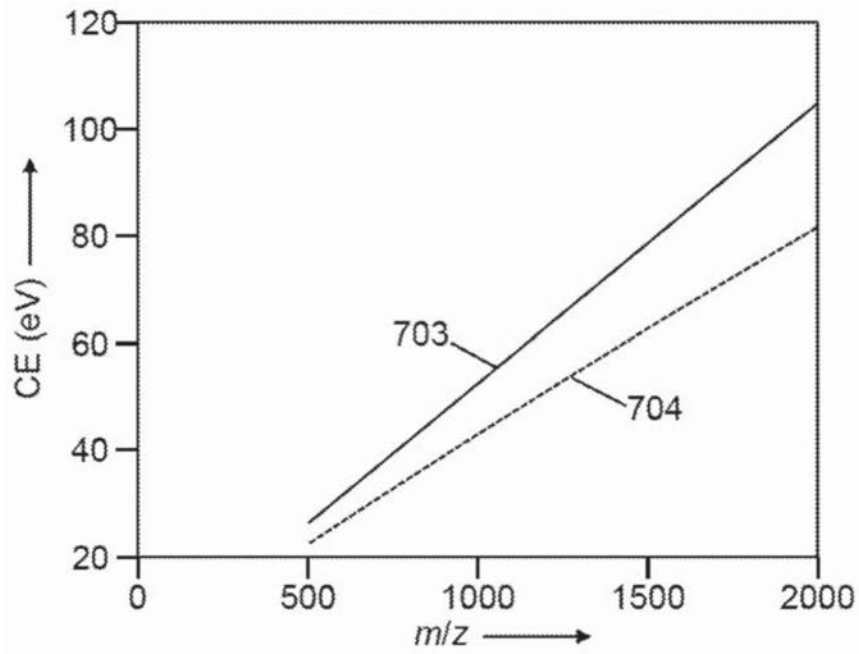


图9A

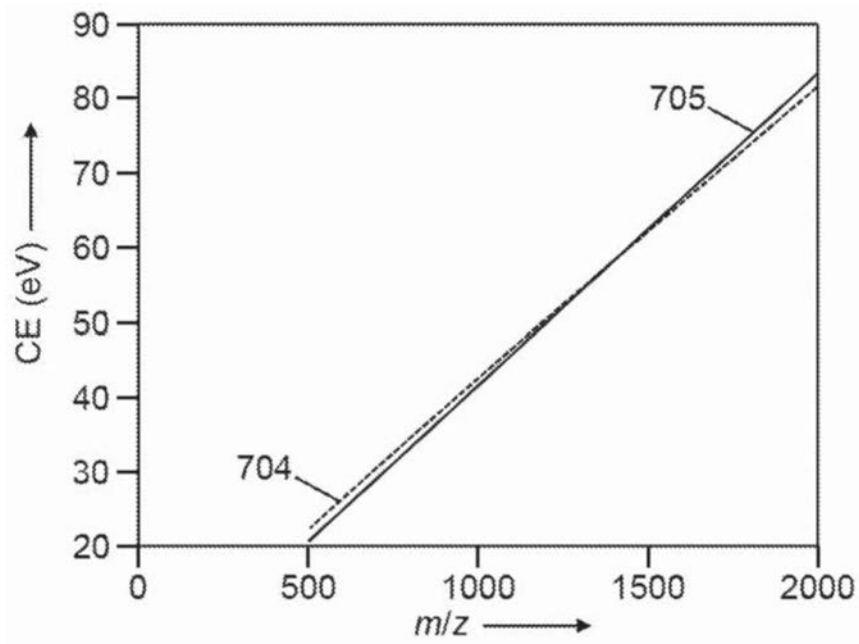


图9B

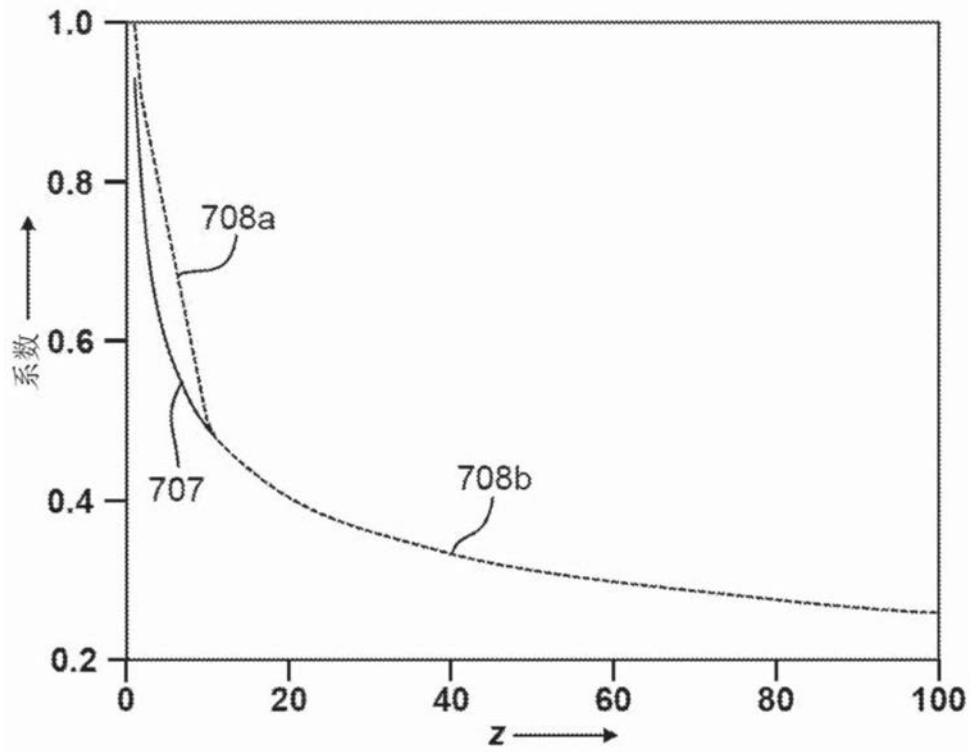


图10

z	$f(z)$	z	$f(z)$	z	$f(z)$	z	$f(z)$	z	$f(z)$
1	1.0000	21	0.3992	41	0.3312	61	0.2964	81	0.2738
2	0.9000	22	0.3941	42	0.3290	62	0.2951	82	0.2729
3	0.8500	23	0.3892	43	0.3268	63	0.2937	83	0.2720
4	0.8000	24	0.3846	44	0.3247	64	0.2925	84	0.2711
5	0.7500	25	0.3803	45	0.3227	65	0.2912	85	0.2702
6	0.7000	26	0.3761	46	0.3207	66	0.2900	86	0.2693
7	0.6500	27	0.3722	47	0.3188	67	0.2887	87	0.2684
8	0.6000	28	0.3684	48	0.3169	68	0.2875	88	0.2676
9	0.5500	29	0.3648	49	0.3151	69	0.2864	89	0.2667
10	0.5000	30	0.3614	50	0.3133	70	0.2852	90	0.2659
11	0.4783	31	0.3581	51	0.3116	71	0.2841	91	0.2651
12	0.4668	32	0.3549	52	0.3099	72	0.2830	92	0.2643
13	0.4565	33	0.3519	53	0.3083	73	0.2819	93	0.2635
14	0.4471	34	0.349	54	0.3067	74	0.2808	94	0.2627
15	0.4386	35	0.3461	55	0.3051	75	0.2798	95	0.2619
16	0.4307	36	0.3434	56	0.3036	76	0.2787	96	0.2611
17	0.4235	37	0.3408	57	0.3021	77	0.2777	97	0.2604
18	0.4168	38	0.3383	58	0.3006	78	0.2767	98	0.2596
19	0.4105	39	0.3358	59	0.2992	79	0.2758	99	0.2589
20	0.4047	40	0.3335	60	0.2978	80	0.2748	100	0.2582

