

基于 LC-ESI-MS/MS 技术与生物信息学的 蚕豆种子贮藏蛋白亚基鉴定

刘玉皎 侯万伟

(青海省农林科学院 青海大学农林科学院/

青海省高原作物种质资源创新与改良利用国家重点实验室培育基地,西宁 810016)

摘要 蚕豆蛋白亚基是蚕豆种子贮藏蛋白重要组成部分,深入鉴定蚕豆蛋白亚基,对了解蚕豆种子蛋白不同亚基的结构和功能具有重要意义。本研究应用液相色谱、电喷雾离子化与串联质谱联用技术和生物信息技术对蚕豆种子贮藏蛋白 6 个特异亚基进行了质谱鉴定。结果表明:这些特异亚基中的 64、47、42 及 38 ku 分别鉴定出 4 个蛋白:豌豆球蛋白、豆球蛋白前体、假定糖结合蛋白、LEGB7_VICFA;97 ku 亚基蛋白为分子伴侣 GroEL;96 ku 亚基蛋白由 6 个具有不同代谢功能的蛋白组成。

关键词 蚕豆; 蛋白亚基; 串联质谱; 鉴定

中图分类号 S 529

文章编号 1007-4333(2014)01-0037-06

文献标志码 A

Identification for seed store protein subunits of faba bean based on LC-ESI-MS/MS and bio-informatics

LIU Yu-jiao, HOU Wan-wei

(Qinghai University Academy of Agricultural and Forestry Science/Cultivating Base of National Key Lab-Plateau Crop Gene Innovated and Utilized in Qinghai Province, Qinghai Academy of Agricultural and Forestry Science, Xining 810016, China)

Abstract The protein subunits are important components for store protein in faba bean seed, since it is essentially the first step to further understand their structures and functions of these protein subunits in detail. Six specific protein subunits of faba bean were identified by using liquid chromatograph (LC), LC-ESI-MS/MS and bio-information approaches in this study. The results indicated that four proteins were identified from each of the 64, 47, 42 and 38 ku specific protein subunits. The four proteins were convicilin, legumin a precursor, putative sucrose binding protein and LEGB7_VICFA; 97 ku encoded the same protein as GroEL; 96 ku was composed of six proteins with different metabolic characteristics.

Key words faba bean; protein subunit; MS/MS; identification

蚕豆种子贮藏蛋白主要由清蛋白、球蛋白组成,而蛋白亚基是蛋白质的重要组成部分^[1-3]。生物质谱(Biological Mass Spectrometry, MS)技术能够准确测定多肽或蛋白质相对分子量、氨基酸序列和翻译后修饰等^[4],是研究蛋白质亚基结构和功能的重要手段,而串联质谱与液相色谱耦联进行数据依赖型扫描,可以分辨出几十种复杂蛋白质的混合物^[5-6]。采用蛋白质库搜索比对是蛋白组学研究中

最主要的方法^[7-8],常用的数据库搜索算法主要有 SEQUEST、Mascot、X! tandem 等^[6],但 SEQUEST 方法鉴定蛋白的结果更可靠^[9-10]。蚕豆不同蛋白的氨基酸组成和功能是不同的,蚕豆球蛋白属贮藏蛋白,而清蛋白是一些功能性蛋白,其中含硫氨基酸较球蛋白高^[11]。本研究拟采用十二烷基磺酸钠-聚丙烯酰胺凝胶(SDS-PAGE)电泳结合液相色谱-电喷雾离子化-串联质谱技术(LC-ESI-MS/MS)和

