

文章编号: 0529-6579 (1999) 03-0057-06

SINPV *p10* 基因特征的研究^{*}

魏永杰, 龙紫新, 郑 姬, 吴文言, 杨 洁

(中山大学昆虫学研究所、生物防治国家重点实验室, 广东 广州 510275)

摘 要: 从病毒感染后期的斜纹夜蛾幼虫脂肪体中提取 mRNA, 反转录合成 cDNA 探针, 并使之与 *SINPV* 基因组的酶切片段杂交, 将 *SINPV* *p10* 基因定位于 *Bam*HI -3.0 千碱基片段上. *SINPV* *p10* 是目前发现的最长 *p10* 基因; 基因的启动子内含有 2 个 ATTGTA 基序, 是目前所发现的 *p10* 基因中唯一含有 2 个 A TTTGTA 基序者. *SINPV* P10 蛋白含有 10 个七肽重复单元, 可形成 1 个相对较长的卷曲螺旋. *SINPV* *p10* 所在的 ORF552-*p10*-ORF945 基因簇与 *SpliNPV* 的 ORF552-*p10*-ORF945 基因簇各基因座位、方向及对应的氨基酸序列十分相似, 表明 *SINPV* 与 *SpliNPV* 有很近的亲缘关系.

关键词: 斜纹夜蛾核多角体病毒; *p10* 基因; ORF552; ORF945; 卷曲螺旋

中图分类号: Q 751 **文献标识码:** A

昆虫细胞感染核多角体病毒 (nucleopolyhedroviruses NPVs) 后期的显著特征为多角体蛋白及 P10 蛋白的大量合成. 多角体蛋白为多角体的主要组分, 而 P10 则与感染后期昆虫细胞核及细胞质中纤维结构的形成有关^[1]. 通过对 *Autographa californica* (*Ac*) MNPV P10 的缺失研究表明, P10 参与了细胞核的裂解及多角体的形态发生过程^[2]. 同多角体基因一样, *p10* 基因为杆状病毒复制的非必需基因, 因此研究者们也利用 *p10* 启动子组建杆状病毒表达载体以表达外源蛋白^[3]. 目前已报道的 *p10* 基因有 *Spodoptera littoralis* (*Spli*)NPV^[4]、*Spodoptera exigua* (*Se*)NPV、*Choristoneura fumiferana* (*Cf*)NPV、*Ae*NPV、*Orgyia pseudotsugata* (*Op*)NPV、*Perina nuda* (*Pn*)NPV (GenBank accession no. U50411) 及 *Bombyx mori* (*Bm*)NPV 的 *p10* 基因. 不同 *p10* 基因的氨基酸序列差异较大, 而它们的二级结构则有较高的相似性; 这表明 P10 蛋白的功能主要取决于蛋白的二级而非一级结构^[3].

斜纹夜蛾核多角体病毒 (*Spodoptera litura* nuclear polyhedrosis virus, *SINPV*) 能感染并杀死斜纹夜蛾, 对斜纹夜蛾有较好的防治效果^[6]. 为了解 *SINPV* 基因组的分子生物学, 并以此为基础构建高效的重组病毒杀虫剂, 本研究首次报道 *SINPV* *p10* 及其上下游基因的核苷酸序列, 并分析了 *p10* 基因的结构特征.

^{*} 基金项目: 国家自然科学基金 (39730030) 资助项目

收稿日期: 1998-04-03 作者简介: 魏永杰, 男, 1972 年生, 博士研究生 通讯联系人: 龙紫新
文中序列已被 GenBank/EMBL 数据库接受, Accession number 为 AF037263.

1 材料和方法

1.1 病毒株系及多角体纯化

SINPV 广州株由本室分离并保存. 用 PBS 缓冲液匀浆感染 *SINPV* 多角体的虫尸, 经数轮差速离心得到较纯净的多角体.

