

棉铃虫核型多角体病毒 p40 基因序列分析*

武家才, 张传溪**

(浙江大学应用昆虫研究所, 浙江杭州 310029)

Sequence Analysis of the *Helicoverpa armigera* Nuclear Polyhedrosis Virus p40 Gene

WU Jia-cai, ZHANG Chuan-xi**

(Institute of Applied Entomology, Zhejiang University, Hangzhou 310029, China)

Abstract: The *Hind* III-H, J fragments of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus (HaSNPV) genome were cloned. By sequencing the terminus of the cloned fragments, an ORF encoding the HaSNPV p40 was identified. The coding region was 966 bp long and potential to encode a 36.6kD peptide. The HaSNPV p40 exhibits 98% identity with HzSNPV p40 at the nucleotide sequence. The two genes diverged significantly from BmNPV p40 and AcMNPV gp41. The predicted amino acid sequences have less than 45% homology between HaSNPV p40, HzSNPV p40 and BmNPV p40 or AcMNPV gp41, but four genes have a conserved hydrophilic domain. In this region they have 62% homology. Further more, we determined that *Hind* III-H fragment links with J fragment in the genome and the two fragments were analysed with four restriction endonucleases (*Bam*H I, *Eco*R I, *Hind* III, *Pst* I).

Key words: HaSNPV; p40 gene; Nucleotide sequence; Restriction pattern.

摘要: 克隆了棉铃虫 (*Helicoverpa armigera*) 单粒包埋型核型多角体病毒 (HaSNPV) 基因组 *Hind* III-H, J 片段, 其长度分别为 10 kb 和 6.7kb。通过对克隆片段末端测序, 得到 HaSNPV p40 基因全序列。p40 基因编码区全长 966bp, 预计可编码 36.6 kD 的多肽。HaSNPV p40 基因核苷酸序列与 HzSNPV p40 基因有 98% 的相同性。这两种基因与 BmNPV p40 (GenBank - L33180) 和 AcMNPV gp41 (GenBank - L22858) 的氨基酸序列相似性 43% 左右, 但四种蛋白对应于 HzSNPV p40、HaSNPV p40 的 28 ~ 277 氨基酸区域, 同源性高达 62%, 并且存在一个保守的亲水性结构域。进一步将在基因组中相邻的 *Hind* III-H, J 两片段用 *Bam*H I、*Eco*R I、*Hind* III、*Pst* I 四种限制性内切酶进行了分析。

关键词: 棉铃虫核型多角体病毒; p40 基因; 序列; 限制性内切酶图谱

中图分类号: Q939.4; S476 文献标识码: A 文章编号: 1003-5125(2001)03-0242-04

棉铃虫 (*Helicoverpa armigera*) 单粒包埋型核型多角体病毒 (HaSNPV) 属于杆状病毒科, 单粒包埋型核型多角体病毒亚属。HaSNPV 是调节棉铃虫种群数

量的一个重要生物因子, 作为杀虫剂具有对环境安全、不使害虫产生抗性等优点, 但天然的病毒杀虫剂存在着寄主范围窄、毒力缓慢等缺点, 所以有必要对

收稿日期: 2000-08-22, 修回日期: 2000-11-22

* 基金项目: 国家自然科学基金资助项目 (39770031, 30070032)

作者简介: 武家才 (1974-), 男, 硕士研究生, 研究方向为杆状病毒分子生物学。

** 通讯作者: 张传溪 (1960-), 男, 教授, 博士生导师, 研究方向为杆状病毒分子生物学与昆虫资源学。Correspondence author.

本文序列已被 GenBank 接收, 接收号为 AF295782

HaSNPV 进行分子生物学研究。

目前对 HaSNPV 基因组已进行了限制性内切酶分析^[1],分析了多角体蛋白^[2]、蛋白激酶(GenBank-U95055)、几丁质酶(GenBank-AF114795)、*egt*(GenBank-AF000009)、*Lef-2*(AF136502)、*p35*(GenBank-AF063105)等等基因的核苷酸编码序列,构建出了其所在片段的物理图谱。本文通过对 HaSNPV 若干 *Hind* III 片段测序分析,测定出了编码 *p40* 基因 ORF 的全序列,并定位了 *Hind* III-H、J 片段在基因组中的相对位置。*p40* 蛋白是包涵体病毒囊膜结构组份,仅存在于有囊膜的包涵体病毒中,胞外型病毒粒子和除去囊膜的病毒粒子中都不存在 *p40* 蛋白^[4],对这种结构蛋白的研究有助于深入了解杆状病毒的感染机制。

