Sequencing, genomic localization and initial characterization of the VPg of pea enation mosaic enamovirus

Christiane E. Wobus, Jihad S. Skaf, Marilou H. Schultz and Gustaaf A. de Zoeten

Department of Botany and Plant Pathology, Michigan State University, East Lansing, MI 48824, USA

The amino acid sequence of the genome-linked viral protein (VPg) of pea enation mosaic enamovirus (PEMV) has been determined. The VPg is encoded by nt 1811–1894 within ORF1 of RNA1 downstream of the protease motif. Direct N terminus sequencing of intact and endoproteinase Asp-N-digested VPg combined with electrospray mass spectroscopy confirmed that the VPg is composed of 28 amino acids with a molecular mass of 3138 Da. The context of the N and C terminus residues as well as the position and size of the VPg suggest that the mature VPg may be generated via post-translational proteolytic processing of the polyprotein arrangement of membrane anchor–proteinase–VPg–polymerase encoded by ORFs 1 and 2. Computer comparisons did not reveal any significant similarity between the VPg of PEMV and any other sequences including those of the VPgs of related subgroup II luteoviruses.

Pea enation mosaic virus (PEMV) is the sole member of the genus Enamovirus (Demler et al., 1996). Its small isometric particles contain two taxonomically distinct, single-stranded, (+)-sense RNAs. The larger RNA (RNA1, 5706 nt) is notably similar to the RNAs of subgroup II luteoviruses in nucleotide sequence and in gene arrangement (Demler & de Zoeten, 1991; Demler et al., 1996). The second RNA (RNA2, 4253 nt) resembles the RNAs of the genus Umbravirus. The polymerase encoded by RNA2 is related to the polymerases of carmoviruses but is distinct from that of RNA1 (Demler et al., 1993). Despite the replicative independence of RNA1 and RNA2 in protoplasts, both are required for wild-type infections. RNA1 provides encapsidation and aphid transmission functions while RNA2 facilitates systemic invasion and mechanical transmission (Demler et al., 1994, 1996).

Many single-stranded viral RNAs encode a protein (VPg) that attaches covalently to the 5′ terminus of the encapsidated RNA (Wimmer, 1982; Vartapetian & Bogdanov, 1987).

Likewise, the 5′ terminus of the encapsidated RNA of PEMV was shown to possess a VPg with an apparent molecular mass of 17.5 kDa as determined by electrophoresis (Reisman & de Zoeten, 1982). The covalent nature of the linkage was determined after the protein–RNA association resisted treatments designed to release non-covalently bound proteins (i.e. heating in the presence of SDS, EDTA and β-mercaptoethanol followed by repeated phenol–chloroform extractions). Sequencing the PEMV genome (Demler & de Zoeten, 1991; Demler et al., 1993) did not reveal a discrete open reading frame (ORF) corresponding to the 17.5 kDa VPg. However, sequence analysis identified a putative linear arrangement of transmembrane region–proteinase–polymerase domains in ORF1 of RNA1 (Demler & de Zoeten, 1991) reminiscent of the membrane binding domain–VPg–proteinase–polymerase arrangement found in picorna-like viruses (Domier et al., 1987). Based on this analogy, Demler et al. (1996) suggested that the PEMV VPg may be encoded by sequences upstream of the proteinase motif. Van der Wilk et al. (1997) determined the 32 N-terminal amino acids of the VPg of potato leaf roll luteovirus (PLRV), a close relative of the RNA1 component of PEMV. The VPg of PLRV was found to be encoded downstream of the putative proteinase motif.

The aim of this study was to take advantage of the relatively high yield of virus that can be purified from pea (Pisum sativum L. cv. 8221) to isolate and analyse the VPg of PEMV and resolve its genomic origin.

The naturally occurring, aphid-nontransmissible ‘deletion’ strain PEMV-AT-∆ (Demler et al., 1996) was propagated in pea and virus particles were purified as previously described (German & de Zoeten, 1975; Reisman & de Zoeten, 1982). RNA–VPg complex was isolated by extracting purified virions three times with phenol–chloroform followed by ethanol precipitation.

To free the VPg for microsequencing, 500 µg aliquots of the RNA–VPg complex were either incubated with 3 units of RNase T1 for 10 h (Zalloua et al., 1996) or hydrolysed in 20% trifluoroacetic acid (TFA) for 48 h at room temperature (Pinck et al., 1991) followed by totally evaporating the TFA in a fume hood. Five separate N terminus amino acid sequencings based on Edman degradation (Protein Sequencer, Procise model 494, Applied Biosystems) revealed that the VPg of PEMV is composed of 28 amino acids encoded by nt 1811–1894 within...
ORF1 of RNA1, downstream of the putative 3C-like proteinase (Fig. 1).

To confirm that the 28 amino acid sequence represents the entire VPg and not just the sequence of the N-terminus and to rule out that a larger protein (either a precursor or a larger VPg) is attached to viral RNA but went undetected due to premature termination of the sequencing reaction at the 28th amino acid, two approaches were used.

(1) Electrospray mass spectroscopy (ESMS, Micromass Platform) of aliquots of RNase T1- or TFA-treated VPg confirmed the absence of a larger protein attached to encapsidated viral RNA.

