The complete nucleotide sequence of pea seed-borne mosaic virus RNA

E. Johansen,[1] O. F. Rasmussen,[2] M. Heide[3] and B. Borkhardt[1]*

[1] *The Danish Research Service for Plant and Soil Science, Lottenborgvej 2, DK-2800 Lyngby,* [2] *Genetic Engineering Group, Building 227, The Technical University of Denmark, DK-2800 Lyngby and* [3] *Plant Protection Division, Novo Nordisk A/S, Novo allé, DK-2880 Bagsværd, Denmark*

The complete nucleotide sequence of the RNA genome of pea seed-borne mosaic virus (PSbMV) was determined from cloned cDNA and by direct sequencing of viral RNA. The PSbMV genomic sequence was determined to be 9924 nucleotides in length excluding the poly(A) tract. The RNA contained an open reading frame (ORF) of 9618 nucleotides with the potential to encode a polyprotein with a calculated M_r of 364000 (364K). The ORF was flanked by a 5' untranslated leader sequence of 143 nucleotides and a 3' untranslated region of 163 nucleotides. A comparison of the PSbMV polyprotein with the polyproteins of the potyviruses tobacco etch virus (TEV), tobacco vein mottling virus (TVMV), plum pox virus (PPV) and potato virus Y (PVY) showed that PSbMV had a similar genome organization. The polyproteins had a high level of amino acid identity except in the N-terminal region, which varied in both sequence and length. Putative proteolytic cleavage sites were identified in the polyprotein of PSbMV by comparison with those identified for other potyviruses. The cleavage site between the 6K protein and the 49K proteinase is proposed to occur at the C-terminal side of glutamic acid and not at the C-terminal side of glutamine as in other potyviruses. In addition to the five proteolytic cleavage sites for the 49K proteinase identified previously, a sixth putative cleavage site was identified internally in the 49K proteinase of PSbMV, as well as in the 49K proteinases of TEV, TVMV, PPV, PVY and soybean mosaic virus. Cleavage at this site in the 49K proteinases of TEV, TVMV and PPV would result in an N-terminal protein of 22K to 24K, which is similar in size to the size determined for their VPgs.

Introduction

Pea seed-borne mosaic virus (PSbMV) is a member of the potyvirus group. The flexuous, rod-shaped particles of the potyviruses contain a positive-sense ssRNA genome of approximately 10 kb (Hollings & Brunt, 1981). The 5' terminus of the RNA is covalently linked to a virus-encoded protein (VPg) (Murphy *et al.*, 1990) and the 3' end terminates with a poly(A) tract (Hari *et al.*, 1979).

The complete nucleotide sequences of the potyviruses tobacco etch virus (TEV; Allison *et al.*, 1986), tobacco vein mottling virus (TVMV; Domier *et al.*, 1986), plum pox virus (PPV; Maiss *et al.*, 1989) and potato virus Y (PVY; Robaglia *et al.*, 1989) have been determined. Analysis of these sequences has demonstrated that the potyvirus genome contains one large open reading frame (ORF) with the potential to encode a polyprotein of between 3005 amino acids (TVMV) and 3063 amino acids (PVY). The polyprotein is proteolytically processed into at least eight proteins (Carrington & Dougherty, 1987*a*, *b*; Carrington *et al.*, 1989, 1990): the N-terminal protein, the helper component-proteinase (HC-Pro), a 42K to 50K protein, the cytoplasmic inclusion protein (CI), a 6K protein, the small nuclear inclusion protein (NI_a), the large nuclear inclusion protein (NI_b) and the coat protein (CP) (Dougherty & Carrington, 1988). The functions of some of the non-structural proteins have been established. HC-Pro is necessary for aphid transmission (Thornbury *et al.*, 1985) and also has proteolytic activity (Carrington *et al.*, 1989); CI has helicase activity (Laín *et al.*, 1990); NI_a, or the N-terminal part of it, is VPg, which is covalently linked to the RNA (Murphy *et al.*, 1990; Riechmann *et al.*, 1989). NI_a also has proteolytic activity (Carrington & Dougherty, 1987*a*) and will be referred to as the 49K proteinase (49K-Pro). NI_b is expected to have RNA polymerase activity owing to its homology with other viral RNA-dependent RNA polymerases (Domier *et al.*, 1987).

We have determined the complete nucleotide sequence of the PSbMV genome. Analysis of the deduced amino acid sequences of PSbMV, TEV, TVMV, PPV, PVY and soybean mosaic virus (SMV) suggests the presence of an additional proteolytic cleavage site in 49K-Pro.

Methods

Virus and RNA purification. The PSbMV isolate (DPD1) originated from a pea seed sample analysed by J. Jørgensen at the Danish Plant Directorate (Lyngby, Denmark). The virus was propagated in *Pisum sativum* L. cv. Dark Skin Perfection. Virus particles were purified 3 weeks after inoculation by the method of Alconero *et al.* (1986), and viral RNA was extracted as described by Maiss *et al.* (1988).

DNA manipulations. DNA was digested with restriction enzymes from Boehringer Mannhein, according to the manufacturer's recommendations. DNA fragments separated on agarose gels were purified using a Prep-A-Gene kit (Bio-Rad). The plasmid vectors pGEM-3Z and pGEM-7Z(+) (Promega) were used for cloning cDNA and for subsequent subcloning for sequence analysis. DNA was ligated under the conditions described by Pfeiffer & Zimmermann (1983), and transformation of *Escherichia coli* was done as recommended by Hanahan (1985).

cDNA cloning. Purified RNA was used as the template for oligo(dT) as well as random primed cDNA synthesis using the cDNA Synthesis System Plus (Amersham). The cDNA was either cut with selected enzymes or methylated with *Eco*RI methylase (New England Biolabs), and ligated to *Eco*RI linkers. Subsequently, the cDNA was ligated to the plasmid vector and cloned into *E. coli* DH5α (Bethesda Research Laboratories).

Nucleotide sequencing and sequence analysis. Sequenase version 2.0 (United States Biochemical) was used in sequence analysis studies. Each nucleotide was determined by sequencing either two independent cDNA clones or by sequencing both strands of a cDNA clone. Computer analyses were performed using the Microgenie sequence software (Beckman).

RNA sequencing. The sequence of the 5′-terminal region was determined by direct sequencing of the RNA using a 18-mer synthetic oligonucleotide [5′ GAGCTTGTTACTTAAGTG 3′ (oligo A)] complementary to positions 120 to 137 of the viral RNA. The RNA sequence reactions were performed with reverse transcriptase in the presence of dideoxynucleotides (Geliebter, 1987). The identity of the first nucleotide at the 5′ terminus was determined by extending the reverse transcriptase reaction products with terminal deoxynucleotidyl transferase (DeBorde *et al.*, 1986).

Cloning of the 5′-terminal region. The 5′-terminal region was cloned as described by Domier *et al.* (1989) using oligo A to prime first strand synthesis. Second strand synthesis was initiated by a second oligonucleotide [5′ TTTAAAATAAAACAACTCAA 3′ (oligo B)] identical to nucleotides 1 to 17 of PSbMV except for the three thymidines at the 5′ end. These were added to generate a *Dra*I site which allows precise excision of the 5′ terminus of PSbMV cDNA.

Results and Discussion

Sequencing of the PSbMV genome

Initial screening of the cDNA clones by restriction enzyme mapping, hybridization analysis and partial sequencing identified four overlapping clones, pPS29 (1·6 kb), pPS128 (1·3 kb), pPS21 (3·0 kb) and pPS13 (5·8 kb), which covered the entire coding region, 94 bases of the 5′ untranslated leader (5′ UTL), the 3′ untranslated region (3′ UTR) of 163 nucleotides and approximately 60

nucleotides of the poly(A) tract. The sequence of the 5′-terminal region not contained in any cDNA clone initially analysed was determined by direct sequencing of RNA. Subsequently, the 5′-terminal region was cloned (pPS34) and sequenced. Thirteen cDNA clones and subclones thereof were used for sequence determination (Fig. 5). The nucleotide sequence of PSbMV is shown in Fig. 1.

