

多肽及蛋白质质谱分析新进展

陈绍农*

(中国科学院兰州化学物理研究所 兰州 730000)

潘远江 陈耀祖

(应用有机化学国家重点实验室 兰州大学 兰州 730000)

[摘要]多肽和蛋白质的结构分析常用化学及生物化学降解法,近年来运用质谱法日益增多,本文简要综述质谱新技术,如快原子轰击电离、串联质谱、电喷雾质谱等应用于多肽和蛋白质分子量测定及序列分析的新进展。

关键词:多肽 蛋白质 质谱

多肽及蛋白质是生命科学中极为重要的研究对象。关于它们的分析研究,一直是化学家及生物化学家极为关注的问题。其研究主要包括分子量的测定、氨基酸的鉴定、氨基酸序列分析以及立体化学分析。随着生命科学的进展,仪器分析手段的更新,尤其是近年来质谱分析在这方面应用的崛起,使这一领域的发展突飞猛进^[1,2],本文将扼要综述质谱分析应用于此的新进展。

质谱分析用于生物活性分子的研究具有如下优点:灵敏度极高,能为亚微克级试样提供信息,能最有效地与色谱联用,适用于复杂体系中痕量物质的鉴定或结构测定。由于质谱法所具有的这些特点使它成为解决生命科学中分析研究的强有力手段,发展前景诱人。

早在1968年,Lederer小组首先对衍生化的多肽进行了电子轰击质谱的研究^[3,4],此后10余年内,尽管质谱应用于多肽和蛋白质的分析研究时有报道^[5,6],但是,这方面的工作真正取得实质性突破是在1981年Barber创立了快原子轰击(Fast Atom Bombardment)质谱(FABMS)^[7,8]和1984年Yamashita和Fenn^[9]等创立了电喷雾(Electron Spray)质谱(ESIMS)之后,随着这些技术手段的广泛应用,多肽与蛋白质分析近年来取得了相当大的成功。

1 快原子轰击质谱(FABMS)用于多肽和蛋白质分析

1.1 FABMS

快原子轰击是1种软电离技术,是将快速惰性原子(如氩)射击存在于底物(液)中样品,使样品离子溅出进入质谱分析器,这种软电离技术非常适于高极性、热不稳定的化合

1994年1月4日收

* 通讯联系人

物的分析,因而 FABMS 曾广泛应用于多肽和蛋白质的分析研究。

肽的 FABMS 中通常有较高丰度的 $(M+H)^+$ 离子,有时也会出现 $(M+Na)^+$ 或 $(M+K)^+$ 甚至更高的离子系列。在实验中,液体基质(底物)的选择是一个较为重要的问题,甘油、硫代甘油、间硝基苄醇以及溶解于甲醇中的三硫赤鲜醇(DTE)和二硫苏糖醇(DTT)的混合物都是几类常用的底物。这些底物的特点是蒸汽压低,在高真空的离子源中能在几分钟甚至更长的时间内仍以液态形式存在。在 FABMS 中,多肽被检测的总灵敏度往往受到多肽在所用基质中的溶解程度以及基质的表面活性等因素的影响,有关研究表明^[10],在这些基质中增加适量的三氟乙酸,往往能使多肽 $(M+H)^+$ 峰大大增强。在早期的实验中,由于受到磁质谱质量范围的限制,测定高质量的多肽分子量时,必须降低电压,因而导致分辨率降低,随着仪器的不断改进,现在已经可以将较高质量的离子峰与其同位素峰分辨。

在肽的 FABMS 中,出现的序列信息碎片主要是通过酰胺键断裂而形成的,这种断裂常伴随着电荷保持在 NH_2 -端基或 $COOH$ -端基上^[11,12],这些系列裂解对质谱的相对贡献是由复杂的气相离子化学所控制的。FABMS 一般能给出肽的分子量,然而在确定多肽序列分析方面,单凭 FABMS 数据还难以达到令人满意的结果,这是因为所必需的碎片系列常不完全^[13,14]。尽管有很多文献报道提出可以仅用 FABMS 数据就能解决肽的序列问题,其实,在几乎所有这样的实例中所研究的多肽都是已知的,因而所需碎片的质量可以通过计算后从出现在与基质相关的化学“噪音”中找出来。如果是一个未知的多肽,只靠 FABMS 数据,有把握建立起的只可能是含有 5~11 个氨基酸的寡肽的序列。因此,在确定多肽的结构时,FABMS 数据必须结合其它的物理化学信息(如 Edman 降解分析结果)。另外,化学衍生物质谱分析可能有助于序列测定。例如,比较一个乙酰化的多肽与未乙酰化的多肽的 FABMS 数据,在乙酰化多肽的裂解中,来自 $-NH_2$ 端基的碎片峰都将质量位移 40Da,借此可以识别肽的氨端^[14]。此外,乙酰化还能用来区分具有相同分子量(质量数)的赖氨酸(MW146)和谷酰胺酸(MW146),因为谷酰胺酸含一个酰胺支链。根据这一原理,曾借反应质谱法鉴定了多肽^[15]。同样,一个多肽的羧基数目可通过甲酯化而加以确认^[16],这类酯化可借以区分羧端和氨端。多肽的二硫键则可以用在碱(三乙胺或 NH_4OH)中还原使质谱图中质量位移 2Da 而加以归属。