收稿日期:2013-05-20

基金项目:国家 973 计划前期专项(2010CB134408);现代农业产业技术体系建设专项(CARS-09);青海省高校“135”高层次人才培养工程

第一作者:刘玉皎,研究员,主要从事蚕豆种质创新与改良利用研究,E-mail:13997058356@163.com

ProSEQUEST方法对蚕豆种子贮藏蛋白特异亚基的结构和功能进行鉴定,结果对于掌握蚕豆种子贮藏蛋白的结构和功能具有重要意义。

1 材料与方法

1.1 材料

试验所用蚕豆品种为青海省农林科学院选育的青海13号;测序级的修饰胰蛋白酶是从Promega公司购买,其他用于高效液相色谱法测定水解和所有化学试剂从Sigma公司购买,普通试剂均为国产分析纯。试验仪器及其制造商为高速浓缩离心设备, Thermo公司;高效液相色谱仪, Agilent公司;离子阱质谱仪(LCQ DECA XP PLUS), Thermo公司。

1.2 蛋白质样品的制备与分离

蚕豆蛋白质提取采用陈毓荃方法^[12]。采用SDS-PAGE^[13]对蚕豆贮藏蛋白进行第一次分离,并用考马斯亮蓝法染色^[14],目标条带进行回收后用于质谱分析。

1.3 蛋白凝胶的酶解

对回收的每个蛋白亚基胶带用DTT还原二硫键脱色和用碘乙酰胺对所有的半胱氨酸残基烷基化。然后,将样品经水清洗,用测序级修饰胰蛋白酶在消化缓冲液(碳酸氢铵100 mmol/L, pH 8.5)中消化凝胶。用乙腈提取消化肽,并在浓缩离心设备彻底干燥下来。最后对干燥的样品在2.0%的乙腈、97.5%的水和0.5%甲酸的溶解液中再溶解^[15]。

1.4 液相色谱-电喷雾-串联质谱分析

液相色谱-电喷雾-串联质谱(LC-ESI-MS/MS)分析是通过高效液相色谱法(HPLC)系统进行的,用直径75 μm,长度为8 cm的反相柱C₁₈, C₁₈的颗粒大小为300 Å的孔径。加样时间为20 min。HPLC溶剂A为97.5%的水、2.0%的乙腈和0.5%的甲酸。高效液相色谱溶剂B是9.5%的水、90.0%的乙腈和0.5%的甲酸。溶剂B从2%到90%的渐变时间为60 min, 20 min的样品装载, 20 min洗柱。柱流量约为800 μL/min后分裂。标准的加样量是3 μL。

HPLC系统是用离子阱质谱仪在线连接,从HPLC色谱柱洗脱的样品直接被电喷雾电离进入质谱仪。电离的电压每次均调在1.2~1.8 kV的有效范围内。毛细管的温度设置在110℃。质谱仪设置在数据依赖模式下通过低通量诱导碰撞解离(CID)处理获得MS/MS数据。默认的碰撞能量为

33%,默认充电状态是3。要求一个完整扫描的质荷比在550~1 800 amu范围内完成,然后,在全质量范围内用3个微扫描进行一个串联质谱离子扫描。动态排斥特征设置如下:在0.3 min内重复计数1,排斥时间为0.4 min,排斥宽度是4 u。

1.5 质谱数据库搜索

在默认状态下,质谱数据在http://www.ncbi.nlm.nih.gov/protein的NR数据库进行搜索。用ProtTech's ProtQuest软件处理数据,各种参数设置如下:

Data acquisition parameters:

Instrument: LCQ DECA XP Plus

Software: Xcalibur 2.0

Centroid mode

Minimum MS signal for Precursor-ion: 5 × 10⁴ counts

Processing parameters:

Software: ProtQuest 2.0

Signal to noise: >=5

De-isotoped: Yes

Filtered: Yes

Missed cleavages: <=5

Modifications: Cys +57

Database: NR database from NCBI's GenBank

Validation parameters:

