Journal of General Virology (1990), 71, 1869–1872. *Printed in Great Britain*

Nucleotide sequence of the coat protein gene of pea seed-borne mosaic potyvirus

Gail M. Timmerman,[1]* Victoria L. Calder[2] and Lorraine E. A. Bolger[2]

[1]*Crop Research Division, Department of Scientific and Industrial Research, Private Bag, Christchurch and* [2]*Department of Plant and Microbial Sciences, University of Canterbury, Christchurch, New Zealand*

The nucleotide sequence of a 1355 bp cDNA representing the 3′-terminal sequences of pea seed-borne mosaic virus (PSbMV) was determined. This sequence contained a single long open reading frame (ORF) of 1189 bp ending with a single TAA termination codon. Downstream from the ORF was an untranslatable region of 189 bp followed by eight bp of polyadenylate. The probable location of the PSbMV coat protein codons within the long ORF was determined by comparing the inferred amino acid sequence with other potyviral coat protein sequences and by examining the sequence for a potyviral polyprotein cleavage cassette sequence. Direct chemical sequencing of the PSbMV coat protein revealed it to be blocked at its amino terminus. A partial amino acid sequence representing the N terminus of the protease-resistant core of the coat protein was determined, however. Alignment of the PSbMV coat protein sequence and the sequences of seven other potyviral coat proteins revealed significant homology, ranging from 53·7% for potato virus Y strain D to 43·2% for tobacco vein mottling virus.

The potyviruses form the largest group of plant viruses, and many cause economically significant disease. Potyviruses have filamentous virions which carry a single-stranded, positive-sense RNA genome of approximately 10000 nucleotides. Their RNA genome is covalently modified at its 5′ end by a virus-encoded protein (VPg) and has a 3′ polyadenylate tail. The complete nucleotide sequences of three potyviruses, tobacco etch virus (TEV) (Allison *et al.*, 1986), tobacco vein mottling virus (TVMV) (Domier *et al.*, 1986) and plum pox virus (PPV) (Maiss *et al.*, 1989) have been determined. Amino acid and/or nucleotide sequences have also been determined for a number of other potyvirus coat proteins or their genes.

Pea seed-borne mosaic virus (PSbMV) is an important member of the potyviral group of plant pathogens. It infects a variety of plant species, with the greatest economic impact occurring when peas (*Pisum sativum*) are infected (Hampton & Mink, 1975). The virus has a broad geographical distribution and has probably been spread throughout the world by infected seed. Aspects of the biology and epidemiology of PSbMV have been reviewed recently (Khetarpal & Maury, 1987).

Recently, transgenic plants expressing plant viral coat protein genes have been shown to be less susceptible to viral disease, a phenomenon termed genetically engineered cross-protection (Nelson *et al.*, 1987). Determination of the sequence of the PSbMV coat protein gene is the first step toward producing transgenic plants expressing this gene, and ultimately towards using this technology to protect grain legume crops from PSbMV-caused disease. This paper reports the nucleotide sequence of the 3′-terminal 1355 nucleotides of PSbMV, which includes the entire coat protein coding region.

PSbMV pathotype P-1 (Alconero *et al.*, 1986), was purified from infected *P. sativum* plants using a modification of the procedure published by Reddick & Barnett (1983). Full-length viral RNA was isolated from freshly prepared virus as described by Brakke & van Pelt (1970). cDNA was synthesized by the single tube reaction described by D'Alessio *et al.* (1987) using an oligo(dT)$_{12-18}$ primer, then cloned into *Sma*I -digested, dephosphorylated pUC19 plasmid using standard methods (Maniatis *et al.*, 1982). The resulting library contained 360 clones and was screened for the length of the inserted cDNA by digesting mini-preparations of plasmid DNA (Birnboim & Doly, 1979) with restriction endonucleases *Eco*RI and *Bam*HI. Three clones, pPSB70, pPSB67 and pPSB13, containing inserts of 1355 and approximately 1270 and 1000 bp respectively were chosen for preliminary nucleotide sequence analysis. Restriction fragments from these clones were subcloned into the polylinker region of M13mp18 or -mp19, and DNA sequences were determined by the dideoxynucleotide chain termination method of Sanger *et al.* (1977).