图11

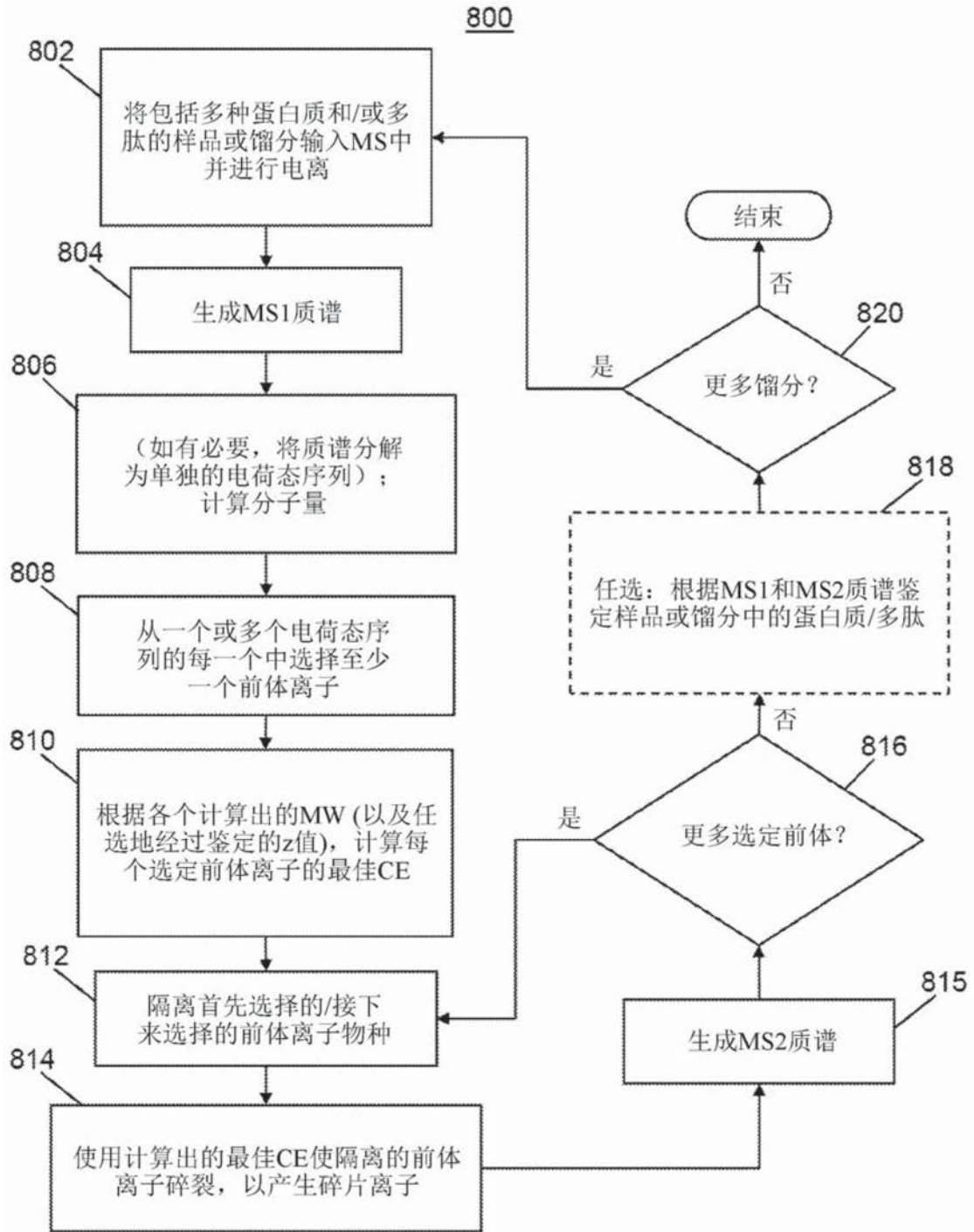


