

大豆 11S 球蛋白 Gy5(A₃B₄)的基因克隆和序列分析

陈三凤

(中国农业大学生物学院 北京 100094)

深泽亲房

(日本国立食品研究所 筑波 日本)

关键词 大豆球蛋白, Gy5(A₃B₄)基因, 外显子, 内含子, DNA 序列分析

中图分类号 Q78 文献标识码 A 文章编号 1000-3061(2000)02-0215-03

大豆 11S 球蛋白(Glycinin)是大豆种子的主要贮藏蛋白,分子量为 360kD,由 6 对相同的蛋白亚基(每对亚基的分子量约 60 kD)构成。每对亚基又是由一个酸性 A 肽(35~45kD)和一个碱性 B 肽(22 kD)通过二硫键连接而成。A 肽和 B 肽源自同一个基因,即首先由一个大的 mRNA 翻译成 A-B 蛋白前体,然后再通过翻译后加工形成酸性 A 肽和碱性 B 肽^[1-3]。

研究表明,大豆 11S 球蛋白共有 5 种 A-B 亚基对,它们分别是由 5 个不同的基因所编码的,即 Gy1(A₁B₂)、Gy2(A₂B_{1a})、Gy3(A_{1a}B_{1b})、Gy4(A₅A₃A₃)和 Gy5(A₃B₃)。根据 cDNA 核苷酸序列的同源性,将该家族分为 I 和 II 两个类群。I 类群包括 Gy1~Gy3, II 类群包括 Gy4 和 Gy5。同一类群的同源性在 85%~90%,而不同类群的同源性只有 50%左右^[4-5]。

美国和日本已先后对这 5 个基因的 cDNA 进行了核苷酸序列分析,并且对 Gy1~Gy4 染色体基因也进行了克隆和序列分析,它们的共同特点是均有 4 个外显子和 3 个内含子组成^[6-10]。由于一直缺少有关 Gy5(A₃B₄)染色体基因组成情况的资料,未能对该基因家族做更加深入的研究,为此,本文报道作者在日本从事客座研究期间有关 Gy5 染色体基因的克隆和序列分析的一些研究工作。

1 材料和方法

1.1 大豆(品种 Williams)基因组文库

购自美国 Stratagene 公司。

1.2 菌株与载体

大肠杆菌 DH5a、XL-1-blue(P₂)、质粒 pUC19 和 pGA₃B₄1425(为 pBR322 含有 1.8 kb Gy5 cDNA)为本室保存;各种限制酶、连接酶购自日本 Takara 公司;自由引物探针标记试剂盒购自日本 Nippon Gene 公司, Autocycle 序列分

析试剂盒购自 Amersham pharmacia 公司。

1.3 探针标记

制备 pGA₃B₄ 1425 质粒 DNA,经 Pst I 单酶切后,琼脂糖凝胶电泳分离出 1.8 kb 的 A₃B₄ cDNA 片段作模板,标记反应完全按照试剂盒说明书提供的方法进行,在 30μL 的反应液中,加入模板 DNA 0.2 μg, [α-³²P]dCTP 5μL, 37℃ 反应 2~3 h,然后过 Sephadex G-75 柱纯化,取高 cpm 的洗脱液做杂交用。

1.4 原位杂交

噬菌斑平板的制备及将 DNA 转移到硝酸纤维素膜的操作参考分子克隆方法^[12]。硝酸纤维素膜的预杂交和杂交在 42℃ 下进行。所用缓冲液:50% 甲酰胺, 5×SSPE, 0.1% SDS, 0.1% Ficoll, 0.2 mg 鳕鱼精 DNA。杂交膜用 2×SSC, 0.1% SDS, 1 mmol/L EDTA 缓冲液室温下洗 3 次,每次 5 min。再在 42℃ 下洗 2~3 次,每次 5~10 min。滤膜在 -80℃ 放射自显影 1~2 d。

1.5 阳性克隆的 Southern blot 和亚克隆

用限制酶: Sal I, Not I, Xba I, Sac I 酶切阳性克隆,然后转移到硝酸纤维素膜上进行 Southern blot 杂交,杂交方法同于原位杂交。然后将含有阳性信号的片段亚克隆到质粒 pUC19 中直接进行 DNA 序列分析。

1.6 DNA 序列分析

DNA 序列分析采用双脱氧末端终止法,以双链质粒 DNA 为模板,按 Autocycle 序列分析试剂盒说明书进行。测序用 Shimadzu DSQ-1000 自动序列仪,并用计算机软件 GENETYX(日本软件发展公司)程序进行序列分析。

2 结 果

2.1 阳性克隆的筛选和亚克隆

以 Gy5(A₃B₄)cDNA(1.8 kb)为探针,通过原位杂交,从

大豆(品种 Williams)基因组文库中筛选 Gy5 染色体基因。在 72 000 个噬菌斑中获得 3 个阳性克隆,分别定名为 FIX-10, FIX-29 和 FIX-61。对 FIX-61 和 FIX-10 两个克隆进行了亚克隆和核苷酸序列分析,发现前者编码 Gy5(A₄B₄)基因,而后者含有 Gy4(A₅A₄B₃)基因(另文报道),这是由于 Gy4 和 Gy5 属同一个类群,其同源性高达 80% 以上,现对 FIX-61 的亚克隆和核酸序列分析作一报道。

因为本试验中所用的大豆品种 Williams 基因组文库是用噬菌体 FIXII 作载体,将叶片 DNA 9~20 kb/Sau 3A 片段插入 FIXII 的 M_{ob}I 位点而构成的。因此,用插入位点附近的限制酶 *Sa*lI、*Not*I、*Xba*I 和 *Sac*I 酶切阳性克隆 FIX-61,琼脂糖电泳表明插入片段约为 20 kb。*Sac*I 酶切插入片段形成 6 kb 和 14 kb 两个片段。Southern blot 杂交发现仅有 6 kb 片段有阳性信号,因此将该片段亚克隆到 pUC19 中,定名为 pGY-6。从该亚克隆 pGY-6 中提取质粒 DNA,直接进行 DNA 序列分析。

2.2 Gy5(A₃B₄)基因的核酸序列分析

DNA 序列分析表明,获得了全长的 Gy5(A₃B₄)染色体基因,其长度为 2819 bp,编码 517 个氨基酸,由 4 个外显子和 3 个内含子组成。外显子 1:内含子 1:外显子 2:内含子 2:外显子 3:内含子 3:外显子 4 的长度为 292bp:358bp:

263bp:425bp:645bp:485bp:351bp。对基因起始密码子 ATG 上游 46 bp 和终止密码子 TAA 下游的 503bp 也进行了测序,总长度为 3368bp,该基因已被收入基因数据库 (DDBJ) 中,其注册号是 AB003680。核苷酸序列比较表明,Gy5 染色体基因的外显子与 Gy5 cDNA 的核苷酸序列完全相同^[7]。外显子与内含子交界处完全遵守 GT-AG 规律。说明我们获得了 Gy5 染色体基因。基因 3'-端有 2 个多聚腺苷酸加尾信号序列 5'-AATAAA-3'。由于基因上游只获得了 46bp,而未发现 TATA 盒和 CAAT 盒。

2.3 Gy5 基因与 Gy1~4 基因在外显子和内含子组成上的比较

Gy5 基因由 4 个外显子和 3 个内含子组成,这与 Gy1~Gy4 的结果一致^[8~11]。说明这 5 个基因的共同特征都是有 4 个外显子和 3 个内含子,3 个内含子在这 5 个基因中的相对位置也大约相同,其特征见表 1。Gy4 和 Gy5 同源于第 II 类群,其 cDNA 同源性高达 80% 以上^[6],其内含子的位置也大致相同,但 3 个内含子的长度却有很大差异,特别是第二个内含子的差异更大。Gy5 基因内含子 2 长达 425 bp,而 Gy4 基因内含子 2 只有 75bp。这种现象也见于第 I 类群 Gy1~Gy3 基因中。

表 1 Gy1~Gy5 基因外显子和内含子组成比较

基因	类群	编码的蛋白亚基	外显子 1	内含子 1	外显子 2	内含子 2	外显子 3	内含子 3	外显子 4
Gy ₁	I	A _{1b} B ₂	286	228	254	291	558	381	387
Gy ₂	I	A ₂ B _{1a}	277	238	254	292	537	624	387
Gy ₃	I	A _{1a} B _{1b}	286	617	245	312	525	439	387
Gy ₄	II	A ₅ A ₄ B ₃	289	332	266	75	747	501	387
Gy ₅	II	A ₃ B ₄	292	358	263	425	645	485	351