1.2 FABMS/MS(串联质谱)

FABMS/MS 能提供多肽的较为详细的分子结构信息。在串联质谱中,MS-1 选择地将肽的 $(M+H)^+$ 离子输送到位于第二无场区的碰撞室内,经惰性气体碰撞活化而裂解,这样可以滤掉原样品中的杂质、基质以及其它背景的信号,从而得到简化的子离子质谱,后者由检测记录,得到所谓的碰撞裂解质谱(CDMS),CDMS 谱图中,一般可获得肽中的氨基酸的肽键序列,而当其分子量超过 2500Da 的肽键,则用一般 FABMS/MS 测定其序列就较困难了。但是,最近曾有人利用 1 台 YG2AB-SE4F 串联双聚焦质谱仪测到质量数高达 4000Da 的肽序列^[17]。

一般认为双聚焦质谱仪都适用于串联质谱的 MS-1 和 MS-2,这是由于它们能对质量数达几千 Dalton 的多肽母离子和子离子提供一致的分辩能力,其次双聚焦质谱仪产生的是带有转化能高达 2~10keV 的离子,这些离子允许高能量碰撞活化,而通过高能量范

围的碰撞所得到的肽的子离子谱具有较好的重复性^[16]。

串联质谱应用于肽及蛋白质的序列分析虽然发展很快,但现在仍处在初级阶段,到目前为止,所报道的应用串联质谱测定多肽序列的大部分实例仍主要是复核已知序列的多肽。一个完全未知的多肽的子离子质谱解析起来仍有相当大的困难,研究的实例也为数不多。主要问题是很难确定产生一套完整而又有可预测性的序列离子的条件,这方面尚待进一步研究。

1.3 绘制肽图

由 FABMS 确定肽的分子量并用串联质谱进行序列分析成为蛋白质测定中 Edman 降解法的理想的补充手段,这整个实践过程一般称之为“绘制肽图”。所谓的绘制肽图就是先将蛋白质进行酶解,将酶解产物经 FABMS/MS 分析,得出各组分的分子量和序列,根据酶解的选择性,以及这些质谱数据与已知结构的肽对照,从而确定整个蛋白质的结构。

在这一过程中,酶解产物通常采取反相 HPLC 先行分离,这样会使问题简化且所得结果更加可靠。蛋白质在酶解以前有时还进行烷化或还原以增加其在质谱中的碎裂程度。将酶解反应条件、肽信号以及预计的蛋白质序列输入计算机就能列出一个表格,其中给出各肽段的序列以及在蛋白质中的位置,70%~95%的氨基酸序列可借此得到进一步证实。然而要得到 100%的序列信息是比较困难的,因为需要 2 种或多种酶解方法。天冬氨酸内切酶和 CNBr 是 2 种常用的切割试剂,能将蛋白质切割成质量数为 10kDa 或更小的肽段。

当得到的子离子质谱符合不止 1 个可能的序列或序列信息不完全时,则必须对自动的 Edman 降解产物进行分离与纯化^[17]。

FABMS 绘制肽图途径不能证实蛋白质中二硫键的位置^[20]。其方法是在尽可能避免二硫键的还原和重组的条件下水解蛋白质,比如在酸性条件下,用骨蛋白酶进行水解,或用 CNBr 切割,以获得每 1 个二硫键合的肽段仅含 1 个—S—S 桥。将这类样品在 DTT/DTE 或硫代甘油基质中加碱还原,比较还原前后多肽的质谱,就能归属含二硫键的半胱氨酸肽段。

2 电喷雾质谱(ESIMS)用于多肽与蛋白质分析

电喷雾技术是近几年来发展最快的 1 种新的电离技术,虽然它起源于早期 Bose 的研究,然而直到 1984 年这种电离方式才同质谱结合起来^[21]。当 Karas M 研究小组^[22]首先用电喷雾作为生物大分子多肽及蛋白质等的电离手段以后,ESIMS 的实践才蓬勃地发展起来。