图12

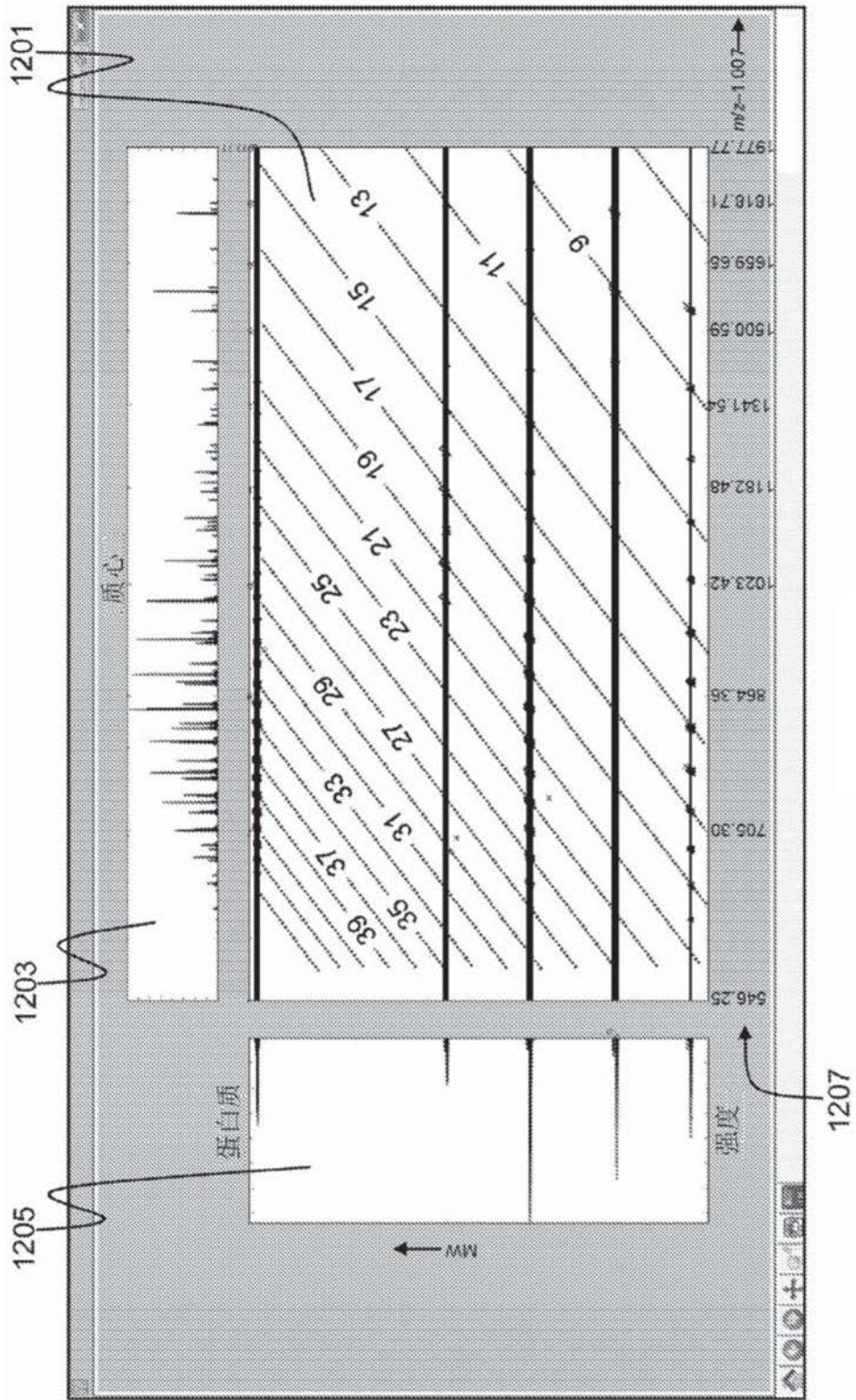


