

印度木薯花叶双生病毒外壳蛋白的研究 II. 计算机分析外壳蛋白的一些特性

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, Dundee, U.K.)

用 NWGCG 软件系统分析了印度木薯花叶双生病毒外壳蛋白的一些性质: (1) 外壳蛋白是碱性蛋白, 其等电点为 10.91; (2) 在外壳蛋白氨基酸序列中分布有 α -helices, β -Sheets 及 turns 等二级结构; (3) 分析了外壳蛋白中可能的表面氨基酸区域, 亲水性和抗原决定簇区域; (4) 在外壳蛋白的氨基酸序列中存在两个糖基化位点 HNT, 同时分析了双生病毒外壳蛋白氨基酸遗传密码的利用频率。

关键词 印度木薯花叶双生病毒; 外壳蛋白; 遗传密码; 计算机分析

外壳蛋白基因工程是植物病毒分子生物学研究的热门课题之一, 多种外壳蛋白的转基因植株具有抗病毒侵染的特性^[1]。双生病毒外壳蛋白的氨基酸序列具有较高的相似性^[2], 但对其生化性质的报道甚少, 木薯花叶双生病毒是木薯花叶病的主要病源, 引起粮食作物木薯严重减产。我们研究了印度木薯花叶双生病毒基因组在寄主植物中存在的形式、外壳蛋白基因和氨基酸的一级结构及基因组 DNA1 和 DNA2 的全序列^[2,3]。本文用 UWGCG 软件通过计算机程序分析了外壳蛋白的一些性质, 计算了双生病毒外壳蛋白中遗传密码的利用频率, 为进一步研究外壳蛋白的人工修饰和生物功能以及在遗传工程中的应用提供了重要的信息。

材 料 和 方 法

(一) 材料

外壳蛋白的氨基酸序列是我们测定的结果^[2,3]。

1	MSKRPADI II	STPGSKVRRR	LNFDSPYSSR	AAVPTVRVTK	RQSWINRPI N
51	RKPRWYRMYR	SPDVPKGCEG	PCKVQSFESR	HDVVHI GKVM	CI SDVTRGI G
101	LTHRVGKRFC	VKSI YI LGKI	WMDENI KTKN	HTNSVMFFLV	RDRRPVDKPO
151	DFGEVFNMF	NEPSTATVKN	MHRDRYQVLR	KWHTATVTGGQ	YASKEQALVR
201	RFFRVNYYVV	YNQQEAGKYE	NHTENALMLY	MACTHASNPV	YATLKI RI YF
251	YDSVSN**				

(二) 方法

UWGCG 是 University of Wisconsin Genetic Computer Group 的简称, 其计算机软件可用于核酸和蛋白质结构和性质分析, 是目前分子生物学中广泛使用的系统程序。

结 果

(一) 外壳蛋白的一些特性

1. 外壳蛋白: 木薯花叶双生病毒的外壳蛋白共由 256 个氨基酸残基组成, 其中含碱性氨基酸 52 个, 酸性氨基酸 21 个, 属碱性蛋白, 其等电点为 10.91(图版 I-A)。

2. 可能存在的二级结构: 二级结构分布有 α -helices, β -sheets 和 turns。CF 和 GDR 两种计算机软件分析预测的相似率约为 60%, 表明结果的可靠性, 在氨基酸序列中 β -sheets 的区域多于 α -helices 的区域, 两种结构形式都比较均匀的分布在序列中, 而 turns 多集中于 N-末端 75 个氨基酸残基之中, 在介于 110—190 氨基酸序列中亦有多数 turns 存在(图版 I-B)。

3. 亲水性和表面氨基酸: 很强的亲水氨基酸序列位于第 50 位、第 125 位、第 150 位和 230 位, 且亲水性区域明显多于疏水性区域(图版 I-C); N-末端是亲水性的, 而 C-末端却是疏水性的, 亲水性区域的氨基酸倾向位于外壳蛋白结构的表面(图版 I-D)。

4. 糖基化位点和抗原决定簇: 外壳蛋白中有两个糖基化位点(图版 I-B), 分别位于第 130 位和第 222 位, 是 NHT 三肽序列, 且都在外壳蛋白的表面。抗原决定簇多分布在 N-末端约 80 个氨基酸之中, 在中部(100—180 区)也存在与抗原决定簇有关的氨基酸序列, 值得注意的是 C-末端约 120 个氨基酸中仅含一个抗原决定簇区域(图版 I-E)。

(二) 外壳蛋白的氨基酸遗传密码

分析了印度木薯花叶双生病毒和其它 13 种侵染双子叶植物以及 7 种侵染单子叶植物的双生病毒或分离株外壳蛋白氨基酸遗传密码的使用频率(表 1)。具有两个遗传密码的氨基酸 D、N、Y 和 F, 其两个密码子在寄主相同的双生病毒中或同一密码子在寄主不同的双生病毒中的使用频率显著不同; C 的两个密码子(TGT, TGC), 在侵染双子叶植物的双生病毒中, TGT 的使用频率显著高于 TGC, 而在侵染单子叶植物的双生病毒中, 二者的使用频率基本一致; 其它具有两个密码子的氨基酸 E、K 和 Q, 其两个密码子在寄主相同的病毒中的使用频率均显著不同, 但就同一密码子在两类寄主中的使用倾向则无差异: 具有三个密码子的氨基酸 I, 四个密码子的氨基酸 A、T、P、G、V 和六个密码子的氨基酸 L、R、S, 其密码子的使用频率均有不同程度的差异。比较植物蛋白中的密码子, 对于同一氨基酸特定密码子的利用而言, 双生病毒和其侵染的寄主植物具有基本相同的倾向性, 另外外壳蛋白基因更多地利用 TAA 为终止密码子, 除 M 和 W 是单一密码子外, 双生病毒外壳蛋白与人蛋白氨基酸的密码子的利用有比较明显的差异。

表 1 外壳蛋白的氨基酸遗传密码使用频率

Table 1 Codon usage in coat protein genes of geminivirus

氨基酸 Amino acid	遗传密码 Codon	双生病毒 Geminivirus(%)		植物 [4] Plant(%)		人 [5] Human(%)
		侵染双子叶的 Infesting dicots n=14		双子叶 单子叶 Dicots Monocots		
		侵染双子叶的 Infesting dicots n=14	侵染单子叶的 Infesting monocots n=17	双子叶 Dicots	单子叶 Monocots	
G	GGG	17.9 ± 9.9	21.0 ± 7.3	12	21	24
	GGA	22.3 ± 13.4	26.3 ± 14.1	38	17	17
	GGT	30.5 ± 12.1	27.4 ± 16.5	34	18	15
	GGC	29.6 ± 10.5	25.4 ± 9.6	16	44	44
E	GAG	60.3 ± 17.1	67.7 ± 21.8	51	75	60
	GAA	39.9 ± 17.1	32.3 ± 21.8	49	25	40
D	GAT	75.6 ± 13.8	37.4 ± 9.2	58	27	38
	GAC	24.4 ± 13.8	62.6 ± 9.2	42	73	62
V	GTG	29.8 ± 12.1	42.3 ± 7.0	29	36	50
	GTA	9.9 ± 5.3	15.9 ± 9.7	12	8	9
	GTT	34.1 ± 12.3	21.7 ± 7.4	39	19	13
	GTC	27.4 ± 7.1	20.0 ± 5.2	20	37	27
A	GCG	13.6 ± 10.5	16.3 ± 1.9	6	22	12
	GCA	24.9 ± 12.8	13.1 ± 10.0	25	16	17
	GCT	31.6 ± 16.2	31.4 ± 13.4	42	24	31
	GCC	29.6 ± 7.9	39.9 ± 12.9	27	38	40
R	AGG	28.6 ± 13.7	17.0 ± 4.6	25	26	23
	AGA	16.1 ± 3.9	7.6 ± 4.5	30	9	24
	CGG	8.0 ± 3.3	24.3 ± 8.9	5	13	15
	CGA	12.5 ± 6.3	9.6 ± 5.6	8	4	10
	CGT	28.1 ± 11.4	18.7 ± 16.0	21	12	9
	CGC	6.8 ± 5.5	23.4 ± 7.2	11	36	19
S	AGT	18.5 ± 8.1	16.7 ± 8.7	14	8	11
	AGC	18.1 ± 8.2	16.4 ± 6.0	18	26	29
	TCG	16.4 ± 5.9	10.0 ± 4.0	6	14	7
	TCA	12.9 ± 9.3	9.1 ± 3.2	19	11	11
	TCT	20.2 ± 12.8	22.0 ± 12.6	25	15	17
	TCC	13.8 ± 11.0	26.1 ± 3.4	18	26	26
K	AAG	77.1 ± 8.9	83.9 ± 8.9	61	86	55
	AAA	22.9 ± 8.9	16.1 ± 8.9	39	14	45
N	AAT	54.8 ± 11.9	37.4 ± 11.7	45	25	34
	AAC	45.4 ± 11.7	62.8 ± 11.8	55	75	66
M	ATG	100	100	100	100	100
I	ATA	29.3 ± 11.5	23.6 ± 9.7	18	11	13
	ATT	38.9 ± 11.4	35.1 ± 15.3	45	24	23
	ATC	31.6 ± 13.7	41.0 ± 13.3	37	65	64
T	ACG	21.8 ± 6.3	22.6 ± 7.2	8	21	12
	ACA	22.5 ± 9.9	12.3 ± 8.6	27	14	21
	ACT	43.4 ± 9.4	25.9 ± 4.2	35	19	20
	ACC	15.2 ± 14.0	39.3 ± 7.3	30	46	47
W	TGG	100	100	100	100	100
END	TGA	0	14.3 ± 35.0	33	34	-
	TAG	0	14.3 ± 35.0	19	36	-
	TAA	100	71.4 ± 45.2	48	30	-
C	TGT	87.9 ± 13.4	52.9 ± 14.1	44	30	30
	TGC	12.1 ± 13.4	47.1 ± 14.1	56	70	70
Y	TAT	62.7 ± 15.4	35.6 ± 17.8	43	21	47
	TAC	38.0 ± 14.9	64.4 ± 17.8	57	79	53
L	TTG	40.0 ± 10.6	19.0 ± 5.6	26	14	0
	TTA	18.4 ± 13.8	4.7 ± 6.1	10	3	5
	CTG	17.6 ± 10.9	38.6 ± 11.2	9	28	46
	CTA	5.9 ± 7.1	6.6 ± 7.3	8	9	7
	CTT	12.9 ± 11.1	11.7 ± 6.1	28	15	11
	CTC	4.9 ± 7.75	19.0 ± 7.3	19	31	22
F	TTT	55.7 ± 24.3	31.4 ± 11.0	45	25	35
	TTC	44.3 ± 24.2	68.8 ± 10.8	55	75	65
Q	CAG	62.0 ± 22.2	75.7 ± 15.0	41	46	26
	CAA	38.3 ± 22.2	24.4 ± 14.9	59	54	74
H	CAT	68.1 ± 22.2	42.7 ± 28.0	54	33	42
	CAC	32.1 ± 22.2	57.3 ± 28.0	46	67	58
P	CCG	5.2 ± 9.2	26.1 ± 8.9	9	23	11
	CCA	31.5 ± 12.5	12.1 ± 10.9	42	34	24
	CCT	27.1 ± 13.0	39.5 ± 18.4	32	17	24
	CCC	36.5 ± 13.7	21.9 ± 10.2	17	26	41

