

银染色测定粘虫核多角体病毒多角体基因序列*

王家旺** 黄永秀 齐义鹏*** 杨复华

(武汉大学病毒学研究所 武汉 430072)

摘要 LsMNPV DNA 用 EcoRV 酶切进行基因组克隆, 用 AcMNPV 的部分多角体基因顺序 DNA 片段作探针, 菌落原位杂交法结合测序筛选到分别含 LsMNPV 部分多角体基因的重组质粒 pLsEV1 和 pLsPH5。用银染色 PCR 线性扩增双脱氧法测序, 发现 LsMNPV 的完整基因即位于这两个片段上。LsMNPV 多角体基因长 741bp, 编码区碱基同源性与 AcMNPV 和 MbMNPV 分别为 80.0% 和 97.0%, 氨基酸同源性分别为 87.8 和 97.5%。氨基酸组成中以谷氨酸 (Glu) 含量最高, 谷氨酰胺和色氨酸含量最低。密码子选用以第三个碱基为嘧啶的密码子频率最高。多角体蛋白 N 端有一类似信号肽结构的 26 个氨基酸的疏水区。

关键词 LsMNPV, 多角体蛋白基因, 序列分析

粘虫 (*Leucania separata*, Walker) 属鳞翅目夜蛾科, 是一种危害禾本科作物、果树、蔬菜和油料作物等的世界性暴食性害虫。粘虫核型多角体病毒 (*Leucania separata* Nuclear Polyhedrosis Virus, 简称 LsNPV) 早已用于粘虫的生物防治, 并进行了形态学等一般物理化学方面的研究, 包括 LsNPV 多角体蛋白的氨基酸组成成分的测定^[1], 其基因组 DNA 大小约为 160kb^[2], 核苷酸序列资料还未见报道。

由于多角体蛋白基因 (polh) 是杆状病毒的非必需结构基因, 且在极晚期能超水平表达, 表达蛋白可占细胞可溶性蛋白的三分之一以上, 因而引起了人们的很大兴趣, 近几年来对多角体基因进行了大量研究^[3]。研究多角体基因具有下列几方面的重要意义: 1. 利用多角体基因的高效启动子构建高效表达载体以大量表达外源有用蛋白。至今已成功利用 polh 基因启动子构建了几十种转移载体, 高效表达了数十种外源基因^[4]。2. 研制高效重组病毒杀虫剂。多年来, 昆虫杆状病毒由于具有特异性强和不污染环境等许多优点, 而一直被用于有效地防治某些害虫, 并成为目前最好的杀虫剂^[5], 但最大的缺点是作用速度太慢。利用多角体基因的高效启动子构建重组病毒杀虫剂已能缩短一些杀虫时间。这使人们深信, 随着研究的深入, 重组杆状病毒杀虫剂有可能发展成为代替化学杀虫剂的主要生物杀虫剂。3. 利用多角体基因序列资料研究杆状病毒的系统进化。杆状病毒是一大类变异较大的昆虫病毒, 基因组大小范围为 88~200kb, 形态学和血清学上差异较小^[6], 从这两方面进行分类较困难, 甚至是不可能的。因此有必要在 DNA 核苷酸和蛋白质氨基酸序列水平上进行分类。polh 基因作为杆状病毒最保守的基因之一, 也是杆状病毒中研

* 国家自然科学基金资助的课题。

** 现为中国医学科学院中国协和医科大学肿瘤医院所分子肿瘤学国家重点实验室博士后研究人员 (北京 100021)。

*** 联系与负责作者。

本文于 1994 年 10 月 19 日收到。

究得最多和最详细的基因,可作为研究杆状病毒分子进化的模型基因。4. 多角体基因超表达和时序表达机制的研究。虽然多角体基因已广泛用于高效表达外源蛋白,但其表达机制还很不清楚。5. 研究多角体基因与其它基因的关系。多角体基因在极晚期超水平表达,动用较多的转录和翻译因子和原料。许多杆状病毒基因也在此时终止表达。因此,多角体基因对其它基因应有较大的影响。确实发现 *polh* 基因的表达与其它基因有密切关系^[7]。6. 包涵体结构的研究。包涵体是晶体状物质,多角体蛋白是如何构成包涵体的还很不清楚。另外最近发现多角体蛋白有抗肿瘤活性^[1]。因此对 *polh* 基因进行序列测定,在研究基因的表达调控,杆状病毒的分子进化和多角体蛋白的功能等方面有着重要的意义,但目前对多角体基因的研究还不很深入。本文定位、克隆和序列分析了 *LsNPV polh* 基因,对 *polh* 基因的表达和结构等进行了初步探讨。

