

## 一种能预测多肽和多肽类似物的质谱碎片质量的程序

朱传钧<sup>1</sup>, 胡长风<sup>1,2</sup>, 崔育新<sup>1,2</sup>

(1. 北京大学药学院, 天然药物与仿生药物国家重点实验室, 北京 100191;  
2. 北京大学医药卫生分析中心, 北京 100191)

**摘要:** 根据已知算法编写了一种预测多肽和多肽类似物质谱碎片质量的程序。采用 C++ 语言编写, 具有良好的移植性, 可与 EXCEL 等表格处理软件联用。它不但可处理天然多肽, 也可以处理含有由用户定义的非天然氨基酸残基的多肽类似物。对多个多肽类似物的质谱碎片质量进行了预测, 与实测结果比较后证明预测全部成功, 说明本程序的结果有效、可靠。使用该程序对多肽和多肽类似物的质谱碎片质量进行预测, 可以大幅简化该类化合物质谱解析过程。

**关键词:** 多肽; 质谱; 预测; 程序

中图分类号: O 657.63 文献标识码: A 文章编号: 1004-2997(2008)06-327-05

## A Computer Program for Predicting MS Fragments' Mass of Peptides or Peptide Analogs

ZHU Chuan-jun<sup>1</sup>, HU Chang-feng<sup>1,2</sup>, CUI Yu-xin<sup>1,2</sup>

(1. *State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Science, Peking University, Beijing 100191, China;*  
2. *Medical and Health Analysis Centre, Peking University, Beijing 100191, China*)

**Abstract:** A computer program was designed for predicting MS fragments' mass of peptides or peptide analogs according to known algorithm. It was written by C++ with high transplantability. The output-file of this program can be processed by EXCEL or other electric-table easily. Not only natural occurring peptides but also peptide analogs with unnatural amino acid residue can be processed. MS fragments' mass of some peptide analogs were predicated by this program. The results show that this program is efficient and dependable. It is convenient and flexible to analysis MS fragments of peptides or peptide analogs supplied by this program.

**Key words:** peptide; mass spectrometry; predicting; program

多肽的测序方法多种多样,质谱测序法是其较为简便快捷的一种<sup>[1-2]</sup>。由于多肽是一种极性化合物,在质谱测试中,传统的 EI 电离源无法使它电离,所以必须采用“软电离”技术,例如 ESI 源<sup>[3-5]</sup>或 MALDI 源<sup>[6]</sup>。由于多肽结构上的特点,质谱裂解规律性较强,但是解析起来较为枯燥。为了解决这一问题,质谱学工作者和质谱仪器生产商均编写过多肽质谱结果预测软件<sup>[7-8]</sup>。但是,这些软件或只能处理天然多肽、或需要特定的软硬件环境、或只能与某种特定型号的质谱仪配套使用,这给使用者带来很大不便。本工作拟编写一种能预测多肽和多肽类似物的质谱碎片质量的程序。采用 C++ 语言编写,具有良好的移植性,可与 EXCEL 等表格处理软件联用。它不但可处理天然多肽,也可以处理含有由用户定义的非天然氨基酸残基的多肽类似物。

## 1 程序部分

### 1.1 算法

目前世界上有 2 种多肽质谱碎片质量命名系统,本工作选取 Roepfstorff 命名法<sup>[9]</sup>,示于图 1,即采用英文字母和数字组合的方式为主链碎