1.2 *SINPV* 基因组 DNA 的制备

在多角体悬液中加入等体积溶液 1 (100 mmol/L Na_2CO_3 , 100 mmol/L NaCl), 室温放置 30 min 后 2 000 r/min 离心 5 min. 上清加等体积溶液 2 (100 mmol/L Tris, 10 mmol/L EDTA, $w=2\%$ SDS, 20 mmol/L β -巯基乙醇, $w=54\%$ 蔗糖) 45 $^{\circ}\text{C}$ 保温 30 min, 酚/氯仿抽提, 乙醇沉淀后溶于 TE 缓冲液, 置 4 $^{\circ}\text{C}$ 备用.

1.3 *SINPV* 基因组 DNA 的酶切及转印

限制性内切酶 *EcoRI*, *XbaI*, *XhoI*, *BamHI*, *PstI*, *SacI*, *HindIII*, *SmaI* 购自 Promega 公司, 高相对分子质量 Marker 购自 GIBCO BRL 公司, SppI DNA/*EcoRI* Marker 为北京天象人生物试剂公司产品. 病毒基因组 DNA 经完全酶切并在 $w=0.5\%$ 琼脂糖凝胶中充分展开后, 用 Southern 转移方法, 印迹到 NC 膜上.

1.4 *SINPV p10* 基因的杂交定位

取感染 *SINPV* 第 7 天的斜纹夜蛾幼虫脂肪体, 利用 QIAGEN 公司 Oligotex Direct mRNA Kits 提取 mRNA. 以所提取的 mRNA 为模板, 以 Oligo (dT)_{12~18} 为引物, 用 Boehringer Mannheim Biochemica 公司的 DIG system 及 cDNA Synthesis Kit 反转录合成 cDNA 探针, 并与 1.3 所印迹的 NC 膜杂交. 杂交条件为 68 $^{\circ}\text{C} \times 8$ h; 洗膜条件为: $2 \times \text{SSC}$, $w=0.1\%$ SDS 室温漂洗 10 min, $0.1 \times \text{SSC}$, $w=0.1\%$ SDS 68 $^{\circ}\text{C}$ 漂洗 30 min.

1.5 基因组片段的克隆及序列分析

将杂交阳性片段 *BamHI* -3.0 千碱基对片段克隆到载体质粒 pUC18 上, 并对插入片段进行双链测序. 测序方法参考美国 Biochemical 公司的 Sequenase Kit 操作手册.

2 结 果

2.1 *SINPV p10* 基因的定位

在感染 NPV 的极晚期, 昆虫细胞内多角体基因及 *p10* 基因的 mRNA 占病毒总 mRNA 的 90% 以上. 本研究提取此期的 poly (A) RNA, 并以之为模板合成 DIG 标记的 cDNA 探针; 以合成的 cDNA 探针与 *SINPV* 基因组 DNA 的 *EcoRI*, *XbaI*, *XhoI*, *BamHI*, *PstI*, *SacI*, *HindIII* 酶解片段在严格条件下进行 Southern 杂交, *BamHI* -3.0 千碱基对片段有明显杂交信号且便于克隆. 鉴于 *SINPV* 多角体基因的 mRNA 不具备 poly (A) (结果将另文发表), 可判断 *p10* 基因位于 *BamHI* -3.0 千碱基对片段上.

2.2 *SINPV p10* 基因的上下游基因

对 *BamHI* -3.0 千碱基对片段的测序结果显示 (见 GenBank/EMBL, AF37263), 在 2 492 碱基对的序列中含有完整的 *p10* 基因, *p10* 基因的上下游分别有 552 碱基对和 945 碱基对的开放读码框 (open reading frame, ORF), 分别称为 ORF552 及 ORF945.

