

雪花莲外源凝集素基因的克隆及序列分析

周岩 田颖川 陶慧中 莽克强

(中国科学院微生物研究所, 植物生物技术开放实验室 北京 100080)

外源凝集素 (lectin) 是自然界中广泛分布的一组蛋白质, 在多种生物中均有发现。外源凝集素为一类能特异地识别并可逆结合糖类复合物的糖基部分而不改变被识别糖基的共价结构的非免疫性蛋白^[1]。它对植物有很重要的生理作用。例如保护功能, 在植物生长的各个阶段以不同的方式保护植物免于害虫的侵害; 作为储藏蛋白, 在植物发芽和幼苗生长阶段, 裂解的外源凝集素为其提供氨基酸^[2]; 外源凝集素还可能参与细胞间的识别, 如在植物建立共生关系中, 根部的外源凝集素可能是宿主特异性的主要决定因素^[3]。

近来, 石蒜科 (*Amaryllidaceae*) 外源凝集素日益引起人们的兴趣。首先由于它们独特的糖结合特性已成功地用于糖结合物的纯化和特征研究^[4]。第二, 由于它们对人类和动物的巨细胞病毒及逆转录病毒, 包括人类免疫缺陷病毒具有强烈及选择性的抑制作用, 这类凝集素正成为艾滋病研究的一个重要工具^[5]。第三, 近来已发现雪花莲外源凝集素对某些嚼食和刺吸式昆虫有明显的毒性^[6], 因此人们希望将它们用于植物保护的目。

雪花莲 (*Snowdrop, Galanthus nivalis*) 外源凝集素对稻褐飞虱 (*Nilaparvata lugens*) 和绿叶蝉 (*Nephotettix cinctipes*) 有毒杀作用^[7-8], 还能抑制桃蚜 (*Myzus persicae*) 的生长^[5], 而其它所试的各种酶、酶抑制剂和外源凝集素对上述昆虫无效。雪花莲外源凝集素对某些害虫有很强的毒性作用, 而对人的副作用很小^[10]。通过基因工程获得表达该凝集素基因的转基因植物, 以期控制稻飞虱、蚜虫一类的汁吸性害虫, 在抗虫植物基因工程方面正引起人们极大关注。

雪花莲外源凝集素由多基因家族编码, 基因内无内含子。已知的雪花莲几个外源凝集素成熟蛋白都含有 105 个氨基酸, 它们的前体含有一个有 23 个氨基酸的信号肽和一个长度在 29~34 个氨基酸之间的 C 末端延伸 (它们分别在翻译过程中和翻译后被切除)。不同雪花莲外源凝集素前体具有高度同源的信号肽序列, 而 C 末端多肽具有异源性。成熟蛋白的氨基酸序列同源性较高^[11]。

本文报道从雪花莲基因组 DNA 通过 PCR 方法对雪花莲外源凝集素基因的分子克隆及序列分析的结果。

1 材料和方法

1.1 菌株和质粒

宿主菌 *Escherichia coli* DH5 α , 克隆载体 pUC19 为本实验室所存。LB 培养基用于细菌培养。

1.2 植物材料

雪花莲小苗由意大利 G. Minato 教授馈赠。

1.3 试剂

限制酶、PCR 试剂盒、Wizard Clean-Up System 购自 Promega 公司。连接酶购自 BRL 公司。T7 Sequencing Kit 购自 Pharmacia 公司。放射性同位素 α -³²P-dATP 为 NEN 公司产品。

1.4 引物的合成及 DNA 的提取

参考文献[11], 根据 LECGNA2 设计四个雪花莲外源凝集素基因的引物, 引物 1、2 为成熟蛋白基

该项研究得到国家 863 计划资助并得到国际科学和文化中心 (ICSC, 瑞士, 洛桑) 世界实验室部分资助。
本文于 1995 年 7 月 5 日收到。

因的引物,引物 3、4 为包括 5'-信号肽、3'-29 个氨基酸残基雪花莲外源凝集素前体基因的引物。从雪花莲中提取总 DNA 参考文献 [12] 进行。质粒 DNA 提取纯化参考文献 [13]。

