

印度木薯花叶双生病毒外壳蛋白的研究

I. 外壳蛋白及其基因的一级结构

洪益国 王小凤 田 波

(中国科学院微生物研究所, 北京 100080)

D. J. Robinson B. D. Harrison

(Scottish Crop Research Institute, U. K.)

在双生病毒 DNA 中存在保守序列 5' TAATATTAC3', 推测也存在于印度木薯花叶双生病毒 (ICMV) 基因组中。以此序列的互补链为引物, 合成 ICMV cDNA, cDNA 的主带大小为 2.7kb。测定了外壳蛋白的基因序列共 768 个核苷酸, 并分析了外壳蛋白的一级结构。ICMV 外壳蛋白由 20 种氨基酸的 256 个残基组成, 其中脂肪族氨基酸含量高, 碱性氨基酸含量高于酸性氨基酸, 含硫氨基酸和亚氨基酸的含量较低。比较 ICMV 与 7 种木薯粉虱 (*Bemisia tabaci*) 传播的双生病毒外壳蛋白, 其 N-末端变异程度大, 而 C-末端基本恒定, 中部存在 3 段较长的保守序列。

关键词 印度木薯花叶双生病毒; 外壳蛋白; DNA 序列

双生病毒是一类单链环状的 DNA 病毒, 含一个或两个基因组^[1]。Stanley 等^[2]首次报道了非洲木薯花叶病毒基因组的全序列, 迄今至少已完成了 17 种双生病毒分离株的 DNA 序列测定^[3]。双生病毒基因组中存在保守序列 5'TAATATTAC3'。在病毒侵染的寄主植物中, 其基因产物仅发现外壳蛋白^[4]。在大多数情况下, 纯化的病毒颗粒中, 通过 SDS-PAGE 也仅能检测到一种蛋白组分, 偶尔情况下可发现另一种小分子蛋白, 但其功能及来源不明^[5]。外壳蛋白的功能主要表现在: (1) 作为结构蛋白, 与病毒基因组 DNA 构成完整的病毒颗粒; (2) 与昆虫介体的特异性有关; (3) 与病毒的侵染性和病毒在植物中的扩散有关^[6,7]。作者研究了 ICMV 外壳蛋白基因克隆和序列分析, 在氨基酸序列水平上比较了 ICMV 与其他 7 种木薯粉虱传播的双生病毒外壳蛋白的差异, 为进一步研究双生病毒外壳蛋白的一级结构与功能的关系提供了分子生物学基础, 现报道如下:

材料和方法

(一) 载体和菌株

pUC19, M13mp18 和 M13mp19, 购自 Pharmacia 公司。DH5 α 和 DH α F 购自 BRL 公司。

(二) 酶和试剂

E. coli DNA 聚合酶 I 和 T4 DNA 连接酶系 BRL 公司产品。³²P 序列试剂盒系

Pharmacia 公司产品。[α - 32 P] dATP (3000Ci/mmol) 系 Amersham 公司产品。

(三) 病毒的纯化和病毒单链 DNA 的提取

参照文献 [8,9] 方法。

(四) 引物的制备

5'TAATATTAC3' 这段保守序列存在于所有已知序列的双生病毒基因组中。我们推测在 ICMV DNA 序列中也同样存在。据此设计出引物序列 5'GTAATATTA3', 在 ABI 391 DNA 合成仪上合成引物, 脱盐后溶于水中, 浓度 $1\mu\text{g}/\mu\text{l}$ 。

(五) cDNA 合成

引物与病毒 ssDNA 模板 (mol:mol = 100:1) 在 100mmol/L KCl, 30mmol/L Tris-HCl (pH7.5), 15mmol/L MgCl₂ 中 65°C 退火 10min, 加 $50\mu\text{mol/L}$ ATP, $25\mu\text{mol/L}$ dNTPs, 5u *E. coli* DNA 聚合酶 I, 1u T4 DNA 连接酶, $100\mu\text{Ci}$ [α - 32 P] dATP, 置室温下 2h, 加 10mmol/L EDTA 终止反应, 酚-氯仿抽提, 乙醇沉淀, 1.0% 琼脂糖凝胶电泳, 放射自显影检测 cDNA。