2.2.1 ORF552 位于 *p10* 基因 5' 端, 方向与 *p10* 基因一致, 编码 183aa 的蛋白质. 将其编码的蛋白质序列与 GenBank/EMBL 数据库已知的蛋白质序列相比较, 结果显示: 它与

SpliNPV 的 ORF552^[4]、*SeNPV* 的 xb187、*OpNPV* 的 ORF26 及 *AcNPV* 的 ORF2^[9] 所编码的蛋白质有一定的同源性, 同源性分别为 95%, 55%, 39%和 33%。这表明 *SINPV* 与 *SpliNPV* 的 ORF552 有很近的亲缘关系。

ORF552 的起始密码子 ATG 上游 22 nt 处有一晚期基因的转录起始基序 TAAG, 因此 ORF552 可能为一晚期基因。通过 Northern blot 分析, 证明 *SpliNPV* 的 ORF552 为一晚期基因, 在感染后 48 h 开始转录^[4]。*SINPV* ORF552 的 TATA 盒(TATA box)与 TAAG 位于同一位置上, 而终止密码子 TAA 则包含在其下游 *p10* 基因的 TAAG 基序中。在 TAA 下游 200 nt 处, 在 *p10* 基因的编码序列中有 AATAAA, 可能为 ORF552 的 poly(A)信号。ORF552 起始密码子 ATG 上游 68~47 nt 处有一特殊的 GGC 重复序列, 在 GenBank/EMBL 数据库尚未发现这段序列, 其功能有待进一步研究。

2.2.2 ORF945 位于 *p10* 基因的 3'端, 其起始密码子 ATG 位于 *p10* 基因的编码序列中, 因此它与 *p10* 基因互为重叠基因。*SpliNPV* 的 ORF945 与其 *p10* 也互为重叠基因^[8], 而已知的 *AcNPV* 表达基因中尚未发现存在重叠基因。ORF945 的 5'端非编码区没有发现 TAAG 基序与 TATA 盒的存在, 而其 poly(A)信号则与终止密码子 TAA 处在同一位置上。*SINPV* 的 ORF945 与 *p10* 方向一致, 编码为 314aa 的氨基酸序列只与 *SpliNPV* ORF945 的氨基酸序列有较高同源性, 达 77%, 而与 GenBank/EMBL 数据库已有的其它序列则没有任何同源性。

2.3 *SINPV* 的 *p10* 基因

SINPV *p10* 基因的 ORF 长 318 碱基对、编码 105aa 的蛋白质, 为目前发现的最长 *p10* 基因。在起始密码子 ATG 上游 67 nt 处有 TAAG 基序, 紧跟着 TAAG 下游 7 nt 及 22 nt 处各有 1 个 TATA 盒。TAAG 与起始密码子 ATG 之间被 2 个 ATTTGTA 基序分隔。A/TTTGTGTA 基序重复地存在于杆状病毒多角体基因的启动子内^[7], 而 *AcNPV*、*BmNPV* 和 *SeNPV* *p10* 启动子中也存在这种基序, 但不重复;*SpliNPV*、*CfNPV*、*OpNPV* 和 *PnNPV* 的 *p10* 启动子则缺乏这种基序(图 1a)。因此 *SINPV* 的 *p10* 基因是目前所发现的 *p10* 基因中唯一具有 2 个 A/TTTGTGTA 基序者。杆状病毒 A/TTTGTGTA 基序的功能尚不清楚。

在 *p10* 终止密码子 TGA 上游 9 nt 及下游 777 nt 处各有 1 个 poly(A)信号 AATAAA。*SpliNPV* *p10* 基因的 2 个 poly(A)信号的位置与 *SINPV* 的相似, 而且只有第 1 个 AATAAA 才是有效的加尾信号^[8]。究竟 *SINPV* *p10* 基因的哪个 AATAAA 才是有效的, 尚需利用 Northern Blot 加以确定。

