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

Gail M. Timmerman,1* Victoria L. Calder2 and Lorraine E. A. Bolger2

1Crop Research Division, Department of Scientific and Industrial Research, Private Bag, Christchurch and 2Department 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 Smal-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 EcoRI and BamHI. 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)
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 a single termination codon (TAA) at nucleotide 1189. It encode the PSbMV coat protein as well as some of the 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 the genome. Direct chemical amino acid sequencing of the PSbMV coat protein was carried out on 200 pmol of purified PSbMV, which produces two protein bands of 40 and 28 kDa (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 Mr 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 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 predicted by examining the amino acid sequence of the mature viral proteins (Shukla et al., 1988a). 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),
shows that the PSbMV coat protein has extensive
region. The amount of homology between the entire
potato virus Y strain D (PVYD) to 43-2~ with TVMV.

The nucleotide sequence for this region is identical for all three of the PSbMV cDNA clones examined, therefore the deletion of these four nucleotides sequenced to date. The nucleotide sequence for this region has been observed in all other potyviral coat proteins sequenced to date. 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 this region 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.

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


(Received 4 December 1989; Accepted 9 April 1990)