电喷雾中所形成的多电荷离子可以直接用来准确、高灵敏度地确定多肽与蛋白质的分子量。电喷雾所产生的多电荷离子特别适用于串联质谱分析,这是因为多电荷离子容易碎裂,使碰撞活化灵敏度提高。

2.1 PESIMS(正离子电喷雾质谱)

大多数蛋白质的 ESIMS 的 1 个特点是平均电荷状态随分子量的增加而呈近似的线性关系。在溶液中,荷电位置的数目是影响 ESIMS 中所能观察到的最大荷电程度的主要因素,对蛋白质而言(制备成水溶液, pH<4),被检测到的最大荷电状态与碱性的氨基酸肽段数目加上—NH₂ 端也成近似的线性关系^[23]。在酸性条件下,大多数碱基肽段在溶液

中质子化。而酸性肽段在溶液的 pH 大于 5 时会失去质子。当 pH 大于 5 时往往能降低所观察到的正离子 ESI 电荷分布状态。例如蛋氨酸人体生长荷尔蒙, 含有 11 个精氨酸、3 个组氨酸、9 个赖氨酸和 1 个 $-NH_2$ 端基的蛋氨酸肽段, 在溶液中有 24 个碱基能作为质子化的位置, 于是在 ESIMS 中, 能观察到的分子离子最高荷电系列为 $(M+24H)^{24+}$ ($m/z=928$)。到目前为止, 已成功地检测到的最大的蛋白质的分子量超过了 190kDa^[24]。

许多蛋白质分子中含有影响蛋白质高级结构的二硫桥, 如牛奶的乳白蛋白由 17 碱基和 4 个半胱氨酸—半胱氨酸二硫桥所构成。在 ESI 质谱中只能看到最多带 13 个正电荷的离子状态, 当用 DTT 将二硫键还原后, 分子离子最高荷电状态达到 19 个, 这样二硫键的数目就可以由分子离子还原前后质谱变化来确定^[25]。具体就 α -乳白蛋白而言, 8 个质量单位的增加对应于 8 个半胱氨酸肽段(4 个二硫键), 在乳白蛋白的质谱中, 实际的质子化程度要比通过累加蛋白质中碱性肽段而估计的多一些。这主要是肽中诸如谷氨酰胺和赖氨酸具有相似的气相质子亲和能而往往未被考虑的缘故, 而且目前尚难肯定单个氨基酸在气相时的质子亲和能与所观察的质子化程度的关系。

以非共价相结合的各蛋白质亚单位一般显示出各亚单位的多电荷离子的特征, 诸如人体血红蛋白是由 2 个 α 链和 2 个 β 链构成的四聚体, 其 ESIMS 显示出 α 与 β 链单位(α : Mr15,126; β : Mr15,867)的 2 种多电荷离子^[26], 当然也有例外。

高度选择性蛋白酶的应用在蛋白质结构解析中是一种很普通的方法(见 1.3), FABMS 和 FABMS/MS 已成为分析复杂的酶解产物的常规工具。然而, 由于有差异的表面活性效应的影响, FABMS 很难对每个蛋白质水解产物进行确定, 相比之下 ESIMS 对于酶解产物无论是亲水性的还是疏水性的组分在 1 个较宽的浓度范围内都能得到较好的结果^[27]。

在迅速发展的生物技术产业中, 蛋白质纯度的确定对于建立产品的质量、效能及稳定性方面都起着重要的作用, 而对蛋白质纯度的定性与定量检测都是分析化学家面临的一个新课题。最近几年来, ESIMS 对这方面进行了有益的尝试。

最近, ESIMS 基于分子离子的峰宽就血清蛋白的相对纯度进行了定性的评估^[28]。分子离子的峰宽一般是指最大峰的半宽度, 往往是在有足够的碰撞活化而除去残留的溶剂和非共价结合的分子系列的条件下测定的。这样的条件可使分子离子峰适当简化, 而这样简化的峰的形状往往能提供纯度信息。这类实验有时还可纠正氨基酸的序列, 如 Hirayama 等^[29]就对发表的奶牛血清蛋白的氨基酸序列进行了考察与纠正。另外, Witkoska 和合作者^[30]曾用 ESIMS 确定了他们所制备的硫酯酶(thioesterase I)的纯度。

2.2 NESIMS(负离子电喷雾质谱)

一般而言, 多肽与蛋白质的 ESIMS 分析总是以正离子方式而进行, 这是因为较多的碱性氨基酸肽段(Arg, Lys 等和 NH_2 端)具有足够大的离解常数以确保蛋白质在酸性溶液中最可能的多重质子化。

负离子 ESIMS 一般用于分析 1 类拥有酸性官能团(诸如羧基、磷酸基等)的小分子^[23,31], 胰岛素 A 链的 ESIMS 中, 观察到了带 5 个负电荷的脱质子化的负离子^[29]。

在碱性水溶液中, 通过电喷雾电离可以产生脱质子化的多电荷分子(负电荷), 这种溶液中碱性肽段一般不带电荷。马的肌红蛋白中含有 13 个 Glu 物 7 个 Asp 肽段, 在 1%

NH_4OH 水溶液中 ESIMS 能出现达到近 $(M-19\text{H})^{19-}$ 电荷状态的球形多电荷状态分布^[32]。

pH 值的变化以及阳离子取代都会对负离子 ESIMS 产生一定的影响。有人对猪的胃蛋白酶分别在 1% (0.15M) NH_4OH 和 $5 \times 10^{-2}\text{M}$ 、 $5 \times 10^{-3}\text{M}$ 、 $5 \times 10^{-4}\text{M}$ NaOH 溶液中的 ESIMS 进行了研究,在 $5 \times 10^{-4}\text{M}$ NaOH 溶中 ($\text{pH} \sim 10.7$),分子离子的信号强度大大减弱,每个峰的 m/z 出现了拖尾(tailing)。这是由于 Na^+ 与 H^+ 的互换所致。另外,与在 1% NH_4OH 溶液中的 ESIMS 相比,被观察的荷电程度也减少了,这或许是由于稍微低的 pH 值所致。而在 $5 \times 10^{-3}\text{M}$ NaOH 溶液中,ESIMS 中出现了较 1% NH_4OH 溶液条件下更高的荷电状态(前者 51^- ,后者 42^-)^[33]。半胱氨酸和酪氨酸肽段的酚羟基出现脱质子化,这是因为它们的 PKa 值分别在 $8.5 \sim 8.9$ 和 $9.6 \sim 10.0$ 之间^[33],因而胃蛋白酶中的 16 个酪氨酸肽段中的一些都可以脱质子化。比较上述几种条件下的质谱可以看出,在较低的 NaOH 浓度下,能观察到更大程度的 Na 加合物。对每一个多重电荷状态而言,都能观察到无法分辨的 Na 加合物的分布而且低电荷状态的离子,比较高的电荷离子含有更多的 Na 取代物。在较大的 NaOH 浓度条件下 ($5 \times 10^{-2}\text{M}$),信号纯度也严重的减弱了,除尽可能更大程度的 Na 取代外,蛋白质可能被降解了。有趣的是,在适中浓度的 NaOH 溶液中,较高的电荷状态 ($>45^-$) 表明仅有少量的 Na 取代。 Na 取代的程度随着降低的电荷而增加,这些结果表明, Na 的取代程度受 pH 值和蛋白质结构的制约。

重组蛋白 A 是有 409 个肽段的蛋白质,其中含有未报道的半胱氨酸一半胱氨酸链,从生产者所提供的序列可知,它含有 63 个碱性肽段和 71 个酸性肽段 ($M_{45,351}$)。正离子 ESIMS 显示超过 65^+ 的多电荷离子,而负离子 ESIMS 显示了超过 52^- 的多电荷离子。这两种电离方法都表明同一分子量 $44,652 \pm 4$ ^[33]。

在二硫键被还原后的羊的血蛋白的负离子 ESIMS 中观察到了大于 52^- 的电荷状态。其质谱表明该蛋白含有 100 个 Glu 和 Asp 肽段。而对于未涉及二硫键的天然羊的白蛋白在 m/z 为 2000 以下没有检测到脱质子化的多重电荷的分子离子。这些结果同用正离子 ESIMS 所得到的结果一致,而区别主要存在于所荷电荷的状态不同^[34]。

综上所述,负离子 ESIMS 对于某些特殊的小分子肽与蛋白质也能达到检测分析的效果,而正离子 ESIMS 和负离子 ESIMS 的联合运用扩展了这些方法的应用,能对溶液中蛋白质结构的分析提供新的定性依据。

多肽与蛋白质的分析是生命科学中的 1 个重要的课题,质谱已成为研究该课题非常重要的工具,除了 FABMS 和 ESIMS 外,其它质谱法同样曾发挥过不可忽视的作用,比如离子喷雾质谱、等离子体解吸质谱以及激光解析质谱,这些软电离技术各具特色,但它们共同的特征是能提供较强的准分子离子峰,可以确定分子量,一般情况下,它们不能提供测定序列的碎片离子峰。