2.4 *SINPV* P10 与其它 NPV P10 蛋白的同源性比较及其结构特征

SINPV P10 含 105 个氨基酸, 为目前发现的最长 P10 蛋白; 其相对分子质量为 11 266, 富含 Leu、Asn 及 Asp 残基。除与 *SpliNPV* P10 有很高同源性外(达 97%), *SINPV* P10 蛋白的氨基酸序列与其它 NPV 的有较大差异(图 1b)。*SINPV* P10 的 C 末端比 *SpliNPV* P10 多出 1 个 Asn 残基。*SINPV* 与 *SpliNPV* P10 蛋白比其它 P10 所多出的氨基酸残基并非杂乱排列, 而是集中于蛋白 N 端的 α 螺旋区域内, 这段 α 螺旋区域可形成 1 种叫卷曲螺旋的结构。卷曲螺旋的显著特征为具有七肽的重复单元^[8]。七肽重复单元的第 1、第 4 位为疏水性氨基酸, 而其它位置的氨基酸则大多是亲水性的; 2~3 个七肽重复单元结合在一起, 就形成了卷曲螺旋结构。*SINPV* P10 具 10 个七肽重复单元, 其中第 4 与第 5 个重复单元被 35 位的 Pro 残基所断开, 而且它们中间有 3 个氨基酸的滑动。同其它已知的 P10 蛋白一

样, *S/NPV* P10 的卷曲螺旋区段后紧接着 1 个 Pro 丰富区 (图 1b 中以单线标出), 其共有序列为 (E/A/V/L/I)P(D/E/N)(V/L/I/P)P. Pro 丰富区的后面则为碱性的 C 末端 (图 1b 中以双线标出).

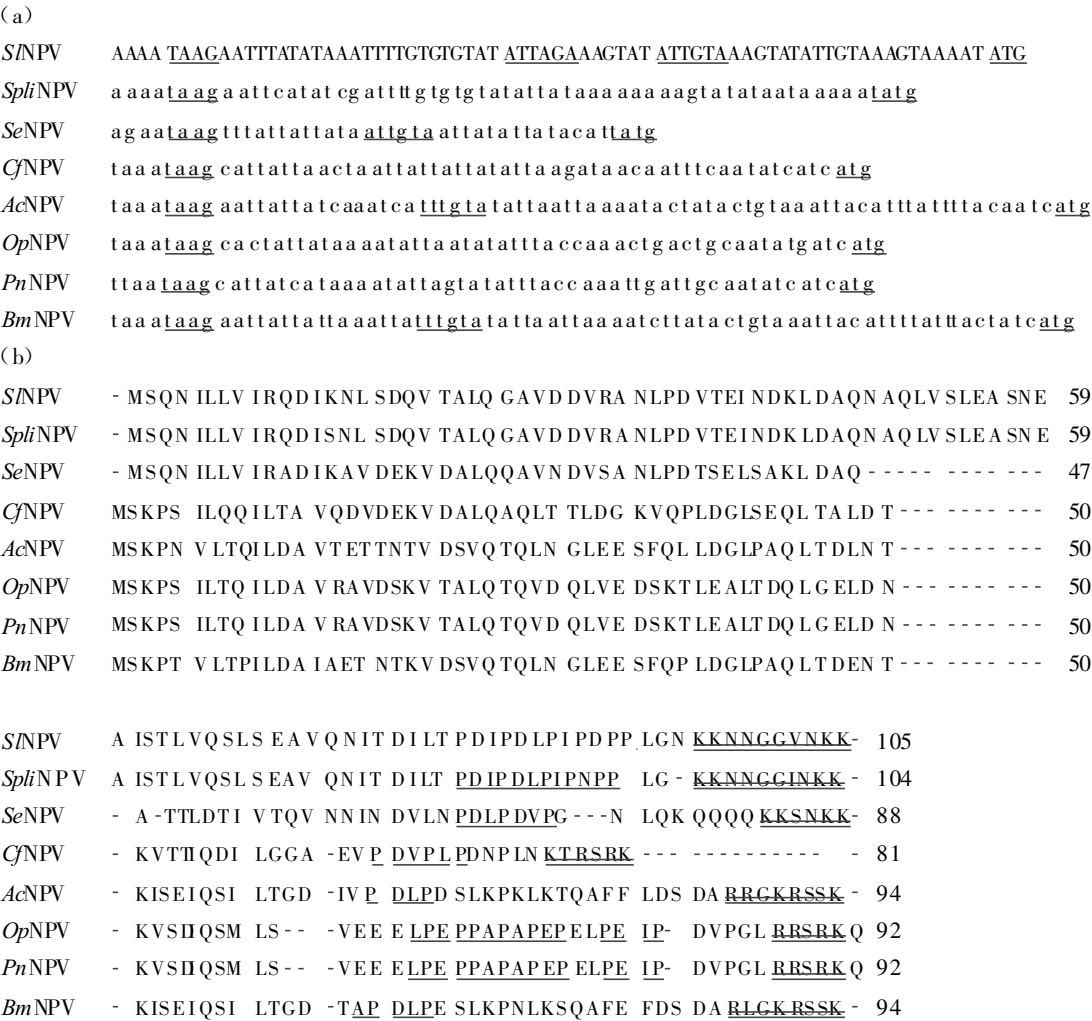


图 1 (a) 8 种 NPV 的 *p10* 基因启动子区域的比较; (b) 8 种 NPV 的 *p10* 基因氨基酸序列的比较

Fig. 1 (a) The promoter region of different *p10* gene species; (b) Alignment of P10

2. 5 *S/NPV* P10 蛋白的卷曲螺旋结构预测

为确定 *S/NPV* P10 蛋白形成卷曲螺旋的可能性, 利用 COILS 软件对其进行分析, 分析中所选用的窗口数为 $2l^{1/3}$. 结果显示, *S/NPV* P10 能形成一相对较长的卷曲螺旋结构. *S/NPV* P10 第 14 位带正电的 Lys 在 *SpliNPV* 的 P10 中被替换成为亲水性的 Ser, 这导致前 4 个七肽重复单元形成卷曲螺旋的可能性从 0.65 降到 0.48. 也就是说, 仅仅 1 个氨基酸的改变就使 *SpliNPV* P10 的前 4 个七肽重复单元不再可能形成卷曲螺旋结构.

3 讨 论

由于病毒感染后期多角体基因及 *p10* 基因的 mRNA 具有很高的丰度, NPV *p10* 基因的

杂交定位多采用此期的 cDNA 为探针. *S/NPV* 的多角体基因不具备 poly(A) 信号, 这就为 *p10* 基因定位带来了方便. *S/NPV* 的 *p10* 基因定位于基因组 *Bam*HI -3.0 千碱基对片段上, 此片段上含有 1 个 ORF552-*p10*-ORF945 的基因簇; 该基因簇各基因的座位与走向与 *SpliNPV* 的 ORF552-*p10*-ORF945 基因簇一致, 且对应基因的氨基酸序列也有很高的同源性, 表明 *S/NPV* 与 *SpliNPV* 具有很近的亲缘关系. 但是 *S/NPV* 的 ORF552-*p10*-ORF945 基因簇的组成与其它 NPV 的 *p10* 基因簇则有很大的差别, 这种差别可能是因为杆状病毒进化过程中基因组的重排所引起的.

曾发现在多角体基因 5' 端的非编码区内有 2 个保守的结构元件 A/TTTGTA 存在, 这种元件的序列与位置都与 5S RNA 及 tRNA 基因 5' 端调控区的共有序列一致^[7]. 5S RNA 及 tRNA 基因的 A/TTTGTA 基序参与结合 class II 转录因子. *S/NPV p10* 基因的启动子区内存在 2 个 ATTGTA, 是目前发现的 *p10* 基因中唯一具有 2 个 A/TTTGTA 基序者. *S/NPV p10* 基因具有 2 个可能的 poly(A) 信号, 因此它有可能产生互相重叠、长度不同的转录产物. *AcNPV*, *OpNPV*, *SeNPV*, *SpliNPV* 的 *p10* 基因也有利用不同的 poly(A) 信号, 产生相互重叠、超过 *p10* 本身长度的转录产物的现象. 重叠 mRNA 在杆状病毒中广泛存在, 这可能是病毒本身的一种表达调控机制.