Coding and noncoding regions

The genome is 9924 nucleotides in length excluding the 3′-terminal poly(A) tract. cDNA clone pPS13 (Fig. 5) has a poly(A) tract of 60 residues, which was the longest determined. Computer analysis revealed an ORF of 9618 nucleotides, starting at position 144, with the potential to encode a polyprotein of 3206 amino acids with a calculated M_r of 364K. The first AUG codon, at positions 144 to 146, is likely to be the initiation codon for the polyprotein of PSbMV because, firstly, this AUG is in the context of a ribosomal binding site and has an A residue in position −3 which is not likely to be passed by the scanning 40S ribosomal subunit (Kozak, 1986). Secondly, the first four amino acids, MSTI, of the PSbMV polyprotein are also the first amino acids of the polyproteins of TVMV and PPV, and are present at amino acid positions 5 to 8 in the polyprotein of PVY (Fig. 6). Thirdly, the overall base composition of PSbMV is 33% A, 18% C, 24% G and 26% T, whereas that of the 143 nucleotide leader sequence is 41% A, 20% C, 7% G and 33% T. This very low G residue content seems to be a common feature of plant virus 5′ leader sequences (Gallie *et al.*, 1987). If the next in-frame AUG codon downstream from the proposed initiation codon is taken as the polyprotein initiation codon, the G residue content of the 5′ leader increases to 18%.

The size of the genome and polyprotein of TEV (Allison *et al.*, 1986), TVMV (Domier *et al.*, 1986), PPV (Maiss *et al.*, 1989), PVY (Robaglia *et al.*, 1989) and PSbMV are listed in Table 1. The PSbMV genome is the largest of those sequenced and, because the 5′ UTL and

Table 1. *Sizes of the genome, 5′ UTL, 3′ UTR and polyprotein of potyviruses*

Virus	Genome*	5′ UTL*	3′ UTR*	Polyprotein†
TEV	9496	144	190	3054
TVMV	9472	206	251	3005
PPV	9741	146	220	3125
PVY	9704	184	331	3063
PSbMV	9924	143	163	3206

* Length in nucleotides.
† Length in amino acids.

the 3′ UTR are shorter than the corresponding regions in other potyviruses, the extension is within the polyprotein of PSbMV, which is between 81 and 201 amino acids longer than the polyproteins of the four other potyviruses.

The initial 30 nucleotides of the PSbMV 5′ UTL were found to have a high degree of sequence identity with other potyviruses (Fig. 2). At present, the function of this part of the 5′ UTL has not been established, but it may play an important role in replication of the virus genome. No significant similarities could be detected in the primary or secondary structure of the 3′ UTR.

Proteolytic processing of the polyprotein

A potential cleavage site separating the N-terminal protein from HC-Pro could not be identified because this cleavage site has not been determined for any of the potyviruses. Evidence has been presented that a cryptic virus-encoded proteinase which is dependent on a host factor or a plant proteinase is responsible for this cleavage (Carrington *et al.*, 1990).

Cell-free expression of defined RNA transcripts *in vitro* has identified a potential cleavage site for the TEV-encoded HC-Pro, between the dipeptide G-G located at amino acid positions 763 and 764 in the TEV polyprotein (Carrington *et al.*, 1989). Alignment of the protein region surrounding this cleavage site in TEV with the corresponding region of PSbMV, TVMV, PPV and PVY showed that the dipeptide G-G is conserved in all five potyviruses (Fig. 3); it is located at amino acid positions 856 and 857 in the PSbMV polyprotein.

A recent comparison of both determined and putative cleavage sites in the potyvirus polyprotein has shown that 49K-Pro cleaves at Q-A, Q-G, Q-S or Q-T dipeptide sequences (Ghabrial *et al.*, 1990). Valine was found in position P4 in approximately 80% of cleavage sites and glutamic acid in half of the P6 positions. Based on these data and on sequence homologies of the polyproteins, five putative 49K-Pro cleavage sites were identified in the PSbMV polyprotein at the expected positions (Table 2). The sequence V(R/K)X(Q/E)↓(S/G/A) is proposed to be a consensus cleavage sequence for the 49K-Pro of PSbMV.

The putative cleavage site between the 6K protein and 49K-Pro differed from the usual potyvirus consensus pattern as it contained glutamic acid instead of glutamine in position P1. However, glutamic acid in position P1 is predicted for several picornavirus cleavage sites (Palmenberg, 1987); these cleavage sites are processed by the 3C proteinase of the picornaviruses, which is homologous to the 49K-Pro of the potyviruses (Domier *et al.*, 1987). To ensure that the deduced glutamic acid was

Table 2. *49K-Pro cleavage sites in the PSbMV polyprotein**

Site	Amino acid position						
	P6	P5	P4	P3	P2	P1	P′1
42K↓CI	Q	E	V	R	C	Q	S
CI↓6K	E	T	V	R	C	Q	S
6K↓49K-Pro	E	P	V	K	L	E	G
49K-Pro↓NI_b	D	G	V	R	T	Q	S
NI_b↓CP	I	K	V	R	L	Q	A
Consensus cleavage site	–	–	V	R/K	–	Q/E	S/G/A

* Sites were identified by comparison with the determined and/or proposed sites in TEV, TVMV, PPV and PVY polyproteins.

not the result of an error during cDNA synthesis, the sequence of this region was confirmed by sequencing cDNA clones obtained from two independent cDNA synthesis reactions.