1 材料与方法

1.1 材料

HaSNPV C1 株由华中师范大学余泽华教授赠送,棉铃虫及人工饲料由浙江省农科院王强副研究员提供,限制性内切酶、连接酶、低熔点琼脂糖均为 GIBCO-BRL 公司产品,载体为 pSK+, 宿主菌 TG 1, P³²-dCTP 为 Amersham 公司产品, Nick Translation 试剂盒为 Promega 公司产品。

1.2 方法

用空斑纯化的 HaSNPV 感染棉铃虫三龄幼虫,收集病毒致死的虫体,根据文献提纯多角体、碱裂解法提取基因组 DNA^[1]。Southern 杂交、低熔点琼脂

糖凝胶电泳分离片段、目的片段克隆、原位杂交筛选阳性克隆、单双酶切构建物理图谱均按文献进行^[3]。测序由 Genecore 公司完成。测序结果通过 BLAST 在 GenBank 中进行同源性检索分析。

2 结果与讨论

2.1 HaSNPV 基因组部分文库的建立

用 *Hind* III、*Xba* I、*Pst* I 和 *Xho* I 四种限制性内切酶酶解 HaSNPV 基因组 DNA,对各种酶切混合片段进行随机克隆,用基因组 DNA 作探针原位杂交筛选出阳性克隆。对阳性克隆通过酶切、电泳分析,大小与基因组 DNA 酶切片段大小基本相符,其中 *Hind* III 片段阳性克隆可覆盖整个基因组 60% 左右。

2.2 *Hind* III-H、J 筛选及限制性内切酶酶切图谱构建

进一步对大小在 6.0kb 以上的 *Hind* III 片段阳性克隆进行 Southern 杂交分析、末端测序。通过将 *Hind* III-H、J 片段测序结果在 GenBank 中检索,鉴别出编码 HaSNPV *p40* 基因 ORF。HaSNPV *p40* 基因编码区 +57nt、+434nt 各存在一个 *Pst* I 和 *Hind* III 位点。限制性内切酶分析 H、J 片段大小分别为 10kb 和 6.7kb,用 *Bam*H I、*Eco*R I、*Hind* III、*Pst* I 四种限制性内切酶分析,该区域内包含了 *Bam*H I-F、E、B、*Eco*R I-D、K 和 *Pst* I-F、B、G、C 片段,具体图谱见下图。

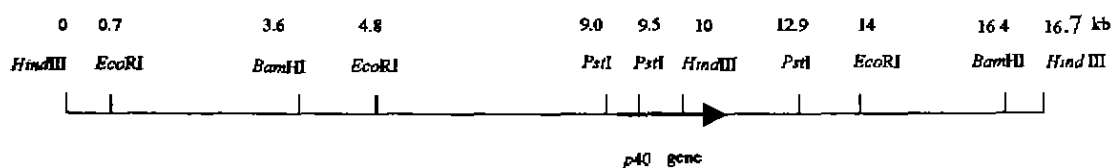


图1 HaSNPV *Hind* III-H、J 片段酶切图

p40 基因粗线表示,箭头表示转录方向。

Fig. 1 Restriction sites of HaSNPV *Hind* III-H、J fragments

The *p40* gene was denoted with broad line, the arrow represented the gene transcription direction

2.3 HaSNPV *p40* 基因序列及同源性比较

通过对 *Hind* III-H、J 片段末端测序,测出了 HaSNPV *p40* 基因的全序列(如图 2 所示),HaSNPV *p40* 基因编码区 966bp,推测编码蛋白分子量 36.6 kD,其编码方向和多角体基因相反。测序结果通过