讨 论

本病毒外壳蛋白为碱性蛋白, 二级结构中有 α -helices、 β -sheets、turns 以及亲水和疏水氨基酸基因的不同分布, 可能与病毒特异的双联体病毒颗粒的形成有关。同时抗原决定簇在氨基酸序列中的不同分布及糖基化点的存在, 说明外壳蛋白某些氨基酸残基顺序具有特殊的功能。

我们也用计算机分析了木薯粉虱 (*Bemisia tabaci*) 传播的其它 7 种双生病毒外壳蛋白 (数据未列出), 其性质与上述的相似; 外壳蛋白的等电点为 10.41-10.91; 分布有类似的二级结构; 抗原决定簇集中在 N 末端, 但分布有所差异, 由于不同双生病毒外壳蛋白 N-末端氨基酸的变化较大^[3], 这也在理论上说明了双生病毒血清学反应的差异; 此外外壳蛋白中都有两个糖基化位点。

双生病毒外壳蛋白和寄主植物蛋白对氨基酸遗传密码的使用具有相似的倾向性, 这可能是病毒侵染植物后利用寄主蛋白合成系统而对寄主适应的结果, 由于生物进化过程中, 生物的遗传背景和环境的选择压力, 使得各种生物对遗传密码子的使用频率显现差异, 密码子的利用频率为合理改造蛋白编码基因进行遗传工程的研究提供了依据。

参 考 文 献

- [1] Beachy, R. N.: Viral Genes and Plant Pathogenesis, p.13-22 (ed. Pirone T. P. et al.), Springer-Verlag Berlin Heidelberg, New York, 1990.
- [2] 洪益国: 博士论文, 第二部分, 中国科学院微生物研究所, 1990 年。
- [3] 洪益国等: 微生物学报, 33(2):98-103, 1993。
- [4] Murray, E. E. et al.: *Nucleic Acids Res.*, 17:447, 1989.
- [5] Lathe, R.: *J. Mol. Biol.*, 183:1, 1985.

COAT PROTEIN OF INDIAN CASSAVA MOSAIC GEMINIVIRUS

II. SOME PROPERTIES OF COAT PROTEIN ANALYZED THROUGH COMPUTER SOFTWARE

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.*)

Some properties of coat protein were predicted by computer analysis (UWGCG). The results showed that (1) coat protein is a basic protein, the pI of which is 10.91; (2) there are secondary structures along the coat protein amino acid sequence including α -helices, β -sheets and turns; (3) some special regions of amino acids of coat protein are characterized by different surface locations, hydrophilicity and antigenic determinants; (4) two glycosylation sites NHT are present in the coat protein sequence. Besides, the frequencies of amino acid codon usage in coat protein of geminiviruses was also analyzed through UWGCG software.

Key words Indian cassava mosaic geminivirus; Coat protein; Codon usage; Computer analysis

图 版 说 明

Explanation of plate

A. 外壳蛋白的等电点曲线, *: pI; B. 外壳蛋白中的二级结构和糖基化位点; C. 外壳蛋白的亲水和疏水性, 红色: 亲水性, 蓝色: 疏水性; D. 外壳蛋白的表面区域 (红色); E. 外壳蛋白的抗原决定簇 (红色); C, D 和 E 的曲线中, 正弦波: α -helices, 锯齿波: β -sheets, 180° 转折: turns.

A. Isoelectric point of coat protein, *: pI; B. Secondary structures & glycosylation sites of coat protein; C. Hydrophilicity (red) and hydrophobicity (blue); D. Surface-region (red) of coat protein; E. Antigen determinants (red) of coat protein. In C, D and E, α -helices shown with a sine wave, β -sheets with a saw tooth wave, turns with 180 degree turns.