All three of the clones examined contained 3′ poly(A)

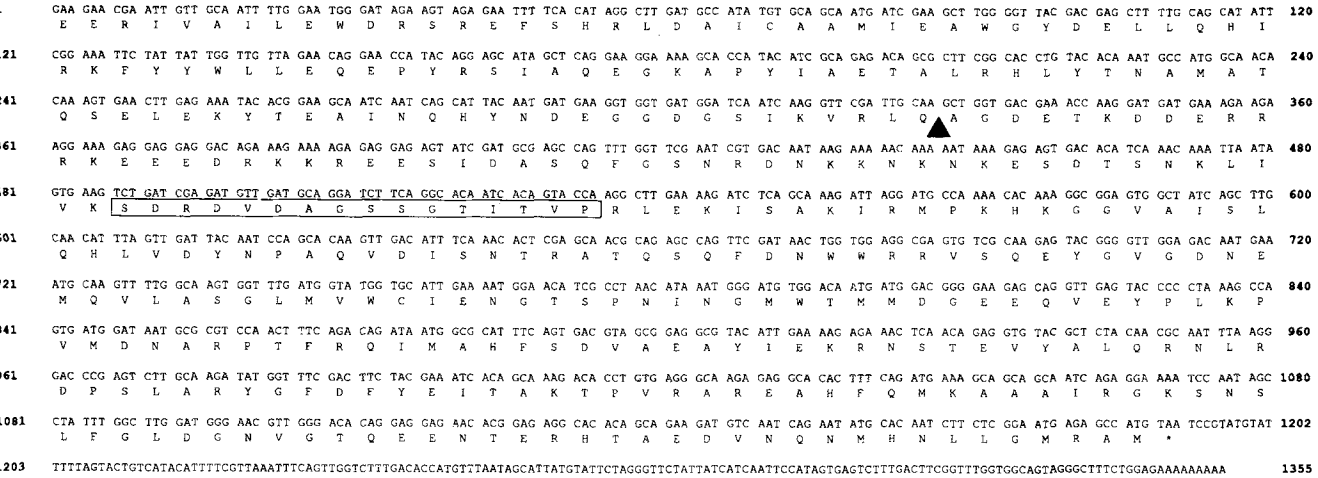

Fig. 1. Nucleotide sequence of the 3′ terminus of PSbMV RNA. The amino acid sequence predicted for the ORF is presented below the nucleotide sequence. An arrowhead indicates the probable proteolytic cleavage site for release of coat protein from the polyprotein translation product. The boxed region indicates the amino acid sequence determined by automated Edman degradation. The single termination codon at the end of the ORF is indicated by an asterisk.

tracts as well as identical sequences adjacent to the polyadenylate tail. The complete nucleotide sequence of the 1355 bp cDNA inserted in plasmid pPSB70 was determined. The resulting sequence was assembled and analysed on an IBM PC-compatible computer using the GENESYS software written by W. Bottomley (CSIRO Division of Plant Industry, Canberra, Australia).

The sequence of the 1355 nucleotides at the 3′ end of PSbMV genomic RNA is presented in Fig. 1. This sequence contains a single long open reading frame (ORF) found on the positive strand and ending with a single termination codon (TAA) at nucleotide 1189. It has a 3′ untranslatable region of 159 nucleotides and ends with eight adenylate residues. The sequence is purine-rich, containing 33·5% adenosine and 26·3% guanosine, as well as 22·9% thymidine and 17·3% cytidine.

The single ORF of 1189 nucleotides is long enough to encode the PSbMV coat protein as well as some of the preceding cistron. No other extended ORF is found by computer analysis of either the positive strand or the negative strand. As for other potyviruses, the primary translation product of the PSbMV genome is probably a polyprotein which is proteolytically cleaved to produce the mature viral proteins (Calder, 1989). As is the case with TEV, TVMV and PPV, the codon that initiates translation is expected to reside near the 5′ end of PSbMV genomic RNA, and therefore is not present in the sequences reported in this paper.

The primary translation products of other potyviruses are proteolytically cleaved at either Q-A, Q-S or Q-G dipeptides to release mature coat protein. The probable amino terminus of the PSbMV coat protein has been predicted by examining the amino acid sequence of the long ORF. The most likely precursor cleavage site for the coat protein is between the Q-A dipeptide found at nucleotides 325 to 330 (Fig. 1). The amino acid sequence at this site (V-R-L-Q-A) closely resembles the consensus cleavage sites described for TVMV polyprotein processing (Domier *et al.*, 1986) as well as the cleavage sites used by other potyviruses, except for TEV (Dougherty *et al.*, 1989). Cleavage at this site produces a coat protein of 287 amino acids, with a calculated M_r of 32651. This agrees well with the results of SDS–polyacrylamide gel electrophoresis of purified PSbMV, which produces two protein bands having mobilities corresponding to 36K and 33K (Calder, 1989). It is common to observe heterogeneity in the size of potyviral coat proteins, and this is often due to proteolytic removal of the hydrophilic N-terminal domain.