1.5 PCR 扩增基因及克隆

以雪花莲总 DNA 为模板,PCR 扩增雪花莲外源凝集素基因,程序如下:

- (1) 95℃5min, 加 TaqDNA 聚合酶;
- (2) 94℃1min, 57℃1min, 72℃1min, 循环 30 次;
- (3) 72℃10min。

产物经电泳检查,用 T4DNA 聚合酶补平, Wizard DNA Clean-Up System 纯化后,与 Sma I 酶切的 pUC19 质粒用 T4DNA 连接酶连接。转化 *E. coli* DH5 α 后在含氨苄青霉素、IPTG 和 X-gal 的 LB 平板上挑选白色菌落。提取质粒 DNA 后,用限制酶酶切分析,选择有正确插入片段的克隆^[13]。

1.6 DNA 序列分析

按试剂盒说明,碱法制备双链 DNA,用双脱氧法测定 DNA 序列。

2 结果与讨论

2.1 雪花莲外源凝集素基因的扩增

引物 1 和 2 为成熟蛋白基因的引物,为了克隆的方便和有利于在植物中表达,在引物 1 中依次加了 Bam HI 限制酶切点, Kozak 序列^[14]和起始密码子,引物 2 增加了 XhoI 限制酶切点和终止密码子。引物 3 和 4 为包括 5'-信号肽, 3'-29 个氨基酸残基雪花莲外源凝集素前体基因的引物。引物 3 依次加了 BamHI 限制酶切点和 Kozak 序列。引物 4 增加了 XhoI 限制酶切点。引物序列如下:

Primer 1: GCGGATCCAACAATGGACAATATTTTGTACTCCGGT

Primer 2: TGCTCGAGTTATCCAGTAGCCCAACG

Primer 3: ACGGAATCCAACAATGGCTAAGGCAAGTCTCCTC

Primer 4: CGCTCGAGTTACTTTGCCGTCACAGGCTT

PCR 产物经电泳分析结果表明 P1、P2 扩增得到约 360bp 片段。P3、P4 扩增得到约 500bp 片段。与实验设计要扩增得到的片段大小相符(见图 1-I),分别命名为 GNA12、GNA34。这一结果又一次证明雪花莲外源凝集素基因中不存在内含子。

2.2 重组质粒的筛选和鉴定

将上述纯化的 GNA12、GNA34 片段平端连接到用 Sma I 酶解的 pUC19 中,转化 *E. coli* DH5 α 细胞,选择白色菌落,提取质粒电泳分析。XhoI、Bam HI 酶切检查(见图 1-II),分别筛选到四个含 GNA12 的重组质粒,命名为 pGNA12。同时筛选到二个含 GNA34 的重组质粒,命名为 pGNA34。

2.3 DNA 序列及同源分析

含雪花莲外源凝集素基因的 pGNA12、pGNA34 质粒,用碱法制备双链 DNA,经变性后用双脱氧法测定 DNA 序列,获得了 GNA12、GNA34 片段的完整序列。DNA12 的碱基序列除了人为加入的起始密码子和终止密码子外由 315 个碱基对组成。GNA12 碱基序列和 GNA34 内部编码成熟蛋白的碱基序列完全一致,与 Van Damme^[11]等人报道的 LECGNA2 编码成熟蛋白的碱基序列相比较同源率为 97%,比较由它们推导出的氨基酸残基同源率为 98%。GNA34 由 483 个碱基对组成,包括 26 个氨基酸残基的信号肽和 29 个氨基酸残基的 C 末端延伸,共 160 个氨基酸。GNA34 的信号肽比 LECGNA2 的信号肽多了 3 个氨基酸,这 3