1 材料和方法

1.1 *LsNPV* DNA 的提取、纯化、酶切和连接转化

LsNPV DNA 的提取方法见前文^[7]。*LsNPV* DNA 用 *EcoRV* 酶切,酶切产物经酚-氯仿抽提和酒精沉淀后,取 1 μ g DNA 与 1 μ g pBluscript SK 质粒 DNA/*EcoRV*,于室温(25 $^{\circ}$ C)连接 8hr 以上。感受态细胞制备参照文献^[10](LB 培养基中 NaCl 不能高于 5g/L,否则制备不出感受态细胞):将单菌落接种于含 5ml 培养基 A(含 10mmol/L $MgSO_4 \cdot 7H_2O$ 和 0.2% 葡萄糖的 LB 培养基)的试管中,37 $^{\circ}$ C,300r/min,培养 2~3h,当可看到菌体时,取出放于冰水中 10min,将菌液倒入 1.5ml 微量离心管,室温离心 8000r/min,30s,去上清,用每管残余的液体将菌体悬开,每管加 100 μ l 冰冷的储存缓冲液 B(含 36% 甘油,12% PEG 7500 和 12mmol/L $MgSO_4 \cdot 7H_2O$ 的 LB 培养基)混匀即可用于转化。转化按常规方法进行。杂交滤膜制备:将菌落转到硝酸纤维膜上,于 37 $^{\circ}$ C 培养,当菌落生长到针尖大小时,用变性液变性 5min,中和 5min,于 80 $^{\circ}$ C 烘烤 2h 后,置于黑暗中备用。

1.2 多角体蛋白基因探针载体的构建和 DNA 放射性探针的制备

用 *HindIII* 酶切质粒 pUC8/6/8 (含 *AcNPV EcoRI*-I 片段,上有完整的 *polh* 基因),切下两条各长 930 和 1042bp 的片段。其中 930bp 片段含有 *polh* 基因部分编码区和 3' 端非编码区,电泳回收这两个片段,克隆进质粒 pBluscript SK 中。用 *KpnI* 筛选含 *polh* 基因的阳性克隆(因为 *polh* 基因具 *KpnI* 位点,另一片段无此位点),得到重组质粒 pAcHV。制备探针时,用 *HindIII* 酶切 pAcHV,电泳回收病毒 DNA,取 100ng 作模板,用 Promega 随机引物标记试剂盒和 ^{32}P 同位素标记。

1.3 杂交

用 30ml 预杂交液,68 $^{\circ}$ C,25r/min,预杂交 12h;倒出预杂交液,加入 8ml 杂交液,68 $^{\circ}$ C 杂交 24h。洗膜:2 \times SSC,0.1% SDS,室温 2 \times 15min;0.1 \times SSC,0.1% SDS,室温 2 \times 15min;0.1 \times SSC,0.1% SDS,55 $^{\circ}$ C,2 \times 15min。待膜洗到放射性残留 100cpm (计数/min) 时即可压 X 光片,-20 $^{\circ}$ C 放射自显影 12~24h。

1.4 序列测定

参照 Promega 银染色测序试剂盒使用说明书。

1.4.1 测序反应: 在每管标记 G, A, T, C 的 500 μ l 硅化的离心管中 (BioRad 产品) 各取 2 μ l 核苷酸混合液, 加入 20 μ l 矿物油。置于 4 $^{\circ}$ C 待用。在另一离心管中加入下列成分: DNA 4 μ g; Primer (正向引物为 5'AGGGTTTTCCAGTCACGAC3', Promega 测序试剂盒提供, 反向引物为 5'AAGCGGATAACAATTTACACAC3', 自己设计合成) 4.5pmol (3.6 μ l); 5 \times 反应缓冲液 5 μ l; Taq 酶 (5u/1 μ l) 1 μ l; 加超纯水至 16 μ l; 各取 4 μ l 分装于上述离心管中, 短时离心, 进行下列 PCR 反应: 95 $^{\circ}$ C 2min 后, 95 $^{\circ}$ C/30s, 42 $^{\circ}$ C/30s, 70 $^{\circ}$ C/60s, 共 60 个循环, 反应完后, 每管加 3 μ l 终止液, -20 $^{\circ}$ C 保存。

1.4.2 变性聚丙烯酰胺凝胶制备: 由 40% 丙烯酰胺母液稀释, 并在每 1L 稀释液中加 460g 尿素, 配制三种浓度 (4%, 5% 和 6%) 的丙烯酰胺储存液, 取 60ml 储存液, 真空抽气 15~20min, 加入 240 μ l 10% 过硫酸铵, 90 μ l TEMED, 灌胶 (长玻板和短玻板分别进行硅化和反硅化处理), 放置过夜即可。

1.4.3 点样和电泳: 用 GIBCO 序列分析仪, 75W 预电泳 0.5~1h, 使凝胶温度升到 55 $^{\circ}$ C 左右。上样前将样品在 70 $^{\circ}$ C 变性 2min, 每个点样孔上样 4.0~4.5 μ l。当第一次所点样品的溴酚兰跑出胶后 10~30min, 再上第二次样。二甲苯青跑到离胶底三分之一处即可结束电泳。

1.4.4 染色和显色: 2L 10% 醋酸固定 30min, 双蒸水洗 3 次, 每次 2min; 2L 染液 (1g/L AgNO₃), 染色 30min; 水洗 10s; 显色液 (10~12 $^{\circ}$ C, 30g/L Na₂CO₃, 0.05% 甲醛, 2mg/L Na₂S₂O₅ · 5H₂O) 显色 4~6min, 以带显出来而背景不太深为准; 将第一步回收的 10% 醋酸倒入, 终止 2min, 水洗 3 次, 每次 2min; 烘干后即可读序。

1.4.5 EDF 正片的制备: 将胶板置于普通 X 光片观察灯箱上, 于暗室中铺上 EDF (Kodak) 正片, 曝光约 20s (曝光时间视胶厚薄而变, 可从几十秒至几分钟不等), 用 D72 显色, F5 固定液固定, 水洗凉干即可。

2 结果

2.1 polh 基因的定位、克隆、亚克隆和序列分析

用 AcNPV DNA 部分 polh 基因作探针, 原位杂交筛选并结合测序鉴定得到含 900bp 和 700bp 的阳性重组子, 即 pLsEV 和 pLsPH5。用多种酶切 pLsEV 后, 发现其上具有 XhoI 和 HincII 单酶切位点, 亚克隆后, 正反向测序即可将整个片段测完, 计算机分析后发现 900bp 片段含有 polh 基因 3' 编码区 571bp 序列。polh 基因编码区一般长 741bp, 因此推测还有 170bp 的编码区和启动子区未测出。用正反向引物从 pLsPH5 两端测序, 结果准确测定了 LsNPV 的 polh 基因及其侧翼的核苷酸序列 (图 1)。DNASIS 分析后发现一个长 741bp 开放阅读框, 与 AcNPV 和 MbNPV 多角体基因分别具有 80.0% 和 97.0% 的碱基同源性以及 87.8% 和 97.5% 的氨基酸同源性。在氨基酸 33—36 位有一 NRKK 顺序, KRKK 在 AcMNPV 中是核定位信号 (图 2)。

在 5' 非编码区—45 位, 高度保守的 14bp Rohrmann 序列^[3] AATGTAAGTAATTT 中, 具有可能的转录起始位点核心序列 GTAAG^[14]; —66 位—77 位分别有两个可能的 TATA 盒 (TATAAC 和 TATAGG), —94 位为可能的 CAAT 序列 (CACAAAT), —28 位和—17 位是另两个保守序列 (TTCGTA 和 TTGTAA); 在 3' 非编码区未发现