(2) To verify that the 28th residue (glutamine) does indeed represent the true C terminus of the VPg, aliquots of RNA-free VPg were resuspended in 100 mM Tris–HCl, pH 8.0, and digested for 10 h at 37 °C with 2% (w/w) of the endoproteinase Asp-N (Sigma), which cleaves on the N terminus of aspartic acid (D). The sequence shown in Fig. 1 indicates the presence of two cleavage sites at amino acids D-13 and D-17. The sequence of RNA1 (Demler & de Zoeten, 1991) shows that the next aspartic acid encoded by ORF1 is 24 amino acids downstream of D-17 (nt 1991–1993). The resulting peptides were separated using a 0.8 × 250 mm C18 column (LC Packings) equilibrated with 0-1% TFA. A 5–50% gradient of acetonitrile–water (90:10, v/v) + 0-1% TFA was developed over 135 min. The separated peptides were subjected to N terminus sequencing as described above. Consistent with the sequence in Fig. 1, two peptides were obtained as a result of Asp-N digestion of the VPg. The sequence of the first peptide was AYSEAETAYEM, which is identical to the 12 N-terminal amino acids of the VPg. The sequence of the second peptide consisted of the 12 amino acids starting at D-17 and terminating unambiguously and reproducibly at Q-28. The tetrapeptide 13-DNFS-16 was not resolved in this system.

Demler & de Zoeten (1991) identified a putative 3C-like proteinase motif in the 84 kDa polyprotein encoded by ORF1, suggesting the processing of expressed viral proteins in infected tissues. The size and location of the VPg suggest that the mature VPg may be released from a putative precursor via the proteolytic activity of that proteinase. This polyprotein processing in PEMV is supported by computer-aided analysis of the secondary structure predictions of the 84 kDa polyprotein using the University of Wisconsin Genetics Computer Group (UWGCG) PEPTIDESTRUCTURE program (Devereux et al., 1984) and the PHD protein mail server at EMBL, Heidelberg, Germany (Rost & Sander, 1993, 1994a, b; Rost, 1996). This analysis identified solvent-accessible surface loops at or near the N and C termini of the VPg of PEMV. Such loops can be utilized by 3C-like proteinases from the genera Picorna-, Como-, Nepo- and Potyvirus (Arnold et al., 1987; Dougherty & Semler, 1993; Gorbalenya & Koonin, 1993; Pallai et al., 1989; Wellink & van Kammen, 1988). For instance, the N terminus of PEMV-AT-VPg can be generated by a cleavage between glutamic acid and alanine (Fig. 1), a cleavage site also used by 3C-like proteinases from picorna- and nepoviruses (Wellink & van Kammen, 1988). Also, the C terminus of the VPg is at the junction between glutamine and serine (Fig. 1). Such a Q–S cleavage site is common for picornaviruses as well as comoviruses and potyviruses (Wellink & van Kammen, 1988). In addition, comparative analysis of protein domain boundaries in polyproteins of luteo- and sobemoviruses predicted cleavage sites at Q, E/G, S, A (Gorbalenya & Koonin, 1993).

It is not clear what factors contributed to the discrepancy between the calculated molecular mass of the VPg reported in this paper (3138 Da) and the 17.5 kDa observed by Reisman & de Zoeten (1982). It is known that other VPgs migrate anomalously in SDS–PAGE due to their cationic character (Daubert & Bruening, 1984). Interestingly, PEMV-VPg is
anionic at neutral pH (−5.02 as determined by UWGCG ISOELECTRIC). An alternative explanation is that residual ribonucleotides remained attached to the VPg after the RNase T1 and RNase A thus contributing to the slower mobility of the VPg in SDS–PAGE described by Reisman & de Zoeten (1982).

Amino acid and nucleotide sequence comparisons using the UWGCG programs FASTA, TFASTA and BLAST did not reveal any statistically significant similarities between the VPg of PEMV and any other viral sequences. This is in sharp contrast with the findings of van der Wilk et al. (1997), who identified significant similarities (53–56%) within the P1 sequences of the luteoviruses beet Western yellows, cucurbit aphid-borne yellows, beet mild yellowing and PLRV. Fig. 1 shows the results of a directed alignment (performed using the UWGCG GAP program) between the consensus sequence of the VPgs of the above-mentioned luteoviruses (van der Wilk et al., 1997) and that of the VPg of PEMV. The identity of 25% and the similarity of 33% found in this alignment need to be contrasted with the identity of 48% and similarity of 62% that are obtained when the same comparison is performed between the consensus sequence and that of the VPg of PLRV. This lack of VPg sequence homology may represent yet another point of divergence between PEMV and related luteoviruses. Only the delineation of the function of these VPgs in the infection cycles of their respective viruses will help clarify the significance of that divergence.

Note added in proof. A recent total amino acid composition analysis verified that the VPg is composed only of the 28 amino acids presented in this report.

The first and second authors contributed equally to this report. We would like to thank Dr Joe Leykam for his assistance in protein sequencing and ESMS. This work was supported by the College of Natural Sciences and by the Agricultural Experimental Station (project 1655) of Michigan State University.

References