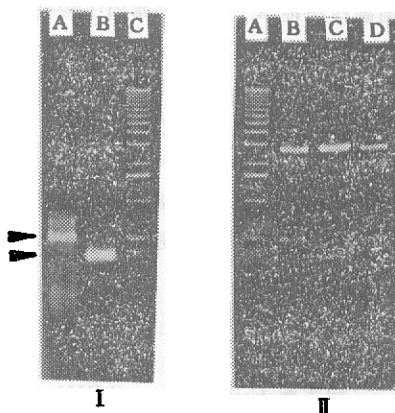


图 1 PCR 产物和重组质粒 (pGNA12、pGNA34) 限制性酶切的电泳图谱 I: A、B 分别为 P3、P4 和 P1、P2 的 PCR 产物,箭头表示扩增的目的基因。C 为标准分子量 DNA (1kb marker)。II: A 为标准分子量 DNA (1kb marker), B、C、D 分别为重组质粒 pGNA34、pGNA12、克隆载体 pUC19 加 XhoI+BamHI

个氨基酸是由于在 GNA34 的信号序列中有一个 9 个核苷酸的正向重复序列 CACCATCTT 而形成 (见图 2)。9 个核苷酸的正向重复序列未导致阅读框的改变。增加的 3 个氨基酸残基为 S、P、S, 是难于形成螺旋结构的氨基酸, 能很好地作为信号肽的 C 区。按照预测信号序列加工位点的方法^[15,16], 预测它的信号肽剪切位点仍然不变, 位于 DNI 之前。增加的 3 个氨基酸残基仅导致信号肽 C 区较 LECGNA2 长。我们筛选到的 2 个 pGNA34 克隆中 9 个核苷酸的正向重复序列都存在, 碱基序列完全一样。看来不可能是 PCR 扩增时造成的偶然差错。与 LECGNA2 比较, GNA34 的 C 末端延伸的氨基酸残基中有二个氨基酸残基不同, 它们在翻译后被切除。除了 9 个核苷酸的正向重复序列外, GNA34 有 14 个核苷酸与 LECGNA2 不同。比较 GNA34 与 LECGNA2 外源凝集素前体基因的碱基序列和氨基酸序列, 同源性均为 95%。GNA12 的碱基序列和 GNA34 编码成熟蛋白的碱基序列完全一样。加之, 雪花莲外源凝集素由多基因家族编码, 所以可以认为 GNA34 不同于以往克隆到的基因, 可能是其多基因家族中新的一员。

(1)	ATGGCTAAGGCAAGTCTCCTCATTTTGGCCGCCATCTTCCTTG	43
(2)	-----	43
(1)	GTGTCATCACACCATCTTCACCATCTTGTCTGGGTGACAATATTTTGTACTCCGGTG	100
(2)	-----C-----A-----	91
(1)	AGACTCTCTCTACGGGGGAATTTCTCAATTACGGAGGTTTCGTTTTATCATGCAAG	157
(2)	-----A-----C-----A-----	148
(1)	AGGACTGCAATCTGGTCTTGTACGACGTAGACAAGCCGATCTGGGCAACAAACACAG	214
(2)	-----G-----A-----	205
(1)	GTGGTCTCTCCCGTAGCTGCTTCCTCAGCATGCAGACTGATGGGAACCTCGTGGTGT	271
(2)	-C-----	262
(1)	ACACCCCATCGAACAAACCGATTGGGCAAGCAACACTGGAGGCCAAATGGGAATT	328
(2)	---A---	319
(1)	ACGTGTGCATCCTACAGAAGGATCGGAACGTTGTGATCTACGGAAGTATCGTTGGG	385
(2)	-----A-----T-----	376
(1)	CTACTGGAAGTCTCGCACCAGGAGCTGTTGGAATTCCTCGCATCGCCACCCTCAGAGAAAT	442
(2)	-----A-----CT-----	433
(1)	ATCCTACTGCTGGAAAGATAAAGCTTGTGACGGCAAAGTAA	483
(2)	-----	474
(B)		
(1)	MAKASLLILAAIFLGVITPSSPSCLDNILYSGETLSTGEFLNYGGFVFIMQED	54
(2)	-----S-----	51
(1)	CNLVLYDVDPKIWATNTGGLSRSCFLSMQTDGNLVVYTPSNKPIWASNTGGQNGNYV	111
(2)	-----N-----	108
(1)	CILQKDRNVVIYGTDRWATGTRTGAVGIPASPPSEKYPYTAGKIKLVTA	160
(2)	-----H--L-----	157