BLAST 在 GenBank 中进行检索,与谷实夜蛾单粒包埋型核型多角体病毒(HzSNPV)结构蛋白 *p40* 基因有 98% 的相同性。其编码区上游 49nt 有一个典型的晚期启动子基序 ATAAG,已确定 HzSNPV 转录起始从 -49nt 的 ATAAG 基序开始^[4]。

```

GCATGCCAACCTGTCCTTAATAACCTATGACTATAGTTGTGCTCATATTTAATAATGATCGGAATTTCTTGGTTCT 75
ATATATTTTTTCATTATATTCATTTTGTATATTATTGTATGTGACACAATTATAIGATAAGTAAATCCGACGATAA 150
ATTACTGTCCACACAAATCGATAAATTATA                                     转录起始位点 180
                                                                    Pst I
ATG TCG CAG CCT CAC GCC GTC ACT ACA GCG TTA CAG CAT CAG CAG CAT CAA AAA CAG CTG 240
M S Q P H A V T T A L Q H Q Q H Q K Q L
CAG GAA TCT AGT TCG GAC GCT TGG ACC AAC AAA TGT GTC GAC TAC GTC GAA CGT ATC ATT 300
Q E S S S D A W T N K C V D Y V E R I I
AGA TTT TAT AGG ACA AAC GAT ATG TCG CAT TTG ACA CCG CAA ATG ATT ATG CTC ATT AAC 360
P F Y R I N D M S H L T P Q M I M L I N
ACT ATA CGC GAT TTG TGC GTG GAA TCG CAT CCG ATT AGC GTG AAC GTG GTC AAA CGG TTC 420
T I R D L C V E S H P I S V N V V K R F
GAT AGC GAT GAA AAT TTA ATT AAA CAT TAT TCA CGT TTG CGC AAA GAA CTG GGC GGT AGT 480
D S D E N L I K H Y S R L R K E L G G S
GAA GTT GCC GAA AAT ATA TTT CAG CCA TCA TTC GTG TAT AAT GTG TTG CCG AGT TAT GCC 540
E V A E N I F Q P S F V Y N V L P S Y A
CAA AAA TTT TAC AAC AAA GGC GCG GAA AAC GTG AGC GGC GAC AGC GTG TCS GAA GCG GCG 600
Q K F Y N K G A E N V S G D S V S E A A
                                                                    HpaIII
CAT GAA TTG GGC GAA GCT TTG CAA TAT CAA ATA GCC GAA GCT GTA GCC AGT AAT ACT CCG 660
H E L G E A L Q Y Q I A E A V A S N T P
ATT CCG TTG CCT GTT CGC CAT CAG CTC GTC AAC ACG TAT ATA ACG TTG CTG TTG CAA CGA 720
I F L P V R H Q L V N T Y I T L L L Q R
GCT AAC AIT CCG CCC AAC GTG CAA GAC GCT GTG TCC AGT CGC AAA TAC CCA ACA TTG AAC 780
A N I P P N V Q D A V S S R K Y P T L N
ATC ATC AAC GAT CTA ATC AAC AAT GTG ATT GAC GAC GTG TTC ACA GGC GTG TAT GGA AAC 840
I I N D L I N N V I D D V F T G V Y G N
TAT TAT TAT TAT GTA CTC AAT GAA AAG AAC CGC GCT CGC ATC GTT ACG CTC AAA GAG AAC 900
Y Y Y Y V L N E K N R A R I V T L K E N
ATA GGC TTT TTA GCG CCA TTG TCT GCG TCC ACG GAT ATA TTT CAA TAT ATT GCC AAT TTG 960
I G F L A F L S A S T D I F Q Y I A N L
GCT ACA CGC GCC GGC AAA AGA CCC AGC CTG GTT CAA GGA GCG ACG TTC CTC AAC GCT CCC 1020
A T P A G K R P S L V Q G A T F L N A P
TCT AGC AAC GGT TCT AAT GTA GAA CAA AAT CGT ACT TCT TGT CAG CAG AGT TTA ACA GAG 1080
S S N G S N V E Q N R T S C Q Q S L T E
TTA GCG TTT CAA AAC GAA GCG TTA CGT CGT TAC ATT TTT CAA AAG TTA AGC TAC AAA CAA 1140
L A F Q N E A L R R Y I F Q K L S Y K Q
AAT TAT TAA AAA AAA ATT TTA GCA TTA TAC ATT 1170
N Y >***
    
```

图 2 HaSNPV p40 核苷酸序列及推测的氨基酸序列

下划线标明转录起始位点、酶切位点,***表示终止密码子(TAA)。

Fig.2 Nucleotide sequence and predicted amino acid sequence of HaSNPV p40 gene

The putative transcription start site and enzyme sites is underlined,*** represents stop code (TAA)