(六) 基因克隆

按文献 [10] 进行。

(七) 菌落杂交, 质粒制备和限制性内切酶分析

按文献 [10] 进行。

(八) M13 单链模板的制备和 DNA 序列测定

按文献 [11] 进行。

(九) 计算机处理序列

用 UWGCG 软件把 DNA 序列翻译成蛋白质的氨基酸序列, 并比较氨基酸的变异。

结果和讨论

(一) cDNA 合成

合成的引物 5'GTAATATTA3' 能与病毒 ssDNA 退火, cDNA 合成产率随时间延长而增高, 最高产率为 26%, 合成的主要产物大小约 2.7kb, 限制性内切酶 BamHI, EcoRI, HindIII, KpnI 及 PstI 均能酶切 cDNA 产物(图版 I-1)。选择 EcoRI 酶切 cDNA, 用于 DNA 克隆。

(二) 基因克隆及其鉴定

EcoRI 酶切的 ds cDNA 与脱磷 EcoRI 酶切 pUC19 连接, 转化感受态细胞 DH5 α , 菌落杂交(图版 I-2) 及小量制备纯化质粒经 EcoRI 分析, 得到了 39 个含病毒 DNA 序列的重组质粒, 插入片段大小为 0.15—2.5kb。

(三) 外壳蛋白基因及相应的氨基酸序列

亚克隆不同大小的插入片段于 M13mp18 或 M13mp19 中, 转染 DH5 α F', 提取单链 DNA 模板, 用 Sanger 双脱氧末端终止法测定序列, 外壳蛋白基因在一个 2.0kb 的插入片段中, 外壳蛋白基因全长序列如图 1a。通过 UWGCG 计算机程序转译为相应的氨基酸序列(图 1b), 证明外壳蛋白由 20 种氨基酸的 256 个残基组成(表 1)。外壳蛋白中脂肪族氨基酸 (G, A, V, L, I) 占 27.7%, 芳香族氨基酸 (F, Y, W) 占 10.9%, 羧基

侧链氨基酸 (S, T) 占 12.9%, 酰胺类氨基酸占 9.4%, 含硫氨基酸和亚氨基酸的含量较低, 分别为 5.5% 和 5.1%, 碱性氨基酸的含量明显高于酸性氨基酸, 前者为 20.3%, 而后者仅为 8.2%。

```

a ATGTCGAAGCGACCAGCAGATATCATAATTTCTACTCCAGGCTCGAAGGTTGCTGCGCGTCTGAACTTCGACAGCCCATACAGC
b M S K R P A D I I I S T P G S K V R R R L N F D S P Y S
a AGCCGTGCGGCTGTCCCTACTGTCCGCGTCACAAAAAGACAATCCTGGACAAACAGACCCATCAATCGGAAGCCAGGTGGTAT
b S R A A V P T V R V T K R Q S W T N R P I N R K P R W Y
a CGGATGTATAGAAGCCAGATGTTCTTAAGGGCTGTGAAGGCCATGTAAGGTCCAGTGGTTCGAGTCGAGACAGATGTGGTC
b R M Y R S P D V P K G C E G P C K V Q S F E S R H D V V
a CATATAGGTAAGGTCATGTGCATCTGTGATGTCACCTCGTGGAAATGGGCTTACACATCGGGTGGGTAAGAGGTTTTGCGTTAAG
b H I G K V M C I S D V T R G I G L T H R V G K R F C V K
a TCCATTTACATCCTGGGCAAGATATGGATGGATGAAAACATTAAGACCAAGAATCACAGCAATAGCGTAATGTTCTTCTCTGTGA
b S I Y I L G A K I W M D E N I K T K N H T N S V M F F L V
a AGGGATCGTAGGCCTGTTCATAAGCCCTCAGGATTTTGGTGAAGTATTAATATGTTGATAATGAGCCCACTACAGCTACCGTG
b R D R R P V D K P Q D F G E V F N M F D N E P S T A T V
a AAGAACATGCATCGTGATCGCTATCAAGTTCTCAGGAAGTGGCATGCCACGGTCACTGGTGGTCAGTATGCGATCAAGGAGCAG
b K N M H R D R Y Q V L R K W H A T V T G G Q Y A S K E Q
a GCTTTAGTTAGCGGTTTTTTTAGGGTTAATAAITATGTTGTGTATAACCAGCAAGAGCGCTGGGAAATATGAGAACCATACTGAG
b A L V R R F F R V N N Y V V Y N Q Q E A G K Y E N H T E
a AATGCATTAATGTTGTACATGGCGTGTACTCATGCCCTCTAATCCTGTATACCGTACGGTTGAAGATTAGAATCTATTCTATGAT
b N A L M L Y M A C T H A S N P V Y A T L K I R I Y F Y D
a TCAGTGAGCAATTAATAA
b S V S N * *
  
```

图1 外壳蛋白基因及其氨基酸序列

a: 核苷酸序列; b: 氨基酸序列

Fig. 1 Nucleic acid and amino acid sequences of coat protein

a: Nucleic acid sequence; b: Amino acid sequence

表1 外壳蛋白的氨基酸组成

Table 1 Amino acid composition of coat protein

氨基酸 Amino acids	数目 Number	氨基酸 Amino acids	数目 Number	氨基酸 Amino acids	数目 Number
A	12	R	26	N	16
D	12	C	5	Q	8
E	9	G	12	F	8
I	13	L	9	K	18
M	9	F	11	P	13
S	18	T	15	W	4
Y	13	V	25	总计 Total	256

70

ICMV	MSKRPADI I I	STPGSKVRRR	LNFDSPYSSR	AAVPTVRVTK	R-QSWTRRFI	NRKPRYRYMY	RSPDVPKGC
ABMV	MPGTSKTSRN	ANYSPRARI GPRVD	KASEWVHRPM	YRKPRY YRTL	RTADVPRGCE
ACMV	MSKRPGDI I I	STPGSKVRRR	LNFDSPYRNR	ATAPTVHVTN	RKRAWVNRPM	YRKPTMYRMY	RSPDI PRGCE
ACMVN	MSKRPGDI I I	STPGSKVRRR	LNEDSPYRNR	ATAPTVHVTN	RKRAWVNRPM	YRKPMYRMY	RSPDI PRGCE
BGMV	MAGTSKVSRS	GNYSPOGNGSKSN	KANAWVNRPM	YRKPRI YRMY	RSPDVPKGC
MYMV	MPKRNVDHRL	LYPDVECAEE	ADLRHPLSLP	ATAGSVPASA	KRRRWVNRPM	WRKPRYRYLY	RSPDVPKGC
TGMV	MPKRDAPWRL	MAGTSKVSRS	ANYSPRCSLP	KRDWVNRPM	YRKPRI YRSL	RSPDVPKGC
TYLCV	MSKRPQDI I I	STPYSKVRRR	LNFDSPYSSR	AAVPI VOGTN	KRSWVYRPM	YRKPRI YRMY	RSPDVPKGC

140

ICMV	GPCKVQSFE	RHDVYHI GKV	MCI SDVTRGI	GLTHRVGKRF	CVKSI YI LK	I WMDENI KTK	NHTNSVMFFL
ABMV	GPCKVQSYEQ	RHDI SHVGKV	MCI SDVTRGN	GLTHRVGKRF	CVKSI YI LK	I WMDENI KLO	NHTNSVMFWL
ACMV	GPCKVQSFEQ	RDDVXHLGI C	KVI SDVTRGP	GLTHRVGKRF	CI KSI YI LK	I WLDETI KKO	NHTNVI FYL
ACMVN	GPCKVQSFEQ	RDDVXHLGI C	KVI SDVTRGP	GLTHRVGKRF	CI KSI YI LK	I WMDENI KKO	NHTNVMFFL
BGMV	GPCKVQSYEQ	RHDI SHVGKV	MCI SDI TRGN	GI THRVGKRF	CVKSVYI LK	I WMDENI MLK	NHTNSVI FWL
MYMV	GPCKVQSFEA	KHDI SHVGKV	I CVYDTRGN	GI THRVGKRF	CVKSI YVTK	I WMDENI KTK	NHTNTVMFKL
TGMV	GPCKVQSYEQ	RHDI SLVGKV	MI CSDVTRGN	GI THRVGKRF	CVKSVYI LK	I WMDENI KTK	NHTNSVMFWL
TYLCV	GPCKVQSYEQ	RDDI KHTGI V	BCVSDVTRGS	GI THRVGKRF	CVKSI YPLK	WMDENI KKO	NHTNQVMFFL

210

ICMV	VRDRRP-VDK	PODFGEVFM	FDNEPSTATV	KNDLRDRYQV	LKRFHATVVG	GOYASKEQAL	VRRFFRVNNY
ABMV	VRDRRPYG-T	PMDFGHVFNM	FDNEPSTATV	KNDLRDRYQV	LKRFYKGVVG	GOYASNEQAI	VKRFWKVNNH
ACMV	LRDRRPYGNA	PODFGQI FNM	FDNEPSTATI	KNDLRDRFQV	LKRFHATVVG	GLYCWKEQAL	VKRFYRLNHH
ACMVN	LRDRRPYGN	PODFGQI FNM	FDNEPSTATI	ENDLRDRFQV	LKRFHATVI G	GPSGMKEQAL	VKRFYRLNHH
BGMV	VRDRRPYG-T	PMDFGQVFNM	FDNEPSTATV	KNDFRDRYQV	MHRFNKAVSG	GOYASNDQAL	VRRFWKVNNH
MYMV	VRDRRPF-G	PODFGOVFM	YDNEPSTATV	KNDLRDRYQV	VRKPOATVVG	GOYASKEQAL	VSKFYRVNNY
TGMV	VRDRRPYG-T	PMDFGQVFNM	FDNEPSTATV	KNDLRDRFQV	IHRFHAKVVG	GOYASNEQAL	VRRFWKVNNH
TYLCV	VRDRRPYGN	PMDFGQVFNM	FDNEPSTATV	KNDLRDRFQV	NRKFHATVI G	GPSGMKEQAL	VKRFYKINSH

260

ICMV	VYV..NQEA	GKYENHTENA	LMLYACTHA	SNPVYATLKI	RI YFYDSVSN
ABMV	VYV..NHQEA	GKYENHTENA	LLLYACTHA	SNPVYATI KI	RI YFYDSLNN
ACMV	VTY..NHQEA	GKYENHTENA	LLLYACTHA	SNPVYATLKI	RI YFYDSIGN
ACMVN	VTY..NHQEA	GKYENHTENA	LLLYACTHA	SNPVYATLKI	RI YFYDSIGN
BGMV	VYV..NHQEA	GKYENHTENA	LLLYACTHA	SNPVYATLKI	RI YFYDSITN
MYMV	VYV..NHQEA	AKYENHTENA	LLLYACTHA	SNPVYATLKI	RI YFYDSISN
TGMV	VYV..NHQEA	GKYENHTENA	LLLYACTHA	SNPVYATLKI	RI YFYDSITN
TYLCV	VTLFI FIOEA	AKYENHTENA	LLLYACTHA	SNPVYATMKI	RI YFYDSISN

图 2 外壳蛋白氨基酸序列比较

Fig. 2 Comparison of amino acid sequences of coat proteins

(四) 外壳蛋白的比较

用 UWGCG 软件系统比较 ICMV 与 10 种双生病毒外壳蛋白的相似性和统一性(表 2), ICMV 及木薯粉虱传播的 7 种双生病毒或分离株外壳蛋白的相似性为 84—87%, 统一性为 74—78%, 而与 BCTV, MSV 和 WDV 有明显的差异, 直接比较外壳蛋白氨基酸序列(图 2), 发现包括 ICMV 在内的 8 种木薯粉虱传播的双生病毒或分离株的外壳蛋白的 N-末端氨基酸残基变异较大, 而 C-末端约 40 个氨基酸残基几乎完全一致, 在中部存在三个较长的保守片段 GCEGPKVQS, THRVGKRFC 和 DNEPSTAT 及两个较短的保守片段 NHTN 及 RDRRP 这些序列可能与病毒传播的特异性介体有关。

表 2 ICMV 和其它双生病毒外壳蛋白相似性和统一性

Table 2 Similarities and identities of coat protein among ICMV and other geminivirus

	AbMV*	ACMVK	ACMVN	BGMV	MYMV	TGMV	TYLCV	BCTV	MSV	WDV
统一性(%) Identity	75.4	77.3	77.7	74.2	72.7	76.5	78.1	24.3	17.3	19.4
相似性(%) Similarity	84.6	87.1	87.1	85.0	84.0	86.2	86.3	49.1	40.3	41.4

* AbMV: Abutilon mosaic virus; ACMVK: African cassava mosaic virus, Kenya; ACMVN: African cassava mosaic virus, Nigerian; BGMV: Bean golden mosaic Virus; MYMV: Mung bean yellow mosaic virus; TGMV: Tomato golden mosaic virus; TYLCV: Tomato yellow leaf curl virus; BCTV: Beet curly top virus; MSV: Maize streak virus; WDV: Wheat dwarf virus.

参 考 文 献

- [1] Harrison, B. D.: *Ann. Rev. Phytopathol.*, 23: 55—82, 1985.
- [2] Stanley, J. et al.: *Nature*, 301: 206—262, 1983.
- [3] 洪益国: 博士论文, 第二部分, 中国科学院微生物研究所, 北京, 1990 年.
- [4] Stanley, J. et al.: *Nucleic Acids Res.* 14: 5981—5998, 1986.
- [5] Goodman, R. M. et al.: *Virology*, 106: 168—172, 1982.
- [6] Briddon, R. W. et al.: *ibid.*, 177: 85—94, 1990.
- [7] Gardiner, W. et al.: *EMBO J.*, 7: 899—904, 1988.
- [8] Sequeira, J. C. et al.: *Ann. Appl. Biol.*, 101: 33—42, 1982.
- [9] Robinson, D. J. et al.: *ibid.*, 105: 483—493, 1982.
- [10] Sambrook, J. et al.: *Molecular Cloning, A Laboratory Manual*. CSH, New York, 1989.
- [11] Sanger, F. S. et al.: *Proc. Natl. Acad. Sci. USA*, 74: 5463, 1977.

COAT PROTEIN OF INDIAN CASSAVA MOSAIC GEMINIVIRUS

I. PRIMARY STRUCTURES OF COAT PROTEIN AND ITS GENE

Hong Yiguo Wang Xiaofeng Tien Po

(*Institute of Microbiology, Academia Sinica, Beijing 100080*)

D. J. Robinson B. D. Harrison

(*Scottish Crop Research Institute, Dundee, U. K.*)

Geminivirus genomes have a conserved nine nucleotide sequence (5'TAATATTAC 3') that was hypothesised to exist in ICMV. Assuming this hypothesis to be correct, cDNA was synthesized by utilizing the complementary sequence as a primer (5' GTAATATTA 3') and the main product was about 2.7kb. cDNA molecules were cut with EcoRI, cloned in vector pUC19 which was used to transform DH5 α competent cells. By colony hybridization, thirty-nine recombinant plasmids were found. The sizes of inserts ranged from 0.15 kb to 2.5kb as shown by EcoRI analysis. Coat protein gene was sequenced and translated into amino acid sequences by UWGCG software. Coat protein consisted of 256 residues belonging to 20 species of amino acids. Direct comparison of the coat protein of ICMV with that of seven other whitefly-transmitted geminiviruses showed that the amino acid sequences of the N-terminal regions tended to be different, but the parts at the C-terminal were very similar.

Key words Indian cassava mosaic geminivirus; Coat protein; DNA sequence

图 版 说 明

Explanation of plate

1. 限制性内切酶分析 cDNA。C: cDNA 未酶切; B: BamHI; E: EcoRI; H: HindIII; K: KpnI; P: PstI。
2. 菌落杂交。→: 阴性对照。
1. Analysis of cDNA by restriction enzymes. C: cDNA, no treatment; B: BamHI; E: EcoRI; H: HindIII; K: KpnI; P: PstI.
2. Colony hybridization. →: Negative control.