Peptide=1, score >=13

Peptide>1, score >=15

1.6 蛋白质定量

采用Noelle M Griffin等人的方法^[16]进行无标记蛋白质定量。蛋白质相对丰度是指样品中每一个蛋白归一化光谱指数(SI_N)占样品总归一化光谱指数的百分比。公式中: p_n: 独特肽段数; s_c: 光谱数; I_k: 肽段强离子片段(k); n: 蛋白总量; L: 每个蛋白中的氨基酸数。

$$SI_N = \left[\sum_{k=1}^{p_n} \left(\sum_{j=1}^{s_c} I_j \right)_k \right] / \sum_{j=1}^n SI_j / L$$

2 结果与分析

2.1 蚕豆不同蛋白亚基不同蛋白肽的质谱分析

试验共分析了97、96、64、47、42和38 ku等6个蚕豆蛋白亚基,结果见表1。表1显示:6个蛋白亚基通过高效液相色谱分离出16个蛋白,蛋白质质量

表 1 蚕豆种子蛋白亚基不同蛋白的 MS/MS 分析结果

Table 1 MS/MS result of different protein of protein subunit in faba bean

亚基 Subunit	蛋白数 Hits	蛋白质量 Protein mass	肽段数 Peptide No.	相对丰度/% Relative abundance	肽质量 Peptide mass	肽段序列 Peptide sequence
97	1	52 426.79	2	100.0	1916.03 2 087.09	AAVEEGIVPGGGTALLYATK TNDIAGDGT TTTATVLAQAIVR
96	1	36 683.17	4	50.2832.48	AAVLANFK 1 440.72 1 549.80 1 913.93	IGYDGIVFASNASK VIEIAGDYTETLAR TSPSAVGVFGLGFYDQNR
	2	42 811.8	2	14.9	904.49 1 936.94	TTDVTGAIK ELLSTYDFPGDDTPVIR
	3	31 418.65	2	15.3	1 903.91 2 133.02	AAVDAGFVPNDMQVGQTGK TAFDPVAAEGGSAAVEAVSEAK
	4	20 626.47	2	9.3	1 338.76 1 502.77	IQIVEINAGGIGR NFDVLI EAAGLADR
	5	47 437.12	1	5.7	1 939.89	MLDEIGTVENVADFMEK
	6	27 456.53	1	4.6	1 915.95	LTDPVEAHDWLHGLSLAR
64	1	57 580.61	10	91.6	1 300.68 1 986.00 1 366.79 2 120.00 2 111.02 1 135.59 1 197.67 1 153.71 1 153.71 1 397.66	SQRQETNALVR VLLEEQDKESQQSIGQR VVDLAISVNRPGK VPAGTTSYLVNQDDEQDLR GNLELVGIQNEQQEQQR NILEASFNTK AILTVLSPNDR AIVVLLVNEGK AIVVLLVNEGK SDLFENLQNYR
	2	52 426.79	2	8.4	1 916.03 2 087.09	AAVEEGIVPGGGTALLYATK TNDIAGDGT TTTATVLAQAIVR
47	1	54 976.93	4	69.2	1 365.80 1 138.63 1 827.87 1 304.64	VKGGLSIITPPER GGLSIITPPER IESEGGLIETWNPNNR DFLEDALNVR
	2	54 979.31	3	30.8	1 399.73 1 204.61 1 474.80	GHGLVAEDQTER FNLEEGDLIR NIENYGLAVLEIK
42	1	54 976.93	3	86.8	1 230.61 1 365.80 1 304.64	NEDEEKGAIVK VKGGLSIITPPER DFLEDALNVR
	2	52 426.79	1	13.2	2 087.09	TNDIAGDGT TTTATVLAQAIVR
38	1	52 426.79	2	5.9	1 916.03 2 087.09	AAVEEGIVPGGGTALLYATK TNDIAGDGT TTTATVLAQAIVR
	2	37 862.21	2	89.1	2 470.11 2 624.20	IINPEGQEEEEQEEEEKQR VFYLGGNPEVEFPETQEEQQR
	3	49 472.45	1	5.0	2 096.03	EFLENYLLTDEGLEAVNK

在 20~57 ku 之间,相对丰度在 5.0%~100.0%。经过胰蛋白酶在精氨酸(R)、赖氨酸(K)切位点将不同蛋白切割成不同的蛋白肽,不同蛋白的肽段数在 1~10,16 个蛋白包含有 42 条肽段。这些肽段经串联质分析后,得到每个蛋白肽段的氨基酸序列,每个

肽段的氨基酸在 8~23 个。

2.2 蚕豆种子贮藏蛋白亚基蛋白质信息

基于肽段 MS/MS 序列基础上,在 <http://www.ncbi.nlm.nih.gov/protein> 的 NR 数据库进行相似序列比对,结果见表 2。可见 16 种蛋白中有

表 2 蚕豆种子蛋白亚基的蛋白信息

Table 2 Information of Protein Subunit of faba bean

亚基 Subunit (ku)	蛋白 Hits	蛋白质名称 Protein Name	覆盖度/% Coverage	登录号 gi number	NBCI Accession No.	来源 Source
97	1	分子伴侣 GroEL chaperonin GroEL	8.10	94497512	ZP_01304082.1	<i>Sphingomonas</i> sp. SKA58
96	1	磷酸 ABC 转运周质底物结合蛋白 phosphate ABC transporter periplasmic substrate-binding protein	15.84	262368531	ZP_06061860.1	<i>Acinetobacter</i> <i>johnsonii</i> SH046
	2	延伸因子 Tu elongation factor Tu	6.63	226953421	ZP_03823885.1	<i>Acinetobacter</i> sp. ATCC 27244
	3	电子转移黄素 α 蛋白亚基 electron transfer flavoprotein subunit alpha	13.50	262369529	ZP_06062857.1	<i>Acinetobacter</i> <i>johnsonii</i> SH046
	4	烷基过氧化氢还原酶 C22 亚基 alkyl hydroperoxide reductase C22 subunit	14.44	262369692	ZP_06063020.1	<i>Acinetobacter</i> <i>johnsonii</i> SH046
	5	柠檬酸合成酶 citrate synthase	4.01	262369421	ZP_06062749.1	<i>Acinetobacter</i> <i>johnsonii</i> SH046
	6	motA/TolQ/ExbB 质子通道家族 蛋白 motA/TolQ/ExbB proton channel family protein	6.39	332186771	ZP_08388513.1	<i>Sphingomonas</i> sp. S17
64	1	豌豆球蛋白 convicilin	26.56	164512572	CAP06335.1	<i>Vicia faba</i>
	2	分子伴侣 GroEL chaperonin GroEL	8.30	94497512	ZP_01304082.1	<i>Sphingomonas</i> sp. SKA58
47	1	豆球蛋白前体 legumin A precursor	10.58	600108	CAA86824.1	<i>Vicia narbonensis</i>
	2	假定糖结合蛋白 putative sucrose binding protein	7.47	12580894	CAC27161.1	<i>Vicia faba</i> var. minor
42	1	豆球蛋白前体 legumin A precursor	7.26	600108	CAA86824.1	<i>Vicia narbonensis</i>
	2	分子伴侣 GroEL chaperonin GroEL	4.25	94497512	ZP_01304082.1	<i>Sphingomonas</i> sp. SKA58
38	1	分子伴侣 GroEL chaperonin GroEL	8.30	94497512	ZP_01304082.1	<i>Sphingomonas</i> sp. SKA58
	2	LEGB7_VICFA HTLV-1 GP21 胞外区的麦芽糖结 合蛋白嵌合体	12.54	126167	P16080.1	<i>Legumin</i> type B
	3	Htlv-1 Gp21 Ectodomain MALTOSE-Binding Protein Chimera	4.00	4930066	IMG1	—