Presence of a potential internal cleavage site in 49K-Pro

Searching the PSbMV polyprotein for the presence of the proposed consensus cleavage site sequence of 49K-Pro, V(R/K)X(Q/E)↓(S/G/A), revealed the presence of an additional potential cleavage site, EHVRTE↓A, in 49K-Pro at amino acid positions 2144 to 2150. Alignment of the amino acid sequence of PSbMV 49K-Pro with those of TEV, TVMV, PPV and PVY and SMV revealed similar potential cleavage sites in their 49K-Pro proteins (Fig. 4). The most prominent feature of these sites when compared with the respective consensus sequences was the presence of glutamic acid instead of glutamine in position P1. However, they all had several features in common with the other determined or proposed 49K-Pro cleavage sites. (i) Serine is the most common amino acid at position P′1; (ii) histidine and phenylalanine are common amino acids at position P2; (iii) a high preference for valine in position P4; (iv) a preference for glutamic acid or glutamine in position P6; (v) the total absence of glycine from the cleavage sequences (see Ghabrial *et al.*, 1990). If these sites are recognized *in vivo* the 49K-Pro proteins of PSbMV, TEV, TVMV, PPV and PVY would be cleaved into two proteins of approximately 22K and 28K. The VPg of TEV has been shown to be 49K-Pro or the N-terminal 24K part of this protein (Murphy *et al.*, 1990). A size of 24K is close to that of 21·2K calculated for the N-terminal part of the potential cleavage product of TEV 49K-Pro. The genome-linked proteins of PPV and TVMV have been determined to have M_rs of 22K and 24K, respectively (Riechmann *et al.*, 1989; Siaw *et al.*, 1985), which are in good agreement with those calculated for the potential

Fig. 1. Nucleotide sequence of PSbMV RNA and the deduced amino acid sequence of the putative polyprotein. The amino acid sequence which starts at nucleotide 144 is shown using the single-letter code. The putative proteolytic cleavage sites for 49K-Pro are underlined.

```
TEV    ¹NAAAUAACAAAUCUCAACACAACAUAUACAAAACA³⁵  -106nt-  ¹⁵GCAAUGG
TVMV   ¹AAAAUAAAACAAAUCAACACAACAUUAUAACGAAC³⁵  -168nt-  ²⁰⁴GCCAUGU.
PPV    ¹AAAAUAUAAAAACUCAACACAACAUACAAAAUUUU³⁵  -108nt-  ¹⁰⁸AAGAUGU
PVY    ¹ AAUUAAAACAACUCAAUCAACAUAAGAAAAACA³⁴  -147nt-  ¹⁴⁸UCAAUGG
PSbMV  ¹AAAAUAAAACAACUCAACAAAACAUAAUCGAAAAU³⁵  -105nt-  ¹⁴¹AAAAUGU
```

Fig. 2. Partial nucleotide sequence of the 5′ UTLs of TEV, TVMV, PPV, PVY and PSbMV showing the initial 34 or 35 nucleotides and the sequence surrounding the translation initiation codon. The length of the sequences between these two regions was between 105 nucleotides, in PSbMV, and 168 nucleotides, in TVMV. Conserved nucleotides are shown in bold.

```
                            ▼
TEV    ²⁴⁴QLIEFANDELESEMKTYNVGGMNRDVVT        QGAIEMLIKSIYKP²⁸⁵
TVMV   ⁴⁰⁰QLEKFASNTLESPMAQYKVGGLVYSENN -1aa-  ASAVKALTQAIFRP⁷⁴⁶
PPV    ⁷⁰⁹QLISFASDTLDSNMKTYLVGGLEVDKCD -1aa-  FKNVKLLIRSIYKP⁷⁵⁰
PVY    ⁷⁵³QLILFANDELESDIKHYRVGGVPNASPE -17aa- SAALKLLLKGIFRP⁷⁷⁹
PSbMV  ⁴¹⁷QLISFAHGALLGEMKMYRVGGTQKMEIN -7aa-  NLLIKQLIRAIYRP⁴⁶¹
```

Fig. 3. Partial sequence alignment of TEV, TVMV, PPV, PVY and PSbMV. The sequences shown are the regions surrounding the dipeptide G-G, shown to be cleaved by the TEV-encoded HC-Pro (amino acid positions 763 and 764, marked by an arrowhead) in vitro. Conserved amino acids are shown in bold.