参 考 文 献

- 1 Biemann K. *Annu Rev Biochem*, 1992, 61, 977
- 2 Carr S A. *Adv Drug Deliv. Rev* 1990, 4, 113
- 3 Vilkas E, Lederer E. *Tetrahedron Lett*, 1968, 3089
- 4 Morris H R. *FEBS Lett*, 1972, 44, 257
- 5 Biemann K. in "Biochemical Applications of Mass Spectrometry." Valler G R Ed. Wiley, New York, 1972, 45—428
- 6 Biemann K. in "Biochemical Applications of Mass Spectrometry, First Supplementary Volume". Waller G R, Demer O C Eds. Wiley, New York, 1980 469—525
- 7 Barber M. Bordoli R S *et al.* *J Chem, Soc Chem Comm*, 1981, 325—327
- 8 Aberth W *et al.* *Anal Chem*, 1982, 54, 2029—2034
- 9 Yamashita M, Fenn J B. *J Phys Chem*, 1984, 88, 4451—4459
- 10 Depauw E. *Mass Spectrom, Rev*, 1986, 5, 191—212
- 11 Bradley C V *et al.* *Bio Chem J*, 1982, 201, 105—117
- 12 Roepstorff P *et al.* *Biomed Mass Spectrom*, 1985, 12, 181—189
- 13 Biemann K, Martin S A. *Mass spectrom Rev*, 6, 1—76
- 14 William D H *et al.* *Biochem J*, 1982, 201, 105—117
- 15 杨厚俊, 陈耀祖等. *Chin J Chem*, 1993, 11(6), 546
- 16 Morris H R *et al.* *Biochem Biophys Res Comm*, 1981, 101, 523
- 17 Carr S A *et al.* *Proceedings of the 35th ASMS Conference on Mass Spectrometry and Allied Topics*, Denver, CO. American Society For Mass Spectrometry, East Lansing, 1987, MI. 830—831
- 18 Gaskell S J *et al.* *Proceedings of the 36th ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, CA, June 5—7, 1988, 27, 4214.
- 19 Anderegg R J *et al.* *Biochem*, 1988, 27, 4214
- 20 Morris H R *et al.* *Biochem Biophys Res Commun*, 1985, 126, 1122
- 21 Yamashita M, Fenn J B. *J Phys Chem*, 1984, 88, 4451
- 22 Karas M, Killenkamp F. *Anal Chem*, 1988, 60, 2299
- 23 Smithe R D *et al.* *Anal Chem*, 1990, 62, 882
- 24 Loo T A *et al.* *Anal Chem*, 1990, 62, 693
- 25 Feng R *et al.* *Anal Chem*, 1993, 65, 645
- 26 Bojesen G. *J Am Chem Soc*, 1987, 5557
- 27 Katla V *et al.* *J Am Chem Soc*, 1991, 113, 8534
- 28 Chowdhury S K *et al.* *Biochem Biophys Res Commun*, 1990, 167, 686
- 29 Hirayama K *et al.* *Biochem Biophys Res. Commun*, 1990, 173, 139
- 30 Witkowska H E *et al.* *J Biochem*, 1990, 265, 5662
- 31 Stutts J T *et al.* *Rapid Commun Mass spectrom*, 1991, 5, 359
- 32 Loo J A *et al.* *Anal Chem*, in press
- 33 Cantoor C R *et al.* *Biophysical Chemistry*, W H Freeman. San Francisco, CA. 1980
- 34 Loo T *et al.* *Anal Chem*, 1991, 63, 2488

Recent Advances in Mass Spectrometry of Polypeptides and Proteins

Chen Shaonong*

(Lanzhou Institute of Chemical Physics, Chinese Academy
of Sciences, Lanzhou 730000, China)

Pan Yuanjiang, Chen Yaozu

(State Key Laboratory of Applied Organic Chemistry,
Lanzhou University, Lanzhou 730000, China)

Received 1994-01-04

Abstract

Structural analysis of polypeptides and proteins has been performed by chemical and biochemical degradation methods. Recently the use of mass spectrometry for this purpose has increased considerably. In present paper, the recent advances in application of new mass spectrometric techniques such as fast atom bombardment ionization, tandem mass spectrometry, electrospray mass spectrometry etc to molecular weight determination and sequence analysis of polypeptides and proteins have been briefly reviewed.

Key Words: polypeptides, proteins, mass spectrometry

* To whom the correspondence should be addressed