图13A

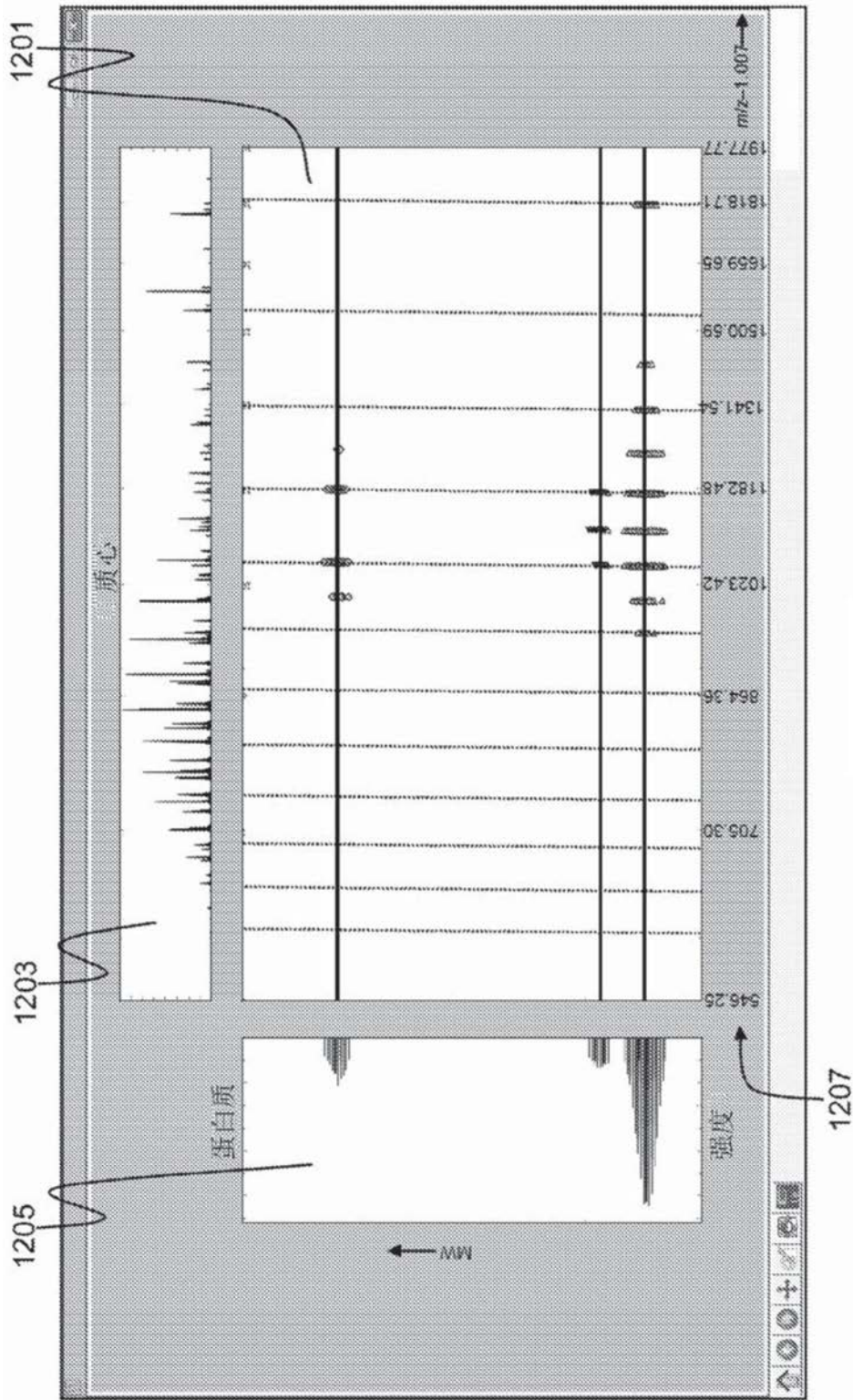