4 个分子伴侣蛋白和 12 个特异蛋白,其中 97 ku 是分子伴侣 GroEL,相对丰度为 100%,64 和 42 ku、38 ku 中分子伴侣 GroEL 的相对丰度较低。12 个特异蛋白中 64 ku 的 1 号蛋白是豌豆球蛋白、47 ku 的 1 号蛋白是豆球蛋白前体和 2 号蛋白是假定糖结合蛋白、42 ku 的 1 号蛋白是豆球蛋白前体、38 ku 的 2 号蛋白是 LEGB7_VICFA,这些蛋白均是蚕豆 (*Vicia faba*) 或巢菜属 (*Vicia*) 或其他豆科植物鉴定出的特有蛋白,相对丰度分别为 91.6%、69.2%、30.8%、86.8% 和 89.1%,被测序列的覆盖度分别为 26.56%、10.58%、7.47%、7.26%、12.54%。96 ku 亚基的 6 个蛋白均为特异蛋白,其被测序列的覆盖度在 4.01%~15.84%。

3 讨论与结论

3.1 关于蚕豆种子贮藏蛋白

种子蛋白质在种子生长发育直至幼苗早期生长的过程中都扮演着极其重要的角色。不仅为种子自身呼吸代谢和萌发提供营养,负责幼苗早期生长发育的养料供给;同时,还调控着机体各种生理生化反应和代谢过程。生物体在不同的发育阶段会合成类型数量不同的蛋白质。这些蛋白质代表着机体某时刻特征性生命活动的基础^[17]。Shigeru Utsumi 等^[11]、李雪琴等^[18]、刘玉皎等^[19] 等分别通过沉降系数法以及 SDS-PAGE 法研究了蚕豆种子贮藏蛋白的组成,认为主要由清蛋白和球蛋白组成,每个蛋白分别由多个亚基组成。本研究通过质谱串联技术 (LC-ESI-MS/MS) 对蚕豆种子贮藏蛋白亚基进行了进一步鉴定,表明不同蛋白亚基由 1~6 个蛋白组成,通过质谱分析蚕豆种子贮藏蛋白亚基的质量分数在 30~60 ku 之间,与李雪琴等^[20] 的结果是一致的。根据生物信息技术找到了这些蛋白的初步信息,其中有 4 个蛋白是从蚕豆中发现的或是从其他豆科植物中发现的。

3.2 关于蚕豆蛋白质消化

蛋白组学研究中消化是通过酶学和化学方法在某些特定氨基酸残基上切割蛋白质,产生最适合于分析的片段,一般为 3~20 个氨基酸的肽片段,而且肽段小于 5 000 u 对质谱分析和数据库比较最为理想,蛋白质消化主要依靠蛋白酶,不同的蛋白酶会在

不同的切位点切割,也会产生不同大小的肽片段,表现不同的质谱分析效果。本研究采用胰蛋白酶消化蛋白,胰蛋白酶具有易纯化,其切位点在赖氨酸 (K)、精氨酸 (R) 的羟基端,而且赖氨酸、精氨酸在许多蛋白质中的距离能产生长度很适合于质谱分析的肽,对于高赖氨酸和精氨酸的蚕豆蛋白是比较适合的,但对在赖氨酸或精氨酸的 C 端方向有脯氨酸则不能切割^[5]。

3.3 串联质谱与生物信息学鉴定

蛋白质鉴定常采用双向电泳分离蛋白,而本研究首先采用 SDS-PAGE 分离目标蛋白,然后高效液相色谱 (LC) 系统通过电喷雾离子化 (ESI),离子源与串联质谱耦联,进行数据依赖型扫描,可以分辨出几十种成分的复杂蛋白质混合物^[5]。但单凭一种分析技术经常无法准确鉴定蛋白,通过使用不同的分析手段,可以起到互为补充的作用,有利于提高蛋白质鉴定的效率和准确度^[21]。因蚕豆基因组没有测序,这些蛋白鉴定是基于蛋白质量和以 SEQUEST 序列查询模式搜索 NCBI/nr 数据库的绿色植物蛋白数据库,被测肽段序列在数据库中预测蛋白的覆盖度在 4.01%~26.56% 之间相对较低。串联质谱搜索 MASCOT 和 EST 数据库的匹配性高,对未知生物蛋白质鉴定是一种可靠的途径^[22]。