(a)
```
                                              ▼
TEV    ²⁰⁵⁵ERENELRQTGMAVPVAYDQLPP-KN----- -EDLTFEGESLFKGPRD²⁰⁸⁷
TVMV   ¹⁹⁵⁴EREGELRRTGPTETLPFDALPPEKQ------E-VAFESKALLKGVRD¹⁹⁹¹
PPV    ²⁰¹⁸EREFELRQTGHPVFVEPNAIPK-INE--EGDEE-VDHESKSLFRGLRD²⁰⁵⁹
PVY    ²⁰⁰⁵ERELELRQTGPAVEVDVKDIPA-Q--------E-VEHEAKSLMRGLRD²⁰⁴¹
SMV    EREDELRQTGLPQVVSKSDVPRAK--------ERVEMESKSVYKGLRD
PSbMV  ²¹¹⁵EREFELRQSGSATLVPYSEVPV-QNEKQEFDEBHVRTEAASLHFGLRD²¹⁵⁹
```

(b)

	Internal 49K-Pro cleavage site sequence						
Potyvirus	P6	P5	P4	P3	P2	P1	P'1
TEV	E	D	L	T	F	E	G
TVMV	Q	E	V	A	F	E	S
PPV	E	E	V	D	H	E	S
PVY	Q	E	V	E	H	E	A
SMV	E	R	V	E	M	E	S
PSbMV	E	H	V	R	T	E	A

(c)

		Cleavage site sequence						
Potyvirus		P6	P5	P4	P3	P2	P1	P'1
TEV	Consensus	E	–	–	Y	–	Q	G/S
	49K-Pro	E	–	–	T	–	E	G
TVMV	Consensus	–	–	V	R	F	Q	S/K/G
	49K-Pro	–	–	V	A	F	E	S
PPV	Consensus	–	–	V	–	H/T	Q	S/T/A
	49K-Pro	–	–	V	–	H	E	S
PVY	Consensus	–	–	V	–	H/E	Q	A/S/G
	49K-Pro	–	–	V	–	H	E	A
SMV	Consensus	E/N	–	V	–	–	Q	S/G
	49K-Pro	E	–	V	–	–	E	S
PSbMV	Consensus	–	–	V	R/K	–	Q/E	A/S/G
	49K-Pro	–	–	V	R	–	E	A

Fig. 4. (a) Amino acid sequence alignment of TEV, TVMV, PVY, SMV and PSbMV. The aligned regions contain the putative internal cleavage site located in 49K-Pro of PSbMV. Conserved amino acids are printed in bold. The putative internal cleavage site is marked by an arrowhead. (b) Comparison of the internal 49K-Pro cleavage site sequences for the six viruses. (c) Comparison of the consensus sequence with the amino acid sequence of the putative internal cleavage site in 49K-Pro.
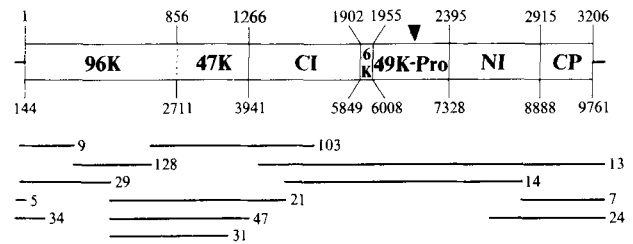

Fig. 5. A tentative map of the PSbMV polyprotein. Vertical lines indicate the proposed cleavage sites for 49K-Pro. The hypothetical cleavage site located within 49K-Pro between amino acids 2149 and 2150 is indicated by an arrowhead. The dotted line indicates the possible cleavage site at the C terminus of HC-Pro. Numbers above the polyprotein are amino acid positions; numbers below the polyprotein are nucleotide positions. Also shown is a map of the PSbMV cDNA clones used for subcloning and sequencing. Clones pPS5, pPS7, pPS9, pPS13, pPS14, pPS21, pPS24, pPS29, pPS31, pPS34, pPS47, pPS103 and pPS128 are marked with their numbers only.