GTACATGATGACACACAATGTGATGGCTTCGGTATAACCGACGGTATTTTATACGTTT	58
TTTGGGAAATGTAAGTAATTTTCTCCTTTCGTAGAAGATTGTAATAAATAAATAATAATG	118
M	1
TATACCCGTTACAGTTACAATCCGTCGTTGGGACGTACCTACGTCTACGACAACAAGTAC	178
Y T R Y S Y N P S L G R T Y V Y D N K Y	21
TACAAAAATCTTGGATCAGTCATAAAAAACGCCAACCGCAAAAAAGCACTTTATCGAACAT	238
Y K N L G S V I K N A N R K K H F I E H	41
EcoRV	
GAACTAGAGGAGAAAAACTCGACCCCTTTAGACAGATATCTGGTGGCCGAAGACCCCTTC	298
E L E E K T L D P L D R Y L V A E D P F	61
CTGGGACCGGGCAAAAAACCAAACTAACTTTGTTTAAGGAAATCAGATATGTTAAGCCC	358
L G P G K N Q K L T L F K E I R Y V K P	81
GACACCATGAAGCTTGTCTGTAAGTGGAGCGGTAAAGAGTTTCTCAGGGAAACTTGGACC	418
D T M K L V V N W S G K E F L R E T W T	101
HindIII	
HincII	
CGTTTCATGGAAGACAGCTTCCCCATTGTTAACGACCAAGAAGTCATGGACGTTTTCCTT	478
R F M E D S F P I V N D Q E V M D V P L	121
GTAATCAACATCGCTCCCACTAGACCCAACCGTTGTTTCAAGTTCCCTGGCTCAACATGCT	538
V I N M R P T R P N R C F K F L A Q H A	141
CTGCGTTGTGATCCCGACTACGTGCCCAACGAAGTCATCCGCATCGTTGAACGGTCTCTGC	598
L R C D P D Y V P H E V I R I V E R S C	161
GTGGGCACGAACAACGAATACCGCGTCAGCTTAGCCAAGCGTGGCGGTGGCTGCCCGGTG	658
V G T N N E Y R V S L A K R G G G C P V	181
ATGAATCTGCACTCTGAGTACCAACTCTTTTGAAGAGTTTCATCAACCGTGTATCTGG	718
M N L H S E Y T N S F E E F I N R V I W	201
GAGAACTTCTACAAGCCAATCGTGTACGTAGGCACAGATTCGGCTGAGGAAGAAGAAATT	778
E N F Y K P I V Y V G T D S A E E E E I	221
XhoI	
CTTCTCGAGGTTTCTCTTGTGTTCAAAATCAAAGAGTTTGGCGCTGATGCCCTCTCTAC	838
L L E V S L V F K I K E F A P D A P L Y	241
AACGGACCGGCATACTAAATAATTAATAACACAAAAATGATACGTAAGTATTTTATTGATT	898
N G P A Y	247
TTACCATTAGATAGTTGTGACTTTATTGATTGCAGTTTCTAGTTTCGTACAATAGTTTTTG	958
TATGGTGTCTTTATCGTTGGTGGTGTGCTGCCGACTGTTTAATGTGTCCGTTATAGTTTT	1018
AGATAGGGCTTGTATATTAGCAGGTATTTCCATTGGAAAACTGCAATTTTTTTTTTGGAA	1078
AACATTAAATTGATTTTTGTAATCAGCAATATCTTCAGGTTCCGCCAACCACTCTTCGGA	1138
ATTTTCGTTTCCATTGTTATCGCTAGAAGACGGGGCGTGAGATGTGCCGACGTGCAGTGT	1198
ATC	1201
EcoRV	

图1 银染法测定的 LsNPV 的 polh 基因及其侧翼的核苷酸和预测的氨基酸序列
预测的氨基酸用单字母表示。启动区调控顺序 CAAT 盒, TATA 盒, GTAAG 晚期启动子核
心顺序及主要酶切位点、多腺苷化识别位点用下线标示。

Fig. 1 Nucleotide sequence and predicated amino acid sequence of polh gene
region of LsMNVP

The predicted amino acid was shown by single letter. Regulation sequence e.g. CAAT
box, TATA box, late promoter sequence GTAAG and majore nuclease recognition
sequence, polyadenylation signal were underlined.