3 讨 论

我们获得的 Gy5 基因由 4 个外显子和 3 个内含子组成,其外显子的序列与已报道的 Gy5(A₃B₄)cDNA 序列完全相同,都编码 517 个氨基酸。这表明我们从大豆基因组中获得了 Gy5 染色体基因。这也是国内外首次正式报道 Gy5 基因序列。

虽然 Gy5 和 Gy4 基因同属于第 II 类群,其 cDNA 同源性高达 80% 以上,但内含子 2 的长度却差异很大。另外,Gy5 基因的 3-端有两个多聚胸腺苷酸加尾信号(AATAAA),

Gy1~Gy3 也同样具有 2 个加尾信号,但 Gy4 基因却有 3 个尾信号。

大豆球蛋白的表达只在种子形成的胚乳期大量表达,这种组织特异性和发育阶段特异性表达是受大豆球蛋白基因启动子的调控。大豆球蛋白基因启动子除具有豆类作物贮藏球蛋白所具有的 Legumin box^[13]调控序列外,Itoh 等人^[14]研究发现在第 I 类群 Gy2(A₂B_{1a})基因的启动子中含有 Glycinin box(TAATAATTT),参与调控球蛋白基因的表达。为深入研究 Gy5 基因启动子的调控序列,分离 Gy5 基因启动子的研究工作正在进行之中。

参 考 文 献

- [1] Kitamura K, Takagi T, Shibasaki K. *Agric Biol Chem*, 1977, 41:833~840
- [2] Turner N E, Thanh V H, Nielsen N C. *J Biol Chem*, 1981, 256:8756~8760
- [3] Staswick P E, Herrnodson M A, Nielsen N C. *J Biol Chem*, 1981, 256:8752~8755
- [4] Nielsen N C, Dickson C D, Cho T J *et al.* *Plant Cell*, 1989, 1:313~328
- [5] Cho T J, Davies C S, Nielsen N C. *Plant Cell*, 1989, 1:329~337

- [6] Scallan B, Thanh V H, Floener L A *et al.* *Theor Appl Genet*, 1985, **70**: 510 ~ 519
- [7] Fukazawa C, Momma T, Hirano H *et al.* *J Biol Chem*, 1985, **260**: 6234 ~ 6239
- [8] Sims T L, Goldberg R B. *Nucl Acids Res*, 1989, **17**: 4386
- [9] Thanh V H, Tumer N E, Nielsen N C. *Nucl Acids Res*, 1989, **17**: 4387
- [10] Cho T - J, Nielsen N C. *Nucl Acids Res*, 1989, **17**: 4288
- [11] Kitamura Y, Arahira M, Itoh Y *et al.* *Nucl Acids Res*, 1990, **18**: 4245
- [12] Sambrook J, Fritsh E F, Maniatis T. In: *Molecular Cloning: a Laboratory Manual*, 2nd Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- [13] Shirsat A H, Meakin P J, Gatehouse J A. *Plant Mol Biol*, 1990, **15**: 685 ~ 693
- [14] Ltoh Y, Kitamura Y, Arahira M, *Plant Mol Biol*. 1993, **21**: 973 ~ 984

Molecular Cloning and Nucleotide Sequence of Soybean Glycinin Gene Gy5(A₃B₄)

CHEN San-Feng

(Dept. of Microbiology, China Agricultural University, Beijing 100094)

Chikafusa Fukazawa

(Genetic Engineering Lab., Division of Applied Microbiology, National Food Research Institute, Tsukuba, Japan)

Abstract The glycinin gene family encoding the glycinin subunits in soybean plants is composed of at least 5 gene members, *i. e.*: Gy1 - Gy5. A genomic clone containing the Gy5 gene from a genomic library of cv. Williams was isolated by using Gy5 cDNA probe. The complete nucleotide sequence of this gene has been determined. It is 2819 bp long consisting of four exons and three introns. These exons and introns are as follows: exon1(292bp), intron1(358bp), exon2(263bp), intron2(425bp), exon3(645bp), intron3(485bp), exon4(351bp). The gene encodes 517 amino acids. This is the first time to report the complete Gy5 gene sequence from a genomic library. The accession number in the DDBJ database is AB003680.

Key words Glycinin, Gy5(A₃B₄) gene, exon, intron, DNA sequence