N-terminal cleavage products of the 49K-Pros, giving further indirect evidence that this cleavage site is processed in vivo. It should be noted that processing at the putative internal cleavage site of 49K-Pro has never been detected in vitro or in vivo. However, this could indicate that only a minor fraction of the total 49K-Pro protein is cleaved or that cleavage at this particular site requires the presence of some factor(s) modifying the specificity of 49K-Pro. Such a factor(s) could provide coupling between internal cleavage of 49K-Pro and initiation of replication.

A tentative map of the PSbMV polyprotein is presented in Fig. 5. The proposed locations of the cleavage sites predicted seven or eight proteins with calculated M_rs of 96K, 47K, 70K, 6K, 51K (or 23K and 28K), 59K and 33K. The 96K protein is expected to be cleaved in vivo into the N-terminal protein and HC-Pro.

Comparison of PSbMV proteins with other potyvirus proteins

Since no potential cleavage site could be identified between the N-terminal protein and HC-Pro, the large 96K protein is expected to contain both of these. Comparison of the N-terminal regions of TEV, TVMV, PPV, PVY and PSbMV showed that this region was variable both in primary amino acid sequence and in length. However, some regions of sequence similarity could be identified (Fig. 6). From this comparison it was evident that the variable region of the N-terminal region of PSbMV was between 93 and 142 amino acids longer than the corresponding regions of the other four potyviruses. The N-terminal protein has been proposed to have a function in cell-to-cell movement of potyviruses (Domier et al., 1987). This suggestion was made on the basis of sequence similarities between the N-terminal

```
TEV    'MALIFGTVNA¹⁰ -199aa- ²¹⁰ASVRHMYGERKRVDLR²²³ -49aa- ²⁷³DLFIVRGK²⁸¹
TVMV   'MSTIHSAVTA¹⁰ -150aa- ¹⁶¹IDVAHAKGHRRRIDCR¹⁷⁶ -48aa- ²²⁵EFFIVRGT²³³
PPV    'MSTIVFGSFT¹⁰ -201aa- ²¹²THVRHLDGSKPRYDLV²²³ -52aa- ²⁷⁶DLFIVRGK²⁸³
PVY    'MATYMSTICF¹⁰ -177aa- ¹⁸⁸VRTAHMMGLRRRVDSD³⁰³ -50aa- ²⁵⁴GLFIVRGS²⁶¹
PSbMV  'MSTLVCQAVA¹⁰ -292aa- ³⁰³VETKHHKGIISQKDAS³¹⁸ -48aa- ³⁶⁷DLFVVRGK³⁷⁴
```

Fig. 6. Partial protein sequence alignment of the N termini of TEV, TVMV, PPV, PVY and PSbMV polyproteins. The sequences shown are the start of the polyprotein, then a region in which some sequence identity became apparent and finally a region of high sequence identity starting 48 to 52 amino acids after the first identity was observed. Conserved amino acids are printed in bold.

region of TVMV and the 30K movement protein of tobacco mosaic virus (TMV). However, the amino acid identities found in the TVMV N-terminal protein and the TMV 30K protein were not observed in TEV, PPV, PVY or PSbMV.

Robaglia *et al.* (1989) have identified a cysteine cluster in the putative N-terminal region of HC-Pro of PVY. This cluster of five cysteines is perfectly conserved in the five potyviruses sequenced, including PSbMV (amino acids 425 to 456 in the PSbMV polyprotein). The proposed 47K protein of PSbMV showed only limited sequence similarity with the corresponding proteins of PVY (26%), TEV and TVMV (28%).

The putative CI protein of PSbMV showed between 52% and 57% sequence identity with the CI proteins of TEV, TVMV, PPV and PVY. A possible nucleotide binding motif, (G/A)XXXXGK(S/T), which has been identified in CI proteins of the potyviruses TEV and TVMV (Gorbalenya *et al.*, 1989), was also present in the proposed CI protein of PSbMV (amino acid positions 1351 to 1359). This motif is also present in the CI protein of PPV and this protein has been demonstrated to have RNA unwinding (helicase) activity (Laín *et al.*, 1990). These data all suggest that the CI protein is involved in replication and could be the membrane-bound protein of a larger replication complex, as suggested by Goldbach (1987).