参 考 文 献

- [1] 刘玉皎,侯万伟,石建斌. 蚕豆蛋白亚基组成分析与特异种质鉴定[J]. 西北植物学报,2012,32(1):54-59
- [2] 石建斌,侯万伟,刘玉皎,等. 蚕豆种质资源清蛋白遗传多样性分析[J]. 植物遗传资源学报,2012,13(1):22-28
- [3] 石建斌,侯万伟,刘玉皎. 蚕豆种子贮藏蛋白组分的比较研究[J]. 植物遗传资源学报,2012,13(2):304-307
- [4] 崔丽娟,黄瑾. 生物质谱技术在蛋白质结构鉴定中的研究进展[J]. 农垦医学,2009,31(4):349-352
- [5] 陈捷主编. 农业生物蛋白质组学[M]. 北京:科学出版社,2009
- [6] 薛庆中. DNA 和蛋白质序列数据分析工具[M]. 2 版. 北京:科学出版社,2010
- [7] 杨阳,蒋爱玲. 液相色谱-质谱/质谱联用技术及其应用[J]. 中国卫生检验杂志,2003,13(6):783-785
- [8] 彭咏波,马永平,丁世家,等. 基质辅助激光解析飞行时间质谱分析条件及准确度对蛋白质数据库搜索结果的影响[J]. 分析化学,2009,37(3):319-324
- [9] 董乃平,李洪东,梁逸曾. 基于串联质谱信息进行蛋白质数据库搜索的结果可靠性分析[J]. 分析化学,2009,37(10):1473-

- 1478
- [10] 刘韬,曾嵘,邵晓霞,等.毛细管区带电泳/串联质谱联用法鉴定多肽和蛋白质[J].生物化学与生物物理学报,1999,31(4):425-432
- [11] Shigeru Utsumi, Zenichi Yokoyama, Tomohiko Mori. Comparative studies of subunit compositions of *Vicia faba* L seeds[J]. Agric Biol Chem,1980,44(3):595-601
- [12] 陈毓荃.生物化学研究技术[M].北京:中国农业出版社,1995
- [13] Coligan J E. Short Protocols in Protein Science(精编蛋白质科学实验指南)[M].李慎涛,译.北京:科学出版社,2007
- [14] 郭尧君.蛋白质电泳实验技术[M].北京:科学出版社,1999
- [15] Shevchenko A, Wilm M, Vorm O, et al. Mass spectrometric [J]. Anal Chem,1996,68:850-851
- [16] Noelle M Griffin, Jingyi Yu, Fred Long, et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis[J]. Nature Biotechnology,2010,28(1):83-89,doi:10.1038/nbt.1592
- [17] Bewley J D. Seed germination and dormancy[J]. The Plant Cell,1997,9:1055-1066
- [18] 李雪琴,苗笑亮,裘爱泳.几种豆类蛋白质结构和组成的比较[J].粮食与油脂,2003(6):19-20
- [19] 刘玉皎,侯万伟,李萍,等.青海不同基因型蚕豆蛋白组成及清蛋白和球蛋白亚基的 SDS-PAGE 分析[J].甘肃农业大学学报,2012,47(2):68-71
- [20] 李雪琴,苗笑亮,裘爱泳.蚕豆蛋白的提取分离及相对分子质量的测定[J].无锡轻工大学学报,2003,22(6):71-74
- [21] 谢苗,成娟,尤民生,等.基于质谱和生物信息学分析的小菜蛾蛋白质鉴定[J].昆虫学报,2009,52(11):1206-1212
- [22] 刘康,高起飞,万振昆,等.蛋白质组学研究中的质谱鉴定与生物信息学分析[J].棉花学报,2008,20(4):281-288

责任编辑:袁文业