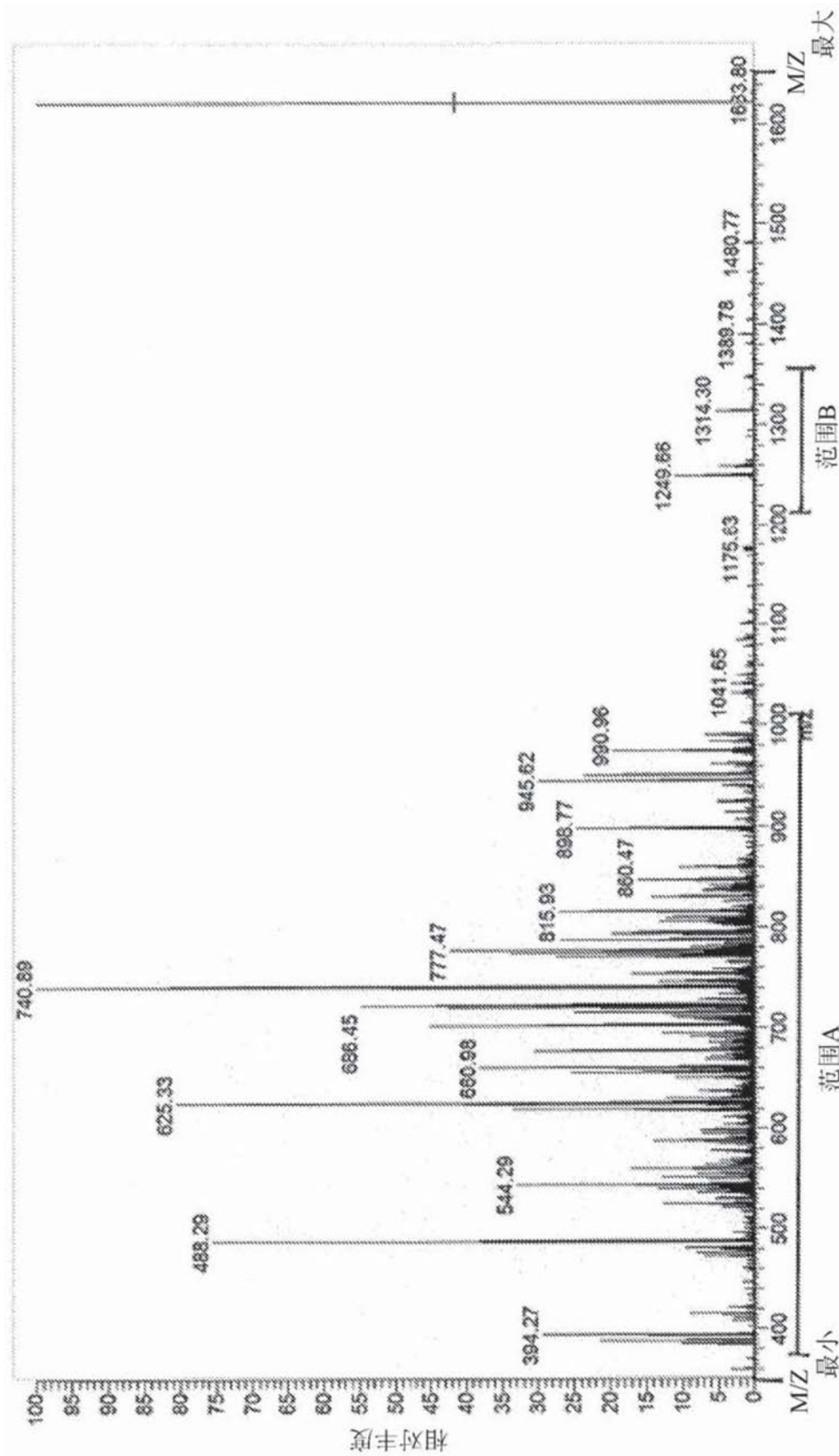


图13B



图A1