The proposed cleavage site, EPVKLE↓G, at the C terminus of the PSbMV-encoded 6K protein is unusual in containing glutamic acid instead of glutamine. Cleavage at this particular site would result in a protein which is identical in size to the 6K proteins of TEV and TVMV, supporting the location of this cleavage site. The PSbMV 6K protein showed 42% and 43% amino acid sequence identity with TEV and TVMV, respectively.

The PSbMV protein corresponding to the 49K-Pro of TEV has a calculated M_r of 51K. It showed from 44% to 50% amino acid sequence identity with the corresponding proteins of TVMV and PPV, respectively.

The proposed NI$_b$ protein of PSbMV showed a high degree of sequence similarity with the corresponding proteins of the four other potyviruses, ranging from 58% with PVY to 63% with TVMV and PPV. The GDD sequence motif found in many replicase enzymes

(Strauss & Strauss, 1988) was also present in PSbMV, starting at amino acid position 2744.

PSbMV CP showed 49% and 60% similarity with the capsid proteins of TVMV and PVY, respectively. Recently, the CP of another isolate of PSbMV has been sequenced (Timmerman *et al.*, 1990); these two CPs showed 98% amino acid identity. The major difference between the two isolates was a deletion in the isolate sequenced by Timmerman *et al.* (1990) encompassing amino acids IPRY located at positions 3120 to 3123 in the polyprotein of our isolate of PSbMV (Johansen *et al.*, 1991). In all the other potyvirus CPs which have been sequenced, these four amino acids are highly conserved.

The high degree of conservation of amino acid sequence in many potyvirus proteins could serve as a guideline for attempts to define and understand their functions. The use of infectious transcripts of full-length cDNA clones may be of use in these studies because they allow the effect of defined mutations to be determined.

References

ALCONERO, R., PROVVIDENTI, R. & GONZALES, D. (1986). Three pea seedborne mosaic virus pathotypes from pea and lentil germ plasm. *Plant Disease* **70**, 783–786.

ALLISON, R., JOHNSTON, R. E. & DOUGHERTY, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* **154**, 9–20.

CARRINGTON, J. C. & DOUGHERTY, W. G. (1987a). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *Journal of Virology* **61**, 2540–2548.

CARRINGTON, J. C. & DOUGHERTY, W. G. (1987b). Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology* **160**, 355–362.

CARRINGTON, J. C., CARY, S. M., PARKS, T. D. & DOUGHERTY, W. G. (1989). A second proteinase encoded by a plant potyvirus genome. *EMBO Journal* **8**, 365–370.

CARRINGTON, J. C., FREED, D. D. & OH, C.-S. (1990). Expression of potyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. *EMBO Journal* **9**, 1347–1353.

DEBORDE, D. C., NAEVE, C. W., HERLOCHER, M. L. & MAASSAB, H. F. (1986). Resolution of a common RNA sequencing ambiguity by terminal deoxynucleotidyltransferase. *Analytical Biochemistry* **157**, 275–282.

DOMIER, L. L., FRANKLIN, K. M., SHAHABUDDIN, M., HELLMANN, G. M., OVERMEYER, J. H., HIREMATH, S. T., SIAW, M. F. E., LOMONOSSOFF, G. P., SHAW, J. G. & RHOADS, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Research* **14**, 5417–5430.

DOMIER, L. L., SHAW, J. G. & RHOADS, R. E. (1987). Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. *Virology* **158**, 20–27.

DOMIER, L. L., FRANKLIN, K. M., HUNT, A. G., RHOADS, R. E. & SHAW, J. G. (1989). Infectious *in vitro* transcripts from cloned cDNA of a potyvirus, tobacco vein mottling virus. *Proceedings of the National Academy of Sciences, U.S.A.* **86**, 3509–3513.

DOUGHERTY, W. G. & CARRINGTON, J. C. (1988). Expression and function of potyviral gene products. *Annual Review of Phytopathology* **26**, 123–143.

GALLIE, D. R., SLEAT, D. E., WATTS, J. W., TURNER, P. C. & WILSON, T. M. A. (1987). A comparison of eukaryotic viral 5′-leader sequences as enhancers of mRNA expression *in vivo. Nucleic Acids Research* **15**, 8693–8711.

GELIEBTER, J. (1987). Dideoxynucleotide sequencing of RNA and uncloned cDNA. *Focus* **9**, 5–8. Gaithersburg: Bethesda Research Laboratories.

GHABRIAL, S. A., SMITH, H. A., PARKS, T. D. & DOUGHERTY, W. G. (1990). Molecular genetic analyses of the soybean mosaic virus NIa proteinase. *Journal of General Virology* **71**, 1921–1927.

GORBALENYA, A. E., KOONIN, E. V., DONCHENKO, A. P. & BLINOV, V. M. (1989). A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Letters* **235**, 16–24.

GOLDBACH, R. (1987). Genome similarities between plant and animal RNA viruses. *Microbiological Sciences* **4**, 197–202.

HANAHAN, D. (1985). Techniques for transformation of *E. coli.* In *DNA Cloning: A Practical Approach*, vol. 1, pp. 109–135. Edited by D. M. Glover. Oxford: IRL Press.

HARI, V., SIEGEL, A., ROZEK, C. & TIMERLAKE, W. E. (1979). The RNA of tobacco etch virus contains poly(A). *Virology* **92**, 568–571.

HOLLINGS, M. & BRUNT, A. A. (1981). Potyviruses. In *Handbook of Plant Virus Infections and Comparative Diagnosis*, pp. 731–807. Edited by E. Kurstak. Amsterdam: North-Holland.

JOHANSEN, E., RASMUSSEN, O. F., HEIDE, M. & BORKHARDT, B. (1991). Nucleotide sequence of pea seedborne mosaic potyvirus coat protein gene. *Virus Genes* (in press).

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

LAÍN, S., REICHMANN, J. L. & GARCÍA, J. A. (1990). RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus. *Nucleic Acids Research* **18**, 7003–7006.

MAISS, E., BREYEL, E., BRISSKE, A. & CASPER, A. (1988). Molecular cloning of DNA complementary to the RNA-genome of plum pox virus (PPV). *Journal of Phytopathology* **122**, 222–231.

MAISS, E., TIMPE, U., BRISSKE, A., JELKMANN, W., CASPER, R., HIMMLER, G., MATTANOVICH, D. & KATINGER, H. W. D. (1989). The complete nucleotide sequence of plum pox virus RNA. *Journal of General Virology* **70**, 513–524.

MURPHY, J. F., RHOADS, R. E., HUNT, A. G. & SHAW, J. G. (1990). The VPg of tobacco etch virus RNA is the 49-kDa proteinase or the N-terminal 24-kDa part of the proteinase. *Virology* **178**, 285–288.

PALMENBERG, A. C. (1987). Picornaviral processing: some new ideas. *Journal of Cellular Biochemistry* **33**, 191–198.

PFEIFFER, B. H. & ZIMMERMAN, S. B. (1983). Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribooligonucleotides by T4 DNA ligase in polymer solutions. *Nucleic Acids Research* **11**, 7853–7871.

RIECHMANN, J. L., LAÍN, S. & GARCÍA, J. A. (1989). The genome-linked protein and 5′ end RNA sequence of plum pox potyvirus. *Journal of General Virology* **70**, 2785–2789.

ROBAGLIA, C., DURAND-TARDIF, M., TRONCHET, M., BOUDAZIN, G., ASTIER-MANIFACIER, S. & CASSE-DELBART, F. (1989). Nucleotide sequence of potato virus Y (N strain) genomic RNA. *Journal of General Virology* **70**, 935–947.

SIAW, M. F. E., SHAHABUDDIN, M., BALLARD, S., SHAW, J. G. & RHOADS, R. E. (1985). Identification of a protein covalently linked to the 5′ terminus of tobacco vein mottling virus RNA. *Virology* **142**, 134–143.

STRAUSS, J. H. & STRAUSS, E. G. (1988). Evolution of RNA viruses. *Annual Review of Microbiology* **42**, 657–683.

THORNBURY, D. W., HELLMANN, G. M., RHOADS, R. E. & PIRONE, T. P. (1985). Purification and characterization of potyvirus helper component. *Virology* **144**, 260–267.

TIMMERMAN, G. M., CALDER, V. L. & BOLGER, E. A. (1990). Nucleotide sequence of the coat protein gene of pea seed-borne mosaic potyvirus. *Journal of General Virology* **71**, 1869–1872.