多腺苷化识别信号 AATAAA, 但离终止密码子 2 个碱基处有一类似 AATAAA 的序列

LsMNPV	MYTRYSYNPSLGRITYVYDNKYKYNLGSVIKNANRKKHFIEHELEEXTLDPLDKYLVAEDP	60
MbMNPVR.Y.....R.....	
AcMNPV	..-PD...R.TI.....A....K....A...I..A.....N.....	
LsMNPV	FLGPGKNQKLTLFKEIRYVKPDTMKLVVNWSGKEFLRETWTRFMEDSFPIVNDQEVMDVF	120
MbMNPVN.....	
AcMNPVN.....G.K....Y.....	
LsMNPV	LVINMRPTRPNRCFKFLAQHALRCDPDYVPHEFIRIVERSCVGTNNEYRVSLAKRGGGCP	180
MbMNPVF.....P.Y...S.....	
AcMNPV	..V.....D.....P.W...S....I....K.....	
LsMNPV	VMNLHSEYTNSPFEFIRVIVENPYKPIVYVGTDSABEEEILLEVSLVFKIKEFAPDAPL	240
MbMNPV	
AcMNPV	I.....Q..D.....I.....V.....	
LsMNPV	YTGPAY	246
MbMNPV	..N....	
AcMNPV	FT....	

图2 粘虫核多角体病毒与其它两种杆状病毒的多角体基因氨基酸同源性比较
“点”表示相同的氨基酸。

Fig. 2 Amino acid sequence identity comparison of three nuclear polyhedrosis virus polyhedrin
Dots show the same amino acids.

AATTAAA, 离终止密码子 207 个碱基处有 AAT9 (AATTTTTTTTTT) 序列, 此序列也存在于 ScMNPV、MbMNPV、PfMNPV 和 SfMNPV 多角体基因中, 并作为转录终止位点。AcMNPV 和 AgMNPV 多角体基因虽然转录也终止于 AT 丰富区, 但并不存在这样的结构。LsNPV 多角体基因下游的另一条链上发现了另一个 ORF, 其可能的多腺苷化识别信号, AATAAA 在 890bp 处。

2.2 粘虫核型多角体病毒多角体蛋白结构分析

2.2.1 氨基酸数目和氨基酸组成: 粘虫核型多角体病毒多角体蛋白由 246 个氨基酸组成, 精确分子量为 28981.70, 其氨基酸组成如表 1。

由表可见, 缬氨酸 (Val)、亮氨酸 (Leu) 和谷氨酸 (Glu) 含量较高, 分别为 8.54 mol%, 8.13 mol%, 8.94 mol%, 色氨酸 (Trp)、谷氨酰胺 (Gln) 和半胱氨酸 (Cys) 较少, 分别为 1.22 mol%, 1.22 mol% 和 1.63 mol%。

酸性氨基酸数目为 34 个, 碱性氨基酸数目为 33 个, 酸性氨基酸和碱性氨基酸含量基本相等, 即多角体蛋白基本为电中性的, 这对于形成的包涵体保持电中性使之稳定是有意

表 1 粘虫核型多角体病毒多角体蛋白的氨基酸组成*

Table 1 Amino acid composition of polyhedrin of LsMNPV

氨基酸 Amino Acid	残基数 Count (CNT)	残基分子量 Residue Weight (RW)	残基数×残基分子量 CNT × RW
Gly	11	57.05	627.55
Ala	9	71.07	639.63
Val	21	99.13	2081.73
Leu	20	113.15	2263.00
Ile	12	113.15	1357.80
Ser	11	87.07	957.77
Thr	11	101.10	1112.10
Cys	4	103.14	412.56
Met	6	131.19	787.14
Asp	12	115.08	1380.96
Asn	17	114.10	1939.70
Glu	23	129.11	2969.53
Gln	3	128.13	384.39
Arg	16	156.18	2498.88
Lys	17	128.17	2178.89
His	5	137.14	685.70
Phe	14	147.17	2060.38
Tyr	16	163.17	2610.72
Trp	3	186.20	558.60
Pro	15	97.11	1456.65
Asx	0	114.59	0.00
Glx	0	28.62	0.00
***	0	0.00	0.00
Xxx	0	118.88	0.00

* 总残基分子量 Total residue weight 28963.68;

多角体蛋白分子量 Molecular weight 28981.70.

义的。强疏水性氨基酸有 86 个,占 35%;强亲水性氨基酸有 74 个占 30.1%,两者共占 65%。在 20 种氨基酸中强疏水性氨基酸有 6 种,强亲水性氨基酸 5 个,共 11 种占氨基酸数的 55%;说明多角体蛋白中富含强亲水性氨基酸和强疏水性氨基酸;能形成转角的 Gly 和 Pro 含量也较高分别为 4.88mol% 和 6.10mol%。

2.2.2 遗传密码的运用: 在 64 个遗传密码子中,除 CGA(Arg)、TAG 和 TGA(终止子)外都被采用。其中下列密码运用频率较高: TTC-Phe, 71%; TAC-Tyr, 81%; CCC-Pro, 60%; CGT-Arg, 72%; ATC-Ile, 75%; AAC-Asn, 82%; GAA-Glu, 63%。可见多角体基因偏向于采用第三个碱基为嘧啶的密码子,且第三位 C>T, A>G; AAA, TTT, AA, TT 较少。

2.2.3 亲疏水性分析: 用 PROSIS 分析 LsNPV 多角体蛋白(图 3),发现两个有两个亲水区,一个在 40,另一个在 220 位置,还可看出亲水和疏水氨基酸数目差不多,其间交叉分布,间隔大致相等,共有四十几个区,每个区平均为 6 个氨基酸左右,分布在整个分子

表 2 粘虫核多角体病毒多角体基因遗传密码子的应用
Table 2 Codon usage of polyhedrin gene of LsMNPV

密码子 Codon	百分含量 %Average	密码子 Codon	百分含量 %Average	密码子 Codon	百分含量 %Average	密码子 Codon	百分含量 %Average
TTT-Phe	1.6	TCT-Ser	1.2	TAT-Tyr	1.2	TGT-Cys	0.8
TTC-Phe	4.0	TCC-Ser	0.4	TAC-Tyr	5.3	TGC-Cys	0.8
TTA-Leu	0.8	TCA-Ser	0.4	TAA-***	0.4	TGA-***	0.0
TTG-Leu	0.8	TCG-Ser	0.8	TAG-***	0.0	TGG-Trp	1.2
CTT-Leu	2.0	CCT-Pro	1.2	CAT-His	0.8	CGT-Arg	3.2
CTC-Leu	1.6	CCC-Pro	3.6	CAC-His	1.2	CGC-Arg	1.2
CTA-Leu	0.8	CCA-Pro	0.4	CAA-Gln	1.2	CGA-Arg	0.0
CTG-Leu	2.0	CCG-Pro	0.8	CAG-Gln	0.0	CGG-Arg	0.4
ATT-Ile	0.8	ACT-Thr	1.2	AAT-Asn	1.2	AGT-Ser	0.4
ATC-Ile	3.6	ACC-Thr	2.0	AAC-Asn	5.7	AGC-Ser	1.2
ATA-Ile	0.4	ACA-Thr	0.8	AAA-Lys	3.6	AGA-Arg	1.2
ATG-Met	2.4	ACG-Thr	0.4	AAG-Lys	3.2	AGG-Arg	0.4
GTT-Val	2.0	GCT-Ala	1.2	GAT-Asp	1.2	GGT-Gly	0.8
GTC-Val	2.8	GCC-Ala	1.6	GAC-Asp	3.6	GGC-Gly	2.0
GTA-Val	1.2	GCA-Ala	0.4	GAA-Glu	5.7	GGA-Gly	1.6
GTG-Val	2.4	GCG-Ala	0.4	GAG-Glu	3.2	GGG-Gly	0.4

注：***示终止密码子
Note: *** indicates termination codon

上。N端有一段 26 个氨基酸的疏水区，其中前 5 个氨基酸有一个碱性氨基酸 R 离 C 端 20 氨基酸左右也是疏水区。

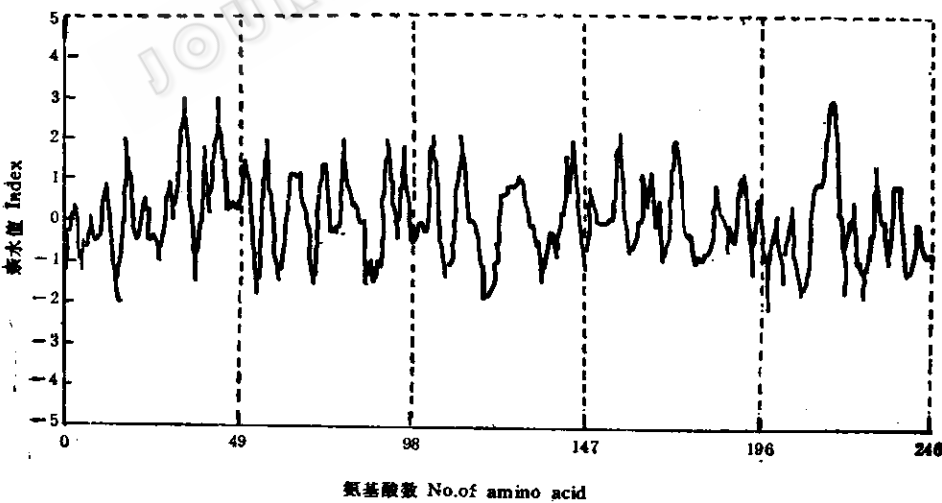


图 3 粘虫核多角体病毒多角体蛋白基因亲水性分析
Fig. 3 Predicted hydrophilicity of LsMNPV polyhedrin

3 讨论

杆状病毒多角体基因是高效表达基因，利用它成功地构建了能高效表达外源基因的

多种表达载体,其超表达机制很令人感兴趣。一般认为超表达的奥秘可能在启动子区^[1]。通过比较分析不同杆状病毒的多角体基因、颗粒体基因和另一高效表达基因 P10 基因的启动子,发现了 14bp 保守序列 Rohrmann 盒,可能与高效表达有关。一些研究者用连接物扫描 (Linker-Scan) 缺失法研究了多角体基因的启动区。由上游一直缺失到 12bp 保守序列时,对报道基因表达几无影响,而 12bp 保守序列缺失后,报道基因不能表达^[11,12]。Rohrmann 盒一般位于-47 到-53 位。启动子顺序位置的保守可能说明转录或翻译调控因子在几何上的需要^[13]。

另一与多角体蛋白高表达相关的因素,可能是密码子的选用。一般认为密码子的选用,与宿主细胞中相应的 tRNA 的浓度成平行关系。用 polh 基因驱动的外源基因的表达水平一般达不到 polh 基因的表达水平,部分原因可能是外源基因密码子的运用与昆虫细胞中相应的 tRNA 不符合。LsNPV 多角体基因表现出较大的密码子选用偏向:偏向于采用第三个碱基为胞嘧啶的密码子。AAA、TTT、AA、TT 选用较少,这可能是为了使病毒 DNA 在自然环境中尽量少受紫外线的损伤,因此尽量避免采用能形成 TT 二聚体的密码子。

多角体蛋白有两个特别的功能,一是形成保护病毒粒子的晶体状包涵体,二是抗溶解(除非是强碱性环境如昆虫的中肠,多角体蛋白是不可溶的)^[14]。这两个功能使病毒粒子在体外得到很好保护,能存活许多年,这种功能肯定具有其结构基础,据认为多角体蛋白是靠盐键形成晶体的^[14]。盐键在强碱条件下可以解离,故能解释多角体的碱可溶性。在 LsNPV 中,酸性氨基酸有 34 个,碱性氨基酸有 33 个。它们的个数基本相等。这表明一方面多角体蛋白基本上是电中性的,另一方面由多角体蛋白互相间通过酸性氨基酸带负电的侧链与碱性氨基酸带正电的侧链间形成离子键,中和了电荷,形成的包涵体也是电中性的。电中性对包涵体在环境中稳定可能是有意义的。LsNPV 多角体蛋白中含量较多的另一类氨基酸是强疏水性氨基酸,疏水键也是一种重要的维持分子间的力;半胱氨酸虽然少,只有 4 个,但因其侧链是唯一能形成化学键的氨基酸,因此除离子键外,疏水键和二硫键也可能在包涵体形成中起着重要作用。可能正是这三种键的作用,而使高度卷曲的多角体蛋白聚合成水不溶和电中性的晶体状包涵体。另一方面多角体蛋白富含能形成转角的 Pro 和 Gly,说明多角体蛋白是高度卷曲的,而且包涵体是不溶于水的。因此除形成分子间离子键的亲水性氨基酸外,其余的亲水性氨基酸可能都位于分子内部;Cys 也易在分子内形成二硫键。

亲疏水性分析表明,LsNPV 多角体蛋白没有大的亲水区和疏水区,亲水区和疏水区基本上是均匀相间排列的,同已测定序列的其它 20 种多角体和颗粒体蛋白一样,N 端为疏水区,且前 5 个氨基酸中有一个带正电荷氨基酸,与跨膜信号肽相似。而在已测序的 CPV 多角体蛋白中,虽 N 端也存在疏水区,但前 5 个氨基酸中没有带电氨基酸。亲水性分析也表明 CPV 多角体蛋白的亲水区和疏水区并不象 NPV 一样均匀相间排列^[15]。NPV 和 CPV 多角体蛋白都在细胞质中合成,后者在细胞质中包装病毒粒子,不需要跨核膜运输;前者在细胞核中包装病毒粒子,需要跨核膜运输。因此多角体蛋白的类信号肽序列和亲疏水区均匀相间排列,可能与跨膜运输有关。33 位和 42 位处于高度变异区,具有种的特异性,因此可用以制备抗体鉴别 NPV,214 位处于保守区,可能与多角体的基

本功能有关。

参 考 文 献

- [1] 刘子夜, 齐义鹏. 中国病毒学, 1993, 8(4): 357~365.
- [2] 黄永秀, 金天全, 李凌云, 等. 华中师范大学学报(自然科学版) 专刊, 1993, 2: 8~13.
- [3] Rormann G F. *J Gen Virol*, 1986, 67: 1499~1513.
- [4] King L A, Possee R D. The baculovirus expressing system, a laboratory guide. London UK, Chapman & Hall, 1991. 181~191.
- [5] 谭业平, 陶涛. 昆虫病毒学. 武汉: 武汉大学出版社, 1991. 405~414.
- [6] Rohrmann G F, Martigoni M E, Beaudreau G S. *J Gen Virol*, 1982, 62: 137~143.
- [7] Chaabihi H, Ogliastro M H, Marthin M *et al*. *J Virol*, 1993, 67(5): 2664~2671.
- [8] 王家旺, 黄永秀, 齐义鹏, 等. 中国病毒学, 1994, 9(2): 130~137.
- [9] Nishimura A, Morita M, Nishimura Y *et al*. *Nuclear Acids Res*, 1990, 18(20): 6169.
- [10] Matsuura Y, Possee R D, Orlerton H A *et al*. *J Gen Virol*, 1987, 68(5): 1233~1250.
- [11] 齐义鹏. 病毒学报, 1993, 6(4): 383~390.
- [12] Lewin B. Building the transcription complex. In: Lewin B ed. *Genes IV*. Cambridge, Massachusetts: Oxford University Press and Cell Press, 1990. 574.
- [13] Vlak J M, Rohrmann G F. The nature of polyhedrin. In: Maramorosch K & Shermen KE ed. *Viral Insecticides for Biological Control*. Oriando: Academic Press, 1985. 489~544.
- [14] Arella M, Lavallo C, Bellonck S *et al*. *J Virol*, 1988, 62(1): 211~217.

SEQUENCING OF POLYHEDRIN GENE OF LEUCANIA SEPARATA MULTIPLE CAPSID NUCLEAR POLYHEDROSIS VIRUS BY SILVER SEQUENCING SYSTEM

Wang Jiawang Huang Yongxiu Qi Yipeng Yang Fuhua
(Institute of Virology, Wuhan University, Wuhan 430072)

Abstract The genome of LsNPV was digested with EcoRV and cloned into pBluescript. With the partial polyhedrin gene of *Autographa californica* as a probe, positive recombinants were screened by colony in situ hybridization, further characterized by sequencing, and designated as pLsEV and pLsPH5. The intact 741 bp polyhedrin gene of LsNPV is located on the two recombinants revealed by Silver Sequencing System, and shares 80.0% and 97.0% nucleotide identity of the coding sequence, 87.8% and 97.5% amino acid identity with AcMNPV and MbMNPV respectively. Glu is the richest amino acid, and Gln and Trp the poorest amino acid in the polyhedrin of LsNPV. The codons ended with pyrimidine were higher in the codon usage. There was a signal-peptide-like hydrophobic region of 26 amino acids in the N end of polyhedrin.

Key words *Leucania separata* Multiple Capsid Nuclear Polyhedrosis Virus, Polyhedrin Gene, Sequencing