片命名。N 端碎片离子用字母表当中开头的几个字母(a, b, c)表示;同时 C 端碎片离子用字母表结尾的几个字母(x, y, z)表示。存在氢转移的离子,可采用重音符号(')明确的标出,也可以不增加任何标记隐式表示。数字脚标表示主链断裂的位置。对于 N 端离子,从 N 端开始计数,在第 1 个氨基酸残基和第 2 个氨基酸残基之间断裂,用数字 1 表示;在第 2 个氨基酸残基和第 3 个氨基酸残基之间断裂,用数字 2 表示,以此类推。对于 C 端离子,从 C 端开始计数,在第 1 个氨基酸残基和第 2 个氨基酸残基之间断裂,用数字 1 表示;在第 2 个氨基酸残基和第 3 个氨基酸残基之间断裂,用数字 2 表示,以此类推。举例来说,对于一个四肽,当连接第二个残基羰基和第三个残基之间的酰氨键断裂时,可产生两个碎片。当电荷保留在 N 端时,称为  $b_2$  离子,同时,包含 C 端的碎片,由于没有电荷,不能被质谱仪检测到。若电荷保留在 C 端,那么将形成  $y_2$  离子,剩下包含 N 端的部分由于没有电荷,是一个中性碎片。

根据上述命名原则,可推导出这 6 类序列特异性离子的质量计算公式<sup>[10]</sup>,示于图 2。

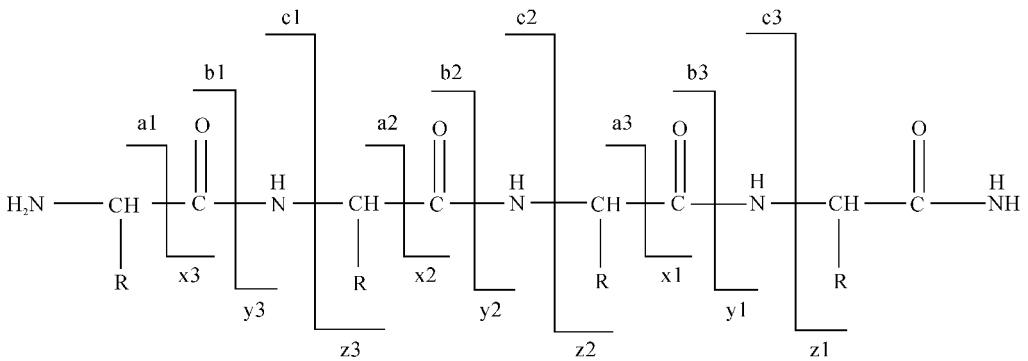


图 1 四肽的质谱骨架碎片命名

Fig. 1 The nomenclature for MS fragment ions of tetra-peptide

$$\begin{aligned}
 a_j &= [N-term] + \sum_1^j aa_1 - [CO] \\
 b_j &= [N-term] + \sum_1^j aa_1 \\
 c_j &= [N-term] + \sum_1^j aa_1 - [NH_2] \\
 x_j &= [C-term] + \sum_{n-j+1}^n aa_1 + [CO] \\
 y_j &= [C-term] + \sum_{n-j+1}^n aa_1 \\
 z_j &= [C-term] + \sum_{n-j+1}^n aa_1 - [NH]
 \end{aligned}$$

图 2 单质子化多肽碎片离子质量计算公式

Fig. 2 Mathematical expressions for the calculation of the mass of singly protonated peptide fragment ions

由于 ESI 或 MALDI 技术均属于“软电离”技术,几乎不能产生碎片,所以为了得到碎片信息,就必须使用其他辅助技术使其解离<sup>[11-12]</sup>,其中最为常见的是碰撞诱导解离技术(CID)。CID 的碰撞能量可以高于 1 000 eV,称为高能 CID;也可以在 100 eV 数量级,称为低能 CID。高能 CID 谱中,由于能量较大使构成多肽的任意化学键都可断裂,a、b、c、x、y、z 六种离子和侧链断裂

离子都可能出现;若采用低能 CID,则以 b、y、a 类型离子为主。

当骨架两端均发生裂解,可形成中间离子。若一端发生 y 类型裂解,另一端发生 b 类型裂解,则形成的离子称为  $y_i b_j$  离子,其他中间离子命名类似。

## 1.2 程序

1.2.1 编译环境 操作系统 Windows XP SP2,编程工具 Microsoft Visual C++ 6.0。

1.2.2 程序界面 采用字符界面,示于图 3,分为“参数设置区”、“非天然氨基酸定义区”、“系统操作区”、“命令区”。

```

----- 参数设置 -----
<S>多肽序列:未定义
<N>N端基团:H1
<C>C端基团:H2N1
<E>检索误差:5ppm
<I>离子选择:-ab--y--im
----- 非天然氨基酸 -----
<B>未定义
<J>未定义
<O>未定义
<U>未定义
<X>未定义
<Z>未定义
----- 系统操作 -----
<R>报告结果
<F>写入文件
<P>谱峰检索
<Q>退出系统
----- 命令区 -----
请选择:_
  
```

图 3 程序界面

Fig. 3 Program interface

## 1.2.3 程序组成

main() 函数,为系统默认的程序入口,总领整个程序;

param\_list() 函数,屏幕显示参数,即程序界面;

init\_residue() 函数,初始化氨基酸残基信息;

report() 函数,用于输出结果到屏幕;

writefile() 函数,用于将结果写入文件,存储为 CSV 文件,可供 EXCEL 打开;

search\_peak() 函数,用于寻找与目标结果匹配的质谱峰;

calculate() 函数,采用前述算法,计算离子的质量;

check() 函数,用于检查非法输入;

residue 类,用于输入、存储、输出残基的名称、元素组成、质量;

ions 类,用于输入、存储、输出离子的名称、元素组成、质量;

param 类,用于输入、存储、输出各参数;

weight 结构体,用于存储各种可能用到的原子质量和电子质量,采用双精度数据类型。

## 1.3 预测举例

某多肽类似物(编号 3-1)的结构示于图 4。

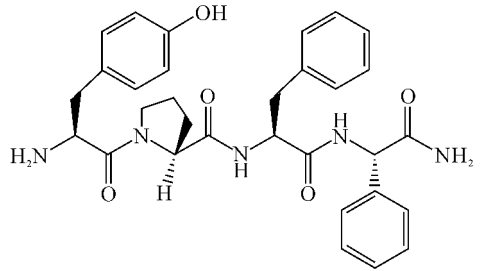


图 4 多肽类似物 3-1 的结构

Fig. 4 Structure of peptide analog 3-1

该化合物第 4 位为非天然氨基酸残基,命名为“氨基酸 B”,元素组成为  $C_8H_7NO$ 。1、2、3 位分别是酪氨酸(Y)、脯氨酸(P)、苯丙氨酸(F),N 端暴露,C 端为  $NH_2$ ,将这些信息输入程序。多肽序列:YPFB;N 端基团: $H_1$ ;C 端基团: $H_2N_1$ ;检索误差:5 ppm;离子选择:ab--y-im;非天然氨基酸定义<B>: $C_8H_7N_1O_1$ 。根据需要可以选择“报告结果”,输出结果到屏幕;“写入文件”输出结果到文件,示于图 5。“谱峰检索”可检索目标峰是否为该化合物所产生的。

	A	B	C
1	Name	Formula	Weight
2	M+H	C31H36N5O5	558. 271 1
3	b1	C9H10N1O2	164. 070 6
4	b2	C14H17N2O3	261. 123 4
5	b3	C23H26N3O4	408. 191 8
6	b4	C31H33N4O5	541. 244 5
7	y1	C8H11N2O1	151. 086 6
8	y2	C17H20N3O2	298. 155 0
9	y3	C22H27N4O3	395. 207 8

图 5 输出的文件(部分)

Fig. 5 Part of the output-file

## 2 验证实验部分

### 2.1 仪器与试剂

Bruker APEX IV 傅立叶变换 (FT-ICR) 质谱仪; 德国 Bruker 公司产品。

乙酸、甲醇 (色谱纯); 天津化工厂产品; 纯净水; 购自娃哈哈集团; Tuning Mix; 德国 Bruker 公司提供。

样品 (化合物 3-1) 由兰州大学生命科学学院生物化学与分子生物学研究所提供, 经液相色谱鉴定, 纯度大于 95%。

### 2.2 实验方法

采用傅立叶变换质谱仪, 配有外部 ESI 源 (Analytica, Branford Inc. USA), 磁场强度 7.0

T。将多肽类似物 3-1 样品溶解于  $V(\text{甲醇}) : V(\text{水}) : V(\text{甲酸}) = 49 : 49 : 2$  的混合溶剂中, 浓度约为  $0.3 \text{ g} \cdot \text{L}^{-1}$ ; 溶液从针状注射器喷雾, 注入速率为  $60 \mu\text{L} \cdot \text{h}^{-1}$ ; 质量扫描范围  $m/z$  32~600。串联质谱实验中, 被选择的离子  $[M+H]^+$  与氩气碰撞产生碎片离子, 外标采用 Tuning Mix 校正。

## 3 结果与讨论

### 3.1 预测结果与实测结果比较

化合物 3-1 的质谱碎片质量的预测结果和实测结果列于表 1。

表 1 多肽类似物 3-1 的 ESI FT ICR MS/MS 结果

Table 1 ESI FT ICR MS/MS fragment ions of peptide analog 3-1

离子名称	元素组成	预测质量 ( $m/z$ )	实测质量 ( $m/z$ )	误差/ppm
M+H	$\text{C}_{31}\text{H}_{36}\text{N}_5\text{O}_5$	558.271 1	558.270 7	-0.7
$b_2$	$\text{C}_{14}\text{H}_{17}\text{N}_2\text{O}_3$	261.123 4	261.122 6	-3.1
$b_3$	$\text{C}_{23}\text{H}_{26}\text{N}_3\text{O}_4$	408.191 8	408.191 5	-0.7
$a_1$	$\text{C}_8\text{H}_{10}\text{NO}$	136.075 7	136.076 0	2.2
$a_2$	$\text{C}_{13}\text{H}_{17}\text{N}_2\text{O}_2$	233.128 5	233.128 2	-1.3
$a_4$	$\text{C}_{30}\text{H}_{33}\text{N}_4\text{O}_4$	513.249 6	513.249 1	-1.0
$y_2$	$\text{C}_{17}\text{H}_{20}\text{N}_3\text{O}_2$	298.155 0	298.154 7	-1.0
$y_3$	$\text{C}_{22}\text{H}_{27}\text{N}_4\text{O}_3$	395.207 8	395.207 9	0.3
$y_1 a_4$	$\text{C}_7\text{H}_8\text{N}$	106.065 1	106.065 0	-0.9
$y_2 a_3$	$\text{C}_8\text{H}_{10}\text{N}$	120.080 8	120.080 7	-0.8
$y_2 a_4$	$\text{C}_{16}\text{H}_{17}\text{N}_2\text{O}$	253.133 5	253.133 9	1.6
$y_2 b_4$	$\text{C}_{17}\text{H}_{17}\text{N}_2\text{O}_2$	281.128 5	281.129 1	2.1
$y_3 a_2$	$\text{C}_4\text{H}_8\text{N}$	70.065 1	70.065 1	0.0
$y_3 a_3$	$\text{C}_{13}\text{H}_{17}\text{N}_2\text{O}$	217.133 5	217.132 6	-4.1
$y_3 b_3$	$\text{C}_{14}\text{H}_{17}\text{N}_2\text{O}_2$	245.128 5	245.128 8	1.2
$y_3 b_4$	$\text{C}_{22}\text{H}_{24}\text{N}_3\text{O}_3$	378.181 2	378.180 9	-0.8

注: 预测到但未观测到的离子省略

凡是主链断裂产生的离子质量都能被正确的预测, 并且得益于傅立叶变换质谱仪的高分辨率、高精度, 使得预测值与观测值相差极小。除此之外, 又进行了 7 个多肽类似物的质谱碎片质量预测, 与实测结果比较后证明预测全部成功, 说明本程序的结果有效、可靠。

### 3.2 本程序优点

3.2.1 移植性好 采用 C++ 语言编写, 具有

良好的移植性。本例运行环境为 Windows 系统, 只要经过简单的处理, 该程序也可运行于 DOS、UNIX、Linux 系统等。无论是个人电脑还是专业的工作站都可以运行本程序, 克服了早年类似功能软件只能运行于特定软硬件系统的不足。

3.2.2 可与电子表格联用 由于 EXCEL 等电子表格软件本身具有强大的索引、排序以及数据

计算功能。在本程序的编写过程中,没有将索引、排序、数据计算功能编入,但是可将数据输出为 EXCEL 等电子表格软件可以处理的 CSV 文档,输出结果由 EXCEL 继续处理,避免手工录入时由于粗心出现的谬误。

**3.2.3 可以处理非天然多肽** 对于药学工作者来说,为了获得良好的生理活性,往往需要设计、合成非天然多肽,例如,掺入非天然氨基酸残基或对端基进行修饰,软件设计者往往没有考虑到这方面的情况,软件只能处理天然多肽,面对非天然多肽无能为力。本程序却可由用户自行添加非天然氨基酸残基,只需选择一个没有被天然氨基酸残基占据的名称,输入元素组成即可。

### 3.3 本程序不足

**3.3.1 界面简陋** 本程序出于可移植性及简化代码的考虑,界面为字符界面,较为简陋,习惯图形界面操作的用户会感到不适应。

**3.3.2 不能处理复杂多肽** 本程序不能处理环肽,用多硫键连接的肽链,由多条肽链构成的蛋白。

## 4 结 论

本程序可以正确的预测多肽和多肽类似物的质谱碎片质量,克服了前人编写的类似功能软件的缺点。经多次实验验证,证明程序的结果有效、可靠。使用该软件对多肽和多肽类似物的质谱碎片质量进行预测,可以大幅地提高该类化合物的质谱解析速度。

### 参考文献:

- [1] MORRIS H R, PANICO M, BARBER M, et al. Fast atom bombardment: a new mass spectrometric method for peptide sequence analysis. [J]. Biochem Biophys Res Commun, 1981, 101(2): 623-631.
- [2] MARTIN S E, SHABANOWITZ J, HUNT D F, et al. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry[J]. Anal Chem, 2000, 72(18): 4 266-4 274.
- [3] Smith R D. Evolution of ESI-MS and fourier transform ion cyclotron resonance for proteomics and other biological applications [J]. Int J of Mass Spectrom, 2000, 200(1/3): 509-544.
- [4] FENN J B, MANN M, MENG C K, et al. Electrospray ionization for mass spectrometry of large biomolecules[J]. Science, 1989, 246 (4 926): 64-71.
- [5] FENN J B, MANN M, MENG C K, et al. Electrospray ionization-principles and practice [J]. Mass Spectrom Rev, 1990, 9(1): 37-70.
- [6] TANAKA K. The origin of macromolecule ionization by laser irradiation (Nobel lecture)[J]. Angew Chem Int Ed Engl, 2003, 42 (33): 3 860-3 870.
- [7] ROZENSKI J, SAMSON I, JANSSEN G, et al. Characterization of modification sites during peptide synthesis using liquid secondary ion/collision-induced dissociation mass spectrometry and a computer program[J]. Organ mass spectrom, 1994, 29(11): 654-658.
- [8] LEE T D, VEMURI S. Macpromass: a computer program to correlate mass spectral data to peptide and protein structures[J]. Biomed environ mass spectrom, 1990, 19(11): 639-645.
- [9] ROEPSTORFF P, FOHLMAN J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. [J]. Biomed Mass Spectrom, 1984, 11(11): 601.
- [10] PAPAYANNOPOULOS I A. The interpretation of collision-induced dissociation tandem mass spectra of peptides [J]. Mass spectrom rev, 1995, 14(1): 49-73.
- [11] WEINMANN W, WIEDEMANN A, EPPINGER B, et al. Screening for drugs in serum by electrospray ionization/collision-induced dissociation and library searching[J]. J Am Soc Mass Spectrom, 1999, 10(10): 1 028-1 037.
- [12] WANG P P, BARTLETT M G. Collision-induced dissociation mass spectra of cocaine, and its metabolites and pyrolysis products[J]. J Mass Spectrom, 1998, 33(10): 961-967.