已知 BmNPV-p40 (GenBank-L33180) 与 AcMNPV-gp41 (GenBank-L22858) 的核苷酸序列有很高的同源性, HzSNPV、HaSNPV p40 基因与这两种同源基因相差较大。BmNPV p40 与 AcMNPV gp41 基因翻译起

始密码子上游核苷酸序列(-51~-1)与 HzSNPV p40、HaSNPV p40 基因的相应区域有 60% 的同源性。HzSNPV、HaSNPV p40 基因前导序列中仅在-49nt 位点有一个晚期启动子基序 ATAAG, 而其它两种同源

基因在 -52 和 -21nt 位点各有一个晚期启动子基序 ATAAG^[5],但四种病毒在该基因转录起始位点(-50nt)附近核苷酸保守,另外启动子上游固定位置都有一个富含 A+T 的保守区域,这两个区域可能对基因的表达具有重要的意义^[6]。

HaSNPV p40 基因编码蛋白与 HzSNPV P40 蛋白仅有 4 个氨基酸不同。HzSNPV、HaSNPV p40 编码区比 BmNPV p40 和 AcMNPV gp41 小 100bp 左右。氨基酸序列相比,HzSNPV、HaSNPV P40 蛋白和其它两种蛋白的同源性 43% 左右。HzSNPV、HaSNPV P40 蛋白与其它两种蛋白比较,在氨基端缺失,羧基端延长了一段,四种蛋白共有的多肽链区域同源性达 62%。HzSNPV、HaSNPV P40 蛋白中的 14 个脯氨酸中有 11 个在其它两种同源蛋白中位置保守,此外 HzSNPV、HaSNPV p40 编码的 3 个半胱氨酸中有 2 个在其它两种同源蛋白中位置保守,这表明这四种蛋白的二级结构或三级结构保守,功能具有相似性^[5]。

2.4 P40 蛋白与病毒结构

核型多角体病毒在感染过程中复制形成两种表型的病毒体:胞外型病毒(ECV)和包涵体病毒(OV)。对 AcMNPV 研究表明:ECV 以出芽方式从细胞质膜获得带有自身编码的 gp64 蛋白的囊膜 OV 囊膜来源于类似于膜状的小泡,在宿主细胞核中由病毒基因合成,这种囊膜参与了多角体的形成,可以识别宿主中肠的微绒毛细胞^[6]。

目前对对病毒 P40 蛋白的确切功能还不明白。

对 AcMNPV、BmNPV 和 HzSNPV P40 蛋白 Western 杂交分析表明:P40 蛋白仅存在于有囊膜的 OV 中,ECV 和除去囊膜的病毒粒子中都不存在 P40 蛋白,推测 P40 蛋白是包涵体病毒囊膜结构组份或者是囊膜与核衣壳之间的一种成份^[4]。SDS-PAGE 分析 HzSNPV P40 蛋白分子量为 40kD,较之核苷酸序列推导的编码蛋白分子量要大,这可能是由于蛋白质进行糖基化修饰的缘故^[4],所以推测 HaSNPV P40 蛋白是一种糖蛋白,与病毒特异性感染宿主细胞有关。

参考文献

- [1] 王根,张传溪,金伟,等.棉铃虫核多角体病毒限制酶切图谱和多角体蛋白基因定位[J].环境与应用生物学报,1996,2(4): 387-392.
- [2] 王根,张传溪,金伟,等.棉铃虫核多角体病毒多角体蛋白基因的克隆和核苷酸序列分析[J].病毒学报,1997, 13(1):82-86.
- [3] 金冬雁,黎孟枫,等.分子克隆实验指南[M],1993,北京:科学出版社.
- [4] Shih-win Ma, Bartholomew G. Corsaro, et al. Cloning and Sequence Analyses of a p40 Structural Protein Gene of *Helicoverpa zea* Nuclear Polyhedrosis Virus[J]. Virology, 1993, 192: 224-223.
- [5] Whitford M, Faulkner P. Nucleotide sequence and transcriptional analysis of a gene encoding gp41, a structural glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus[J]. J Virol. 1992a, 66: 4763-4768.
- [6] Fraser M J. Ultrastructural observations of virion maturation in *Autographa californica* nuclear polyhedrosis virus infected *Spodoptera frugiperda* cell cultures [J]. J Ultrastruct Mol Struct Res 1986, 95: 189-195.