图 2 雪花莲外源凝集素前体基因 DNA 序列 (A) 及由此推导出的氨基酸序列 (B)

(1) GNA34 基因 (2) LECGNA2 基因

下划线代表重复序列。小点代表相对位置的缺失。(一) 表示序列的一致。箭头表示信号序列 (肽) 和 C 末端序列 (多肽) 可能的加工位点。

雪花莲外源凝集素基因的成功克隆, 为利用该基因进行抗虫基因工程研究奠定了基础。我们已用该基因构建了植物表达载体, 通过农杆菌转化植物的工作正在进行中。

参 考 文 献

- [1] Goldstein I J, Hayes C E Adv Carbohydr Chem Biochem. 1978, 35: 127~140.
- [2] Chrispeels M J, Raikhel N V Plant Cell. 1991, 3: 1~9.
- [3] Diaz C, Melchers L S, Hooykaas P J J *et al.* Nature, 1989, 338: 579~581.
- [4] Shibuya N, Berry J E, Goldstein I J *et al.* Arch. Biochem. Biophys. 1988, 267: 676~680.

- [5] Balzarini J, Schols D, Neyts J *et al.* Antimicrob. Agents Chemother. 1991, 35: 410~416.
- [6] Hilder V A, Gatehouse A M R, Gatehouse J A *et al.* Third International Congress from Plant Molecular Biology. 1991. Tucson, AZ, USA.
- [7] Hilder V A, Brough C, Gatehouse A M R *et al.* In proceedings of the Brighton Crop Protection Conference: Pests and Diseases, Lavenham; Lavenham Press. 1992, 2: 731~740.
- [8] Powell K S, Gatehouse A M R, Hilder V A *et al.* Entomol. Exp. Appl., 1993, 66: 119~126.
- [9] Hilder V A, Powell K S, Gatehouse A M R. *et al.* Transgenic Res., 1995, 4: 18~25.
- [10] Pusztai A, Ewen S W B, Grant G. *et al.* Digestion, 1990, 46: 308~316.
- [11] Van Damme E J M, De Clercq N, Claessens F *et al.* Planta, 1991, 186: 35~43.
- [12] Manning K. Anal Biochem. 1991, 195: 45~50.
- [13] Sambrook J, Fritsh E F, Maniatis T. Molecular Cloning, 1989. Cold Spring Harbor Lab. Press, USA.
- [14] Lutcke H A, Chow K C, Mickel F S *et al.* EMBO, 1987, 6: 43~48.
- [15] Von Heijne G. Eur J Biochem. 1983, 133: 71~21.
- [16] Von Heijne G. J Mol Biol. 1985, 184: 99~105.

Molecular Cloning and Sequence Analysis of Snowdrop (*Galanthus nivalis*) Lectin Gene

Zhou Yan Tian Yingchuan Tao Huizhong Mang Keqiang

(Laboratory of Plant Biotechnology, Institute of Microbiology, Academic Sinica, Beijing 100080)

Abstract Snowdrop lectin genes were cloned from DNA extracted from snowdrop young leaves by PCR. Sequence analysis revealed that the cloned gene GNA12 (*Galanthus nivalis* agglutinin) which encode mature lectin and GNA34 which encode precursor lectin have a high homology with previously reported LECGNA2 gene. The sequence coding for mature lectin showed 97% and 98% homology at nucleotide and amino acid level respectively. However a 9-base direct repeat in the sequence coding for signal peptide of GNA34 makes it different from other previously reported LECGNA genes.

Key words Snowdrop (*Galanthus nivalis* L.) lectin, molecular cloning, sequence analysis