Direct chemical amino acid sequencing of the PSbMV coat protein was carried out on 200 pmol of purified whole virus. The sequence of the N terminus could not be determined directly, probably due to an acetylated amino acid. An amino acid sequence representing 6·5% of the total sample was determined, however. This sequence is *-D-*-D-V-D-A-G-S-*-G-*-I-*-V-P, where * represents residues which could not be identified positively. The location of this sequence is identified by the boxed region in Fig. 1, and it coincides with the N terminus of the trypsin-resistant core protein described for other potyviral coat proteins (Shukla *et al.*, 1988*a*). Other potyviral coat proteins have blocked N termini, including Johnson grass mosaic virus (JGMV) (Gough *et al.*, 1987), three strains of sugarcane mosaic virus (Shukla *et al.*, 1987),

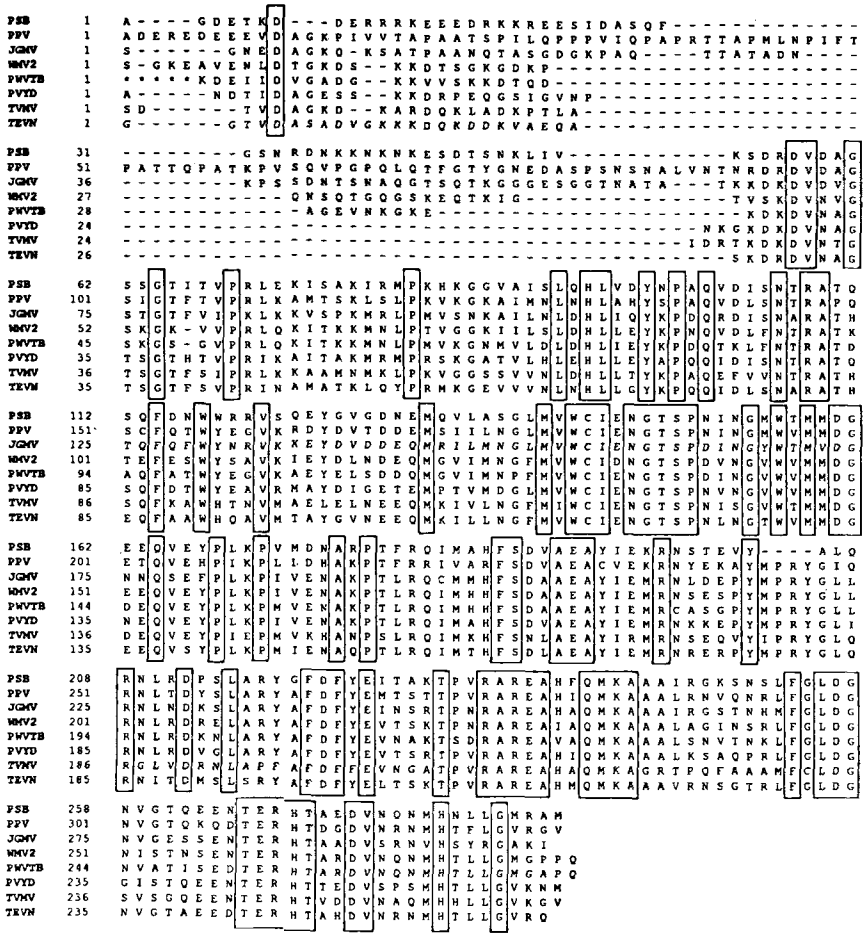

Fig. 2. Alignment of the amino acid sequences of eight potyviral coat proteins. Amino acids identical in all sequences are boxed. Literature references not cited in the text are: PPV, plum pox virus (Ravelonandro *et al.*, 1988); WMV-2, watermelon mosaic virus (Yu *et al.*, 1989); PWVTB, passionfruit woodiness virus strain TB (Shukla *et al.*, 1988*b*); PVYD (Shukla *et al.*, 1988*c*), TEVN, tobacco etch virus aphid non-transmissible (Allison *et al.*, 1985).

and the highly aphid-transmissible strain of TEV (Allison *et al.*, 1985). These N-terminally blocked potyviral coat proteins all start with serine. The coat protein of PSbMV is, however, predicted to start with an alanine.

An alignment of the PSbMV coat protein sequence with those of seven other distinct members of the potyvirus group is presented in Fig. 2. This alignment shows that the PSbMV coat protein has extensive sequence similarity to other potyviral coat proteins throughout its middle and C-terminal regions but, like other potyviruses, has little similarity in its N-terminal region. The amount of homology between the entire PSbMV coat protein and other potyviral coat proteins varies from 53·7% amino acid sequence identity with potato virus Y strain D (PVYD) to 43·2% with TVMV. The percentage sequence identity between these proteins is greater when only their middle and C-terminal regions are compared. The homology between the 238 C-terminal amino acids of PSbMV and the aligned amino acids from the other seven potyviral coat proteins in Fig. 2 varies from 65·1% for PVYD to 52·5% for TVMV. The PSbMV coat protein gene has a deletion of the codons encoding the sequence M-P-R-Y which is found from amino acids 244 to 247 in the PPV sequence (Fig. 2). This sequence has been observed in all other potyviral coat proteins sequenced to date. The nucleotide sequence for this region is identical for all three of the PSbMV cDNA clones examined, therefore the deletion of these four codons is unlikely to represent a cloning artefact or an anomaly occurring during cDNA synthesis.

The authors are grateful to Adrian Russell for providing infected plant material and to W. Bottomley for sequence analysis software. This research was funded in part by the New Zealand University Grants Committee.

References

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

ALLISON, R. F., SORENSON, J. C., KELLY, M. E., ARMSTRONG, F. B. & DOUGHERTY, W. G. (1985). Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: evidence for the synthesis of a polyprotein in potyvirus genome expression. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 3969–3972.

ALLISON, R., JOHNSON, 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.

BIRNBOIM, H. C. & DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513–1523.

BRAKKE, M. K. & VAN PELT, N. (1970). Properties of infectious ribonucleic acid from wheat streak mosaic virus. *Virology* **42**, 699–706.

CALDER, V. (1989). *A translation strategy and gene map for pea seedborne mosaic virus*. Ph.D. thesis, University of Canterbury.

D'ALESSIO, J. M., NOON, M. C., LEY, H. L. & GERARD, G. (1987). One-tube double-stranded cDNA synthesis using cloned M-MLV reverse transcriptase. *Focus* **9**, 1–4. Gaithersburg: Bethesda Research Laboratories.

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.

DOUGHERTY, W. G., CARY, S. M. & PARKS, T. D. (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* **171**, 356–364.

GOUGH, K. H., AZAD, A. A., HANNA, P. J. & SHUKLA, D. D. (1987). Nucleotide sequence of the capsid and nuclear inclusion protein genes from Johnson grass mosaic virus RNA. *Journal of General Virology* **68**, 297–304.

HAMPTON, R. O. & MINK, G. I. (1975). Pea seed-borne mosaic virus. *CMI/AAB Descriptions of Plant Viruses*, no. 146.

KHETARPAL, R. K. & MAURY, Y. (1987). Pea seed-borne mosaic virus: a review. *Agronomie* **7**, 215–224.

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.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

NELSON, R. S., POWELL-ABEL, P. & BEACHEY, R. N. (1987). Lesions and virus accumulation in inoculated transgenic tobacco plants expressing the coat protein of tobacco mosaic virus. *Virology* **158**, 126–132.

RAVELONANDRO, M., VARVERI, C., DELBOS, R. & DUNEZ, J. (1988). Nucleotide sequence of the capsid protein gene of plum pox potyvirus. *Journal of General Virology* **69**, 1509–1516.

REDDICK, R. R. & BARNETT, O. W. (1983). A comparison of three potyviruses by direct hybridization analysis. *Phytopathology* **73**, 1506–1510.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SHUKLA, D. D., GOUGH, K. H. & WARD, C. W. (1987). Coat proteins of potyviruses. 3. Comparison of amino acid sequences of the coat proteins of four Australian strains of sugarcane mosaic virus. *Archives of Virology* **96**, 59–74.

SHUKLA, D. D., STRIKE, P. M., TRACY, S. L., GOUGH, K. H. & WARD, C. W. (1988*a*). The N and C termini of the coat proteins of potyviruses are surface-located and the N terminus contains the major virus-specific epitopes. *Journal of General Virology* **69**, 1497–1508.

SHUKLA, D. D., MCKERN, N. M. & WARD, C. W. (1988*b*). Coat proteins of potyviruses. 5. Symptomatology, serology, and coat protein sequences of three strains of passionfruit woodiness virus. *Archives of Virology* **102**, 221–232.

SHUKLA, D. D., THOMAS, J. E., MCKERN, N. M., TRACY, S. L. & WARD, C. W. (1988*c*). Coat proteins of potyviruses. 4. Comparison of biological properties, serological relationships, and coat protein amino acid sequences of four strains of potato virus Y. *Archives of Virology* **102**, 207–219.

YU, M., FRENKEL, M. J., MCKERN, N. M., SHUKLA, D. D., STRIKE, P. M. & WARD, C. W. (1989). Coat proteins of potyviruses. 6. Amino acid sequences suggest watermelon mosaic virus 2 and soybean mosaic virus-N are strains of the same potyvirus. *Archives of Virology* **105**, 55–64.

(*Received 4 December 1989; Accepted 9 April 1990*)