Unusual amino-terminal sequence repeat characterizes the capsid protein of dasheen mosaic potyvirus

S. S. Pappu, H. R. Pappu, E. P. Rybicki and C. L. Niblett

1 Plant Pathology Department, University of Florida, Gainesville, Florida 32611-0680, U.S.A. and 2 Department of Microbiology, University of Cape Town, Rondebosch 7700, South Africa

The 3'-terminal region of a Florida isolate of dasheen mosaic potyvirus (DMV-LA) genome including the coat protein (CP) gene was cloned and sequenced. Protease digestion was predicted to occur between the glutamine and alanine residues at positions 79 and 80 of the 408 residue long polypeptide to produce a CP of 329 amino acids with an estimated Mr of 36229. Following the putative protease recognition site is a DAG sequence, which is conserved among aphid-transmitted potyviral CPs. There is an unusual and unique stretch of 52 amino acids after the DAG that is repetitive and rich in threonine and asparagine. A sequence of 10 residues (GNNTNTNTT) was repeated three times in tandem within this stretch and was followed by six proline residues. Several potential glycosylation sites were found clustered within this region. Expression in Escherichia coli and Western blotting of the CP confirmed its size and serological identity. Sequence comparisons and phylogenetic reconstructions indicated that DMV is a distinct potyvirus within the passionfruit woodiness virus subgroup cluster.

Dasheen mosaic virus (DMV) is an aphid-transmitted potyvirