

# 磺基异硫氰酸苯酯化学辅助方法对 新蛋白质进行从头测序

周新文<sup>1</sup>, 严 钦<sup>2</sup>, 周 珮<sup>2</sup>, 杨芑原<sup>1</sup>

(1. 复旦大学生物医学研究院, 2. 药学院, 上海 200032)

**摘要** 利用基质辅助激光解吸电离-串联飞行时间(MALDI-TOF-TOF)质谱结合磺基异硫氰酸苯酯化学辅助的方法对一种从拟青霉(*Paecilomyces bainier*)分离纯化到的新人参皂苷 Rb1 水解酶的部分肽段进行了从头测序. 共获得了这个新蛋白质 8 条肽段的序列, 一些磺化后信噪比非常低的肽段也获得了比较完整的序列. 同时通过从头测序分析确定了一对甲硫氨酸非氧化和氧化肽段的序列. 结果表明, 磺化后的肽段离子化效率大大增强, 在 PSD(源后裂解)过程中只有肽键断裂产生的 C 端的碎片离子系列( $\gamma$  离子系列)出现在质谱图中, 图谱背景清晰, 信噪比高, 单纯的  $\gamma$  离子系列使得图谱解析变得非常容易. 将这 8 条序列在 NCBI(美国国立生物技术信息中心)数据库中进行 BLAST(蛋白质序列比对工具)检索印证这种  $\beta$ -葡萄糖苷酶是一个新蛋白质, 发现的两条相对保守的序列为进一步研究奠定了基础.

**关键词** 磺基异硫氰酸苯酯; 基质辅助激光解吸电离-串联飞行时间质谱; 新蛋白质; 从头测序

**中图分类号** O657; Q51

**文献标识码** A

**文章编号** 0251-0790(2009)04-0706-06

对于大多数基因组序列未知或者测序不全, 蛋白质序列数据库有限的非模式物种的蛋白质, 将其定义为序列完全未知的新蛋白质<sup>[1]</sup>. 要获得这一类新蛋白质序列, 确定其结构和功能的相互关系, 一般是采用基于生物质谱分析的跨物种鉴定方法<sup>[2]</sup>. 但由于物种之间同类蛋白质序列上的差异, 这种跨物种鉴定方法往往得不到按一般标准衡量的可靠结果. 此外, 基于 N 端或 C 端进行从头测序和相似序列搜索结合的方法, 由于需要的蛋白质量大, 灵敏度和通量都较低, 对于凝胶分离的蛋白质点或者条带, 这种方法获得的数据非常有限.

人参中起主要药效的是人参皂苷<sup>[3,4]</sup>. 人参皂苷 Rg3 因在诱导肿瘤细胞凋亡、抑止肿瘤细胞增殖和肿瘤转移等方面具有很强的抗肿瘤活性而备受关注<sup>[5~9]</sup>. 与已有的抗肿瘤药物相比, 由于其副作用微小且温和, 人参皂苷 Rg3 是一种非常有潜力的抗肿瘤药物. 然而, 由于其在人参中的含量很低(低于 0.003%)而无法大量制备. 另外一种与人参皂苷 Rg3 结构非常相似的人参皂苷 Rb1 在人参中含量较高(1%~3%), 选择合适的酶可将人参皂苷 Rb1 有效地转化为人参皂苷 Rg3 从而间接大量地获得该抗肿瘤药物. 最近从拟青霉(*Paecilomyces bainier*)分离纯化到一种新型的人参皂苷 Rb1 水解酶, 命名为  $\beta$ -葡萄糖苷酶, 这种酶是一种新蛋白质. 前期研究发现, 这个蛋白质含有 3 个分子量大约为 102000 相同的亚基, SDS-PAGE 验证该蛋白质的整体分子量大约为 305000, 酶稳定性和活性实验发现, 该酶在 pH = 3~7 和温度低于 50 °C 时比较稳定, 在 pH 3.5 和温度为 55 °C 时具有最佳酶活性, 它能特异性地将人参皂苷 Rb1 转化为人参皂苷 Rg3<sup>[10]</sup>, 对制备抗肿瘤药物人参皂苷 Rg3 非常有用. 为了提高这种新型  $\beta$ -葡萄糖苷酶作用的效率, 对于蛋白质一级序列的测定是研究其结构与功能相互关系的基础.

本文报道了一种基于磺基异硫氰酸苯酯化学辅助对该新蛋白质进行从头测序的方法. 该方法得到串联谱图只包含简单的  $\gamma$  系列离子, 背景干净, 可得到准确可信的序列. 应用该方法成功测定了  $\beta$ -葡萄糖苷酶 8 条胰酶酶解肽段的序列, 并通过 BLAST 工具将这些序列提交 NCBI 数据库进行同源性蛋白质分析, 为下一步基于已知肽段序列设计引物获得该新蛋白质全长序列, 并进行结构与功能相互关系

收稿日期: 2008-05-21.

基金项目: 国家“九七三”计划(批准号: 2007CB914100)资助.

联系人简介: 杨芑原, 男, 博士, 教授, 博士生导师, 主要从事蛋白质组学研究. E-mail: pyyang@fudan.edu.cn

的研究奠定了基础.

## 1 实验部分

### 1.1 试剂与仪器

分离纯化后的 $\beta$ -葡萄糖苷酶由复旦大学药学院提供; 磺基异硫氰酸苯酯(4-Sulfophenyl isothiocyanate, SPITC)、TPCK 处理的胰蛋白酶(TPCK-trypsin)、三氟乙酸(TFA)和基质 $\alpha$ -氰基-4-羟基肉桂酸( $\alpha$ -Cyano-4-hydroxycinnamic acid, CCA)购于 Sigma 公司; 乙腈购自 Merck 公司; 测序级胰蛋白酶购自 Promega 公司; C18 ZipTip<sup>TM</sup>购自 Millipore 公司; 其它试剂均为分析纯.

4700 型 MALDI-TOF-TOF 飞行时间质谱仪(美国 ABI 公司); 恒温箱购自其林贝尔公司.

### 1.2 实验过程

色谱分离后的粗酶( $\beta$ -Glucosidase)经 SDS-PAGE 进一步纯化, 割下目标蛋白质条带. 蛋白质条带的胶内酶解、酶解肽段的 SPITC 磺化反应及 C18 ZipTip<sup>TM</sup>脱盐参照 Leon 等<sup>[11]</sup>方法进行. 酶解肽段和磺化肽段采用 4700 型 MALDI-TOF-TOF 飞行时间质谱仪进行 PSD 分析<sup>[11,12]</sup>. 质谱配备波长为 355 nm、频率为 200 Hz 的 Nd:YAG 激光器. 在进行串联(MS/MS)分析过程中, 关闭 CID(碰撞诱导解吸), MALDI 源产生的离子在 8 kV 加速电压下通过 6.7 kV 的栅格电压进入自由场漂移区的飞行管. MS/MS 数据进行手工分析得到肽段的序列. 新蛋白的序列采用 BLAST 工具对 NCBI 数据库进行比较分析(<http://www.ncbi.nlm.nih.gov/BLAST>).

## 2 结果与讨论

### 2.1 肽段序列分析

按照常规蛋白质鉴定的方法对 $\beta$ -葡萄糖苷酶进行了分析, 发现在 NCBI 数据库中未匹配到在统计学上有意义相关蛋白质. Edman 降解方法对 $\beta$ -葡萄糖苷酶 N 端测序也未得到 N 端序列, 可能是因为该酶的 N 端被封闭了. 因此发展了一种基于磺基异硫氰酸苯酯(SPITC)化学辅助方法针对新蛋白从头测序的技术平台<sup>[13]</sup>.

$\beta$ -葡萄糖苷酶经胰蛋白酶酶解, MALDI-TOF-TOF 分析后得到该酶的肽指纹图谱[图 1(A)]. 由于在 MALDI 条件下 C 端为 Arg 的肽(Arg-肽)离子化效率明显高于 C 端为 Lys 的肽(Lys-肽), 所以在经

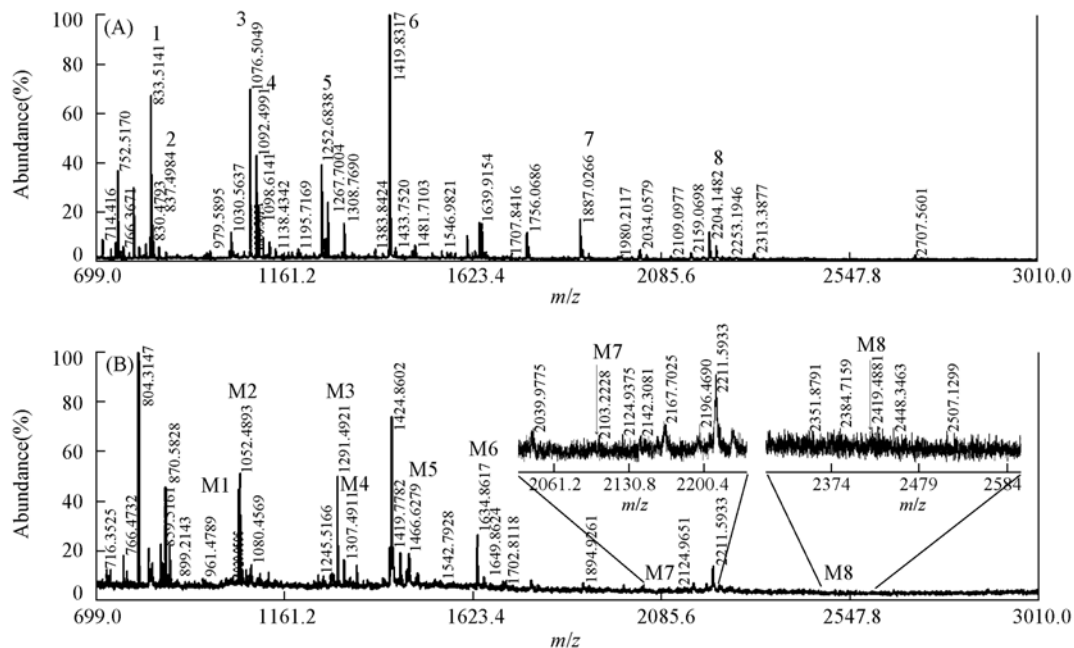


Fig. 1 Mass spectra of tryptic digest of novel protein (A) and derivatized tryptic digest of novel protein (B)

M1—M8 denoting the sulfonated peptides of 1—8.

过 SPITC 磺化反应后, 选取 8 条离子化效率相对较高的 Arg-肽进行分析, 如图 1 中 1~8 号肽段直接在肽段的 N 端修饰上 SPITC 基团(分子量增加 215), 得到 M1~M8 号磺化肽段[图 1(B)].

将酶解后的肽段(未磺化)直接进行串联分析, 得到的串联谱图[图 2(A)]非常复杂. 图 2(A)为酶解肽段  $m/z$  1076.5049 的 MS/MS 图. 除了一系列不完整的 N 端断裂离子(a, b, c 离子)和 C 端断裂离子(x, y, z 离子), 还含有内部断裂和侧链断裂离子. 因此无法进行精确的从头测序. 而经修饰后的磺化肽段在低能的 PSD 模式下非常容易断裂, 且只在肽键处断裂, 即只得到 b 和 y 系列离子; 而在 b 离子 C 端的正电荷被 b 离子 N 端的磺酸负电荷中和后整个肽段呈中性, 在质谱分析中不会出现信号, 只有带正电荷的 y 离子系列方能被检测, 因此可得到单纯的 y 系列离子<sup>[12]</sup>, 图 2(B)为酶解肽段  $m/z$  1076.5049 反应后磺化肽段  $m/z$  1291.4921 的 MS/MS 图. 与图 2(A)不同, 图 2(B)上清晰地展现出一系列 y 离子, 分别为  $y_1(+1): 175.0885$ ;  $y_2(+1): 322.1259$ ;  $y_3(+1): 451.1388$ ;  $y_4(+1): 508.1567$ ;  $y_5(+1): 565.1686$ ;  $y_6(+1): 696.2097$ ;  $y_7(+1): 833.1991$ ;  $y_8(+1): 1019.3085$ ;  $y_9(+1): 1076.3220$ . 通过手工计算相邻 y 系列离子之间的质量差可以得到该肽段的精确的序列为 GWHMGGEFR.

在进行序列分析的过程中发现了一对质量相差 16 的峰( $m/z$  1076.5049,  $m/z$  1092.4991)同时被修饰成一对质量增加 215 的峰( $m/z$  1291.4921,  $m/z$  1307.4911), 手工分析磺化后肽段的序列, 发现图 2(C)从  $y_6$  到  $y_9$  的质荷比比图 2(B)的相应位置的质荷比增加了 16, 它们的序列都为 GWHMGGEFR.  $m/z$  1092.4991 为该序列在甲硫氨酸处发生了氧化. 这意味着  $\beta$ -葡萄糖苷酶在该残基上可能存在甲硫

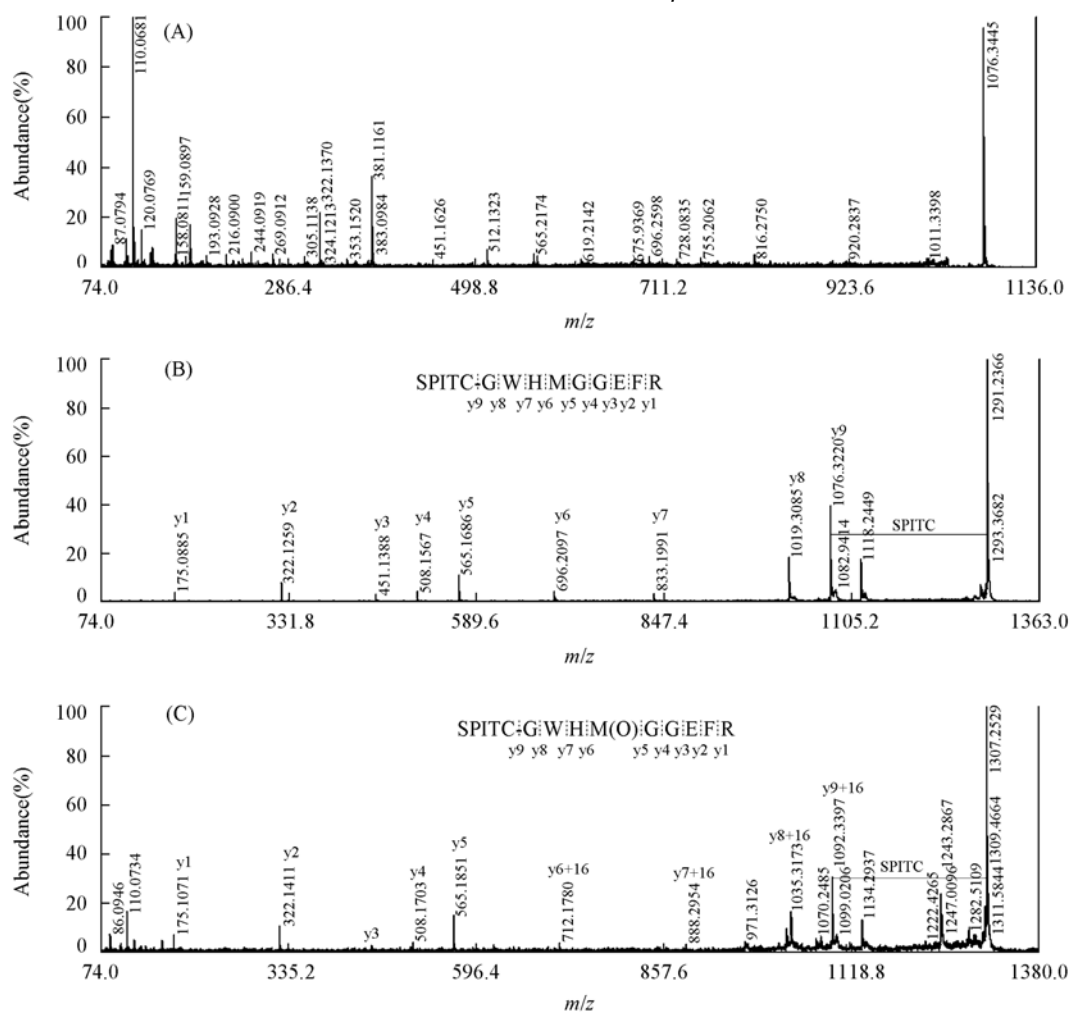


Fig. 2 Example of PSD spectra, acquired on a 4700-proteomics analyzer mass spectrometer, of the same peptide, either (A) non-sulfonated or (B) sulfonated with SPITC (M3) and (C) PSD spectra of sulfonated peptide (M4) with oxidation of methionine compared with M3

氨酸氧化和非氧化两种状态, 也有可能是在样品处理的过程中该甲硫氨酸发生了氧化。

另外, 如图 1(B) 局部放大图中箭头所指, 在对这 8 个磺化后的肽段进行 PSD 分析的过程中发现磺化肽段 M7 和磺化肽段 M8 的信噪比非常低, 但是得到的 PSD 谱图上 y 系列离子非常纯净完整。如图 3 所示, 磺化肽段 M7 和 M8 的信噪比仅为 4.1 和 2.3, 而经 PSD 分析后得到的 y 系列离子却非常完整, 肽段的序列可以完整的解析为 W[I/L][I/L][Q/K]SGSYNVFVIGSSSR 和 AYFPNTVVMWDTSPYAR, 由此可以说明磺化后的肽段离子化效率大大增强, 肽段的酰氨键更易断裂。同时, 在对这个磺化肽段从头测序的时候也发现质量数非常接近的氨基酸无法确认, Q 和 K 分别为 128.05858 和 128.09496, I 和 L 都为 113.08406。如图 3 所示, y16 与 y15 之间, y15 和 y14 之间的质量差都为 113, 无法区分 I 和 L, 而 y14 和 y13 之间的质量差为 128, 无法区分 Q 和 K。但通过 BALST 进行同源性分析可对这些氨基酸进行初步估算。

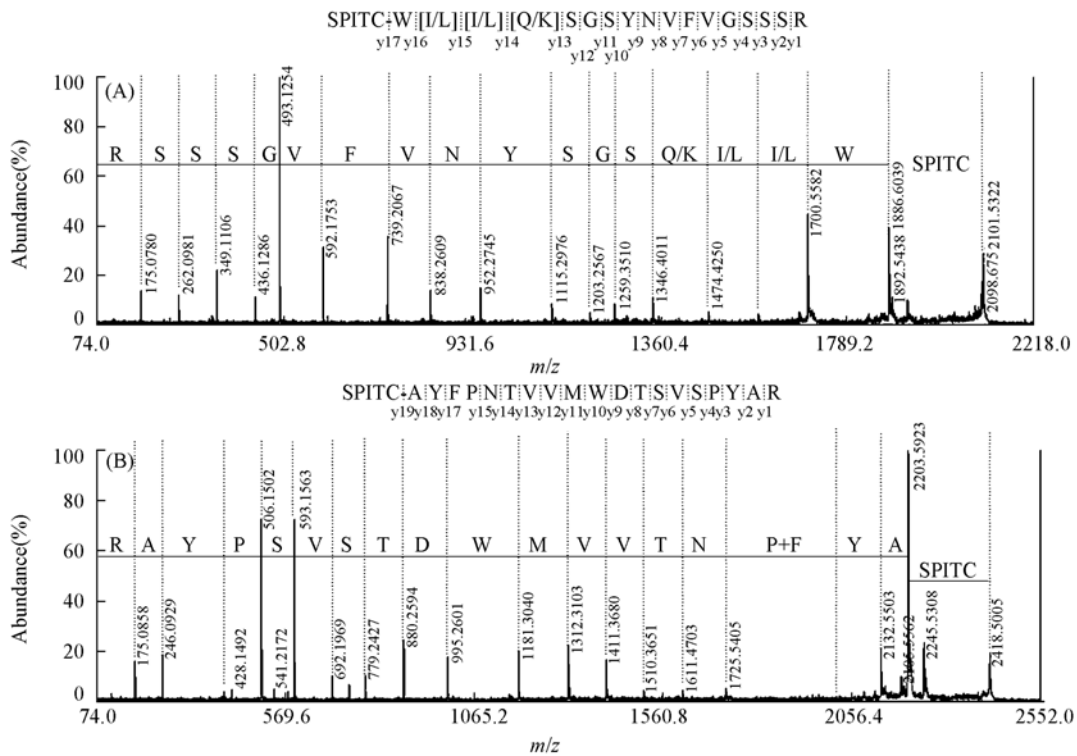


Fig. 3 PSD spectra of M7 (A) and M8 (B) with very low signal/noise ratio

## 2.2 肽段序列的同源性分析

通过计算 M1 ~ M8 磺化肽段的串联数据的 y 系列离子的差值, 得到新蛋白质部分精确的氨基酸序列, 结果如表 1 所示。

Table 1 Peptide fragments identified from tryptic digest of novel protein

No.	Precursor ion ( $m/z$ )	Modified precursor ion ( $m/z$ )	Peptide sequence
1	833.5141	1048.5028	VTFP[I/L]TR
2	837.4984	1052.4893	A[I/L]MPH[I/L]R
3	1076.5049	1291.4921	GWHMGGEFR
4	1092.4991	1307.4911	GWHM(O)GGEFR
5	1252.6838	1466.6279	Y[I/L]P[I/L]GAYV[I/L]SR
6	1419.8317	1634.8617	GV[Q/K]VA[I/L]GPVVG[S/L]GR
7	1887.0266	2102.4619	W[I/L][I/L][Q/K]SGSYNVFVIGSSSR
8	2204.1482	2419.3499	AYFPNTVVMWDTSPYAR

将从头测序得到的序列通过 BLAST 工具提交 NCBI 数据库进行同源性蛋白质检索, 结果没有检索到任何一个蛋白质包含这个新蛋白质的 8 条序列中任何 2 条序列, 即这个新蛋白质是一种未报道的新  $\beta$ -葡萄糖苷酶, 与采用常规蛋白质鉴定方法得到的结果一致。如表 2 所示, 通过与数据库中已知的  $\beta$ -

葡萄糖苷酶比较发现,在这8条序列中有2条序列GV[Q/K]VA[L/L]GPVVG[S/L]GR和W[L/L][L/L][Q/K]SGSYNVFVGSSSR与已知的 $\beta$ -葡萄糖苷酶序列具有较高的同源性,说明它们在 $\beta$ -葡萄糖苷酶序列中相对保守.同时,通过序列同源性分析也可推断序列GV[Q/K]VA[L/L]GPVVG[S/L]GR为GVQVALGPVVGSLGR,序列W[L/L][L/L][Q/K]SGSYNVFVGSSSR可能为WLIQSGSYNVFVGSSSR,没有一个已知的 $\beta$ -葡萄糖苷酶同时包含这2条序列.下一步研究可根据这2条保守序列设计引物从微生物中调取这个酶基因,以获得该新蛋白质的全长蛋白质序列,并进行基因克隆和表达.

Table 2 Comparison of two peptide sequences of novel protein with known  $\beta$ -glucosidase\*

Gen bank accession No.	Source	Start seq.	Sequence	End seq.
	Paecilomyces bainier		<b>W[L/L][L/L][Q/K]SGSYNVFVGSSSR</b>	
ABN73102	Aspergillus niger	827	VETQDWEITSYPKMFVVGSSSRKPLR	853
AAX07690	Magnaporthe grisea	842	VKVQDWWISEHPKKVVGSSSRKHLHS	869
BAA19913	Aspergillus kawachii	827	VEKQDWEITSYPKMFVVGSSSRKPPLR	853
YP_119169	Nocardia farcinica	629	TRVDRWLVEGGDYEVFVGASSRDIRLT	655
XP_001218668	Aspergillus terreus	751	TEAQQWKLQSGDYQVYVGRSSRDPLK	777
	Paecilomyces bainier		<b>GV[Q/K]VA[L/L]GPVVG[S/L]GR</b>	
XP_001261853	Neosartorya fischeri	145	EFKRKGVNVALGPVVGPLGRTARGG	169
XP_746996	Aspergillus fumigates	145	EFRRKGVNVALGPVVGPLGRTARGG	169
XP_001543147	Ajellomyces capsulatus	126	EHRDKGVDVVLGPVVGPLGRSPDGG	150
XP_001551828	Botryotinia fuckeliana	61	EFRKKGTHVALGPVAGPLGRHPLGG	85
AAT95376	Phaeosphaeria avenaria	147	EHRKGVVDVQLGPVVGPLGRNPKGG	171
AAT95381	Phaeosphaeria nodorum	147	EHRKGVVDVQLGPVVGPLGRNPKGG	171
ABX84365	Periconia	140	EHRKGVVDVQLGPVVGPIGRHPKGG	164
ABX79553	Thermoascus aurantiacus	132	EHRKGVVDVQLGPVAGPLGRHPDGG	156
XP_001591700	Sclerotinia sclerotiorum	162	EHKAKGVTVGLGPVAGPLGRSPQGG	186
XP_001244447	Coccidioides immitis	131	EHRDKGVDVQLGPAVGLGRSPDGG	155
AAL69548	Talaromyces emersonii	131	EHRKGVVDVQLGPVAGPLGRSPDAG	155
XP_001218668	Aspergillus terreus	141	EFHKKGVNVLLGPVVAPLGRVVEGG	165
AAA91297	Kuraishia capsulate	141	EFRHKGVNVLGPVVGPIGRVATGG	165
XP_001387646	Pichia stipitis	78	EHKKKGVHIALGPVVGPIGLKAAGG	102
AAX07690	Magnaporthe grisea	141	EAKDKGINVLLGPVAGGLGRVAAGG	165
XP_956104	Neurospora crassa	145	EAKGDGVTVLLGPVAGPLGRAPEGG	169

\* The identical amino acids are highlighted in black.

### 3 结 论

结合磺基异硫氰酸苯酯化学辅助对新蛋白质( $\beta$ -葡萄糖苷酶)部分肽段进行了从头测序,精确地确定了该蛋白质的8条肽序列.通过BLAST工具将这些序列提交NCBI数据库进行同源性蛋白质分析,找到了2条保守序列,为下一步基于已知肽段序列设计引物获得该新蛋白质全长序列,并进行结构与功能相互关系的研究奠定了基础.上述分析证明磺基异硫氰酸苯酯化学辅助方法对新蛋白质进行从头测序,方法精确可靠,值得推广.

### 参 考 文 献

- [1] MA Jie(马洁), WU Song-Feng(吴松锋), ZHU Yun-Ping(朱云平). Progress in Biochemistry and Biophysics(生物化学与生物物理进展)[J], 2007, **34**(8): 791—799
- [2] Liska A. J., Shevchenko A.. Proteomics[J], 2003, **3**(1): 19—28
- [3] SUN Guang-Zhi(孙光芝), LI Xiang-Gao(李向高), LIU Zhi(刘志), et al.. Chem J. Chinese Universities(高等学校化学学报)[J], 2007, **28**(7): 1316—1318
- [4] WANG Yu-Tang(王玉堂), LI Xu-Wen(李绪文), JIN Hai-Yan(金海燕), et al.. Chem J. Chinese Universities(高等学校化学学报)[J], 2007, **28**(12): 2264—2269
- [5] Mochizuki M., Yoo Y. C., Matsuzawa K., et al.. Biol. Pharm. Bull. [J], 1995, **18**: 1197—1202
- [6] Iishi H., Tatsuta M., Baba M., et al.. Clin. Exp. Metastasis[J], 1997, **15**: 603—611

- [ 7 ] Liu W. K. , Xu S. X. , Che C. T. . Life Sci. [J] , 2000 , **67** : 1297—1306
- [ 8 ] Min J. K. , Kim J. H. , Cho Y. L. , *et al.* . Biochem. Biophys. Res. Commun. [J] , 2006 , **349** : 987—994
- [ 9 ] Shinkai K. , Akedo H. , Mukai M. , *et al.* . Jpn. J. Cancer Res. [J] , 1996 , **87** : 357—362
- [ 10 ] Yan Q. , Zhou W. , Li X. W. , *et al.* . Biosci. Biotechnol. Biochem. [J] , 2008 , **72** : 352—359
- [ 11 ] Leon I. R. , Neves-Ferreira A. G. , Valente R. H. , *et al.* . J. Mass Spectrom. [J] , 2007 , **42** : 781—792
- [ 12 ] Chen P. , Nie S. , Mi W. , *et al.* . Rapid Commun. Mass Spectrom. [J] , 2004 , **18** : 191—198
- [ 13 ] Keough T. , Youngquist R. S. , Lacey M. P. . Proc. Natl. Acad. Sci. USA [J] , 1999 , **96** : 7131—7136

## De Novo Sequence of Novel Protein Based on Chemically-assisted Method Using 4-Sulfophenyl Isothiocyanate Derivatization by Mass Spectrometry

ZHOU Xin-Wen<sup>1</sup> , YAN Qin<sup>2</sup> , ZHOU Pei<sup>2</sup> , YANG Peng-Yuan<sup>1\*</sup>

(1. Institutes of Biomedical Sciences, 2. School of Pharmacy, Fudan University, Shanghai 200032, China)

**Abstract** Eight tryptic peptides of a novel ginsenoside Rb1 hydrolase ( $\beta$ -glucosidase) purified from *Paecilomyces bainier* were identified with chemically assisted de novo sequencing followed 4-sulfophenyl isothiocyanate (SPITC) derivatization by MALDI-TOF-TOF mass spectrometry. Some relatively complete sequences were obtained from sulfonated peptides with very low signal/noise ratio. A pair of peptides were identified with non-oxidation and oxidation at methionine, respectively. The results indicated that the sulfonated peptide by SPITC significantly enhanced MALDI-TOF-PSD fragmentation signals and produced a spectrum containing only  $y^+$  ions. This method facilitated de novo sequencing and spectrum interpretation. By comparing the eight identified peptide sequences with the NCBI database, this purified  $\beta$ -glucosidase proves to be a novel protein that has not yet been reported. Two relatively conserved peptides are remarkable contribution to further research.

**Keywords** SPITC; MALDI-TOF-TOF MS; Novel protein; De novo sequence

(Ed. :Y, G)

### 第九届亚太国际微分离分析学术会议 (APCE2009) 及 第一届亚太国际芯片实验室学术会议 (APLOC2009) 征稿通知

由中国科学院大连化学物理研究所主办的“第九届亚太国际微分离分析学术会议 (APCE2009)”和“第一届亚太国际芯片实验室学术会议 (APLOC2009)”将于 2009 年 10 月 28 ~ 31 日在上海召开。林炳承、秦建华教授任大会主席。本次会议将致力于推动各种微型分离分析技术的发展, 同时也将对以微流控芯片或芯片实验室 (Lab on a Chip) 为代表的新型科学技术及其在生命科学领域的应用予以特别关注。会议将邀请一大批来自国内外的知名学者、专家参会并做邀请报告和大会报告, 热忱欢迎大家踊跃投稿并到会交流。会议投稿、注册等相关事宜请登陆会议网站。

本次会议主题涉及微纳生物化学分离分析和微流控芯片实验室的各个方面, 主要内容包括微纳流控理论、微纳流体系统、微纳分析与合成系统、微纳材料及表面、微纳加工技术、微流控芯片集成技术、微流控芯片接口技术、微流控芯片控制、检测技术等及其在化学、生命科学中的应用等方面。

凡与上述主题相关且尚未公开发表的关于新理论、新方法、新技术和新应用等研究论文均在应征之列。应征稿件将提交大会学术委员会审定, 接受论文将编入本届学术会议论文集。应征稿件请于 2009 年 7 月 31 日前发送到 shiw@dicp.ac.cn。

欢迎国内外分析仪器公司、厂商到会参展。

通讯地址: 大连市中山路 457 号, 大连化学物理研究所, 邮编: 116023

联系人: 时蔚; 电话/传真: 0411-84379059

E-mail: shiw@dicp.ac.cn; 会议网站: <http://www.biochem.dicp.ac.cn/apce2009/index.html>