不同的 P10 蛋白的氨基酸序列差异很大, 但它们的蛋白结构却十分相似. P10 蛋白的序列中有 3 个保守的结构域, 即 N 端的卷曲螺旋、碱性的 C 末端以及 1 个 Pro 丰富区. 卷曲螺旋与蛋白质之间形成寡聚体有关, 所形成的寡聚体可能是 P10 纤维结构的前体^[8]. *S/NPV P10* 具有 10 个七肽重复单元, 能形成相对较长的卷曲螺旋. P10 的 Pro 丰富区可能与多角体从细胞核中的释放有关^[4]. 为研究 P10 蛋白碱性的 C 末端的功能, 将 *AcNPV P10* C 端的 8 个氨基酸缺失, 截短的 P10 能形成聚集体但不能像完整的 P10 一样形成有序的纤维结构^[2]. 而碱性的 C 端则与 P10 或宿主因子的负电氨基酸残基结合从而把聚集体连接成有序的纤维结构.

参考文献:

- [1] VLAKE J M, KLINKENBERG F A, ZAAL K J M, et al. Functional study on the *p10* gene of *Autographa californica* nuclear polyhedrosis virus using a recombinant expression a *p10* β -galactosidase fusion gene [J]. J Gen Virol, 1988, 69: 765~776.
- [2] van OERS M M, FLIPSEN J M T, REUSKEN C B E, et al. Functional domains of the *p10* protein of *Autographa californica* nuclear polyhedrosis virus [J]. J Gen Virol, 1993, 74: 563~574.
- [3] VLAKE J M, SCJOUTEN A, USMANY M, et al. Expression of cauliflower mosaic virus gene I using a baculovirus vector based upon the *p10* gene and a novel selection method [J]. Virology, 1990, 179: 312~320.
- [4] FAKTOR O, TOISTER-ACHITUV M, NAHUM O, et al. The *p10* gene of *Spodoptera littoralis* nucleopolyhedrovirus: nucleotide sequence, transcriptional analysis and unique gene organization in the *p10* locus [J]. J Gen Virol, 1997, 78: 2119~2128.
- [5] van OERS M M, FLIPSEN J M T, REUSKEN C B E, et al. Specificity of baculovirus *p10* function [J]. Virology, 1994, 200: 513~523.
- [6] 蒲蛰龙, 陈其津, 刘复生, 等. 斜纹夜蛾核型多角体病毒杀虫剂中试生产及应用 [A]. 全国生物防治学术讨论会论文集 [C]. 北京: 中国农科院生物防治所, 1991. 141~146.
- [7] de ZANOTTO P M, SAMPAIO M J A, JOHNSON D W, et al. The *Anticarsia gemmatilis* nuclear polyhedrosis

- virus polyhedrin gene region, sequence analysis, gene products and comparison [J]. J Gen Virol, 1992, 73: 1049~1056.
- [8] LUPAS A, van DYKE M, STOCK J. Predicting coiled-coils from protein sequences [J]. Science, 1991, 252: 1162~1164.

Characterization of the *p10* Gene of *Spodoptera litura* Nuclear Polyhedrosis Virus

WEI Yong-jie, LONG Qi-xin^{*}, ZHENG Ji, WU Wen-yan, YANG Jie

Abstract: The *p10* gene was localized within the 3.0 kb *Bam*H I fragment of the *S/NPV* genome by Southern hybridization. With a coding sequence of 318 nucleotides, corresponding to a protein of 105 amino acids, the *p10* gene of *S/NPV* is the longest *p10* gene identified so far in NPVs. The *S/NPV* *p10* gene is also distinct in having two A/TTTGTA motifs in the promoter region. With 10 heptad repeats, the *S/NPV* P10 protein is probably forming a relatively large coiled-coil structure. The organization of *S/NPV* ORF552-*p10* gene-ORF945 cluster is colinear with that of *SpliNPV* and their corresponding amino acid sequences of each ORF share high similarity. This indicates that *S/NPV* and *SpliNPV* belongs to a homologous baculovirus group.

Keywords: *S/NPV*; *p10* gene; ORF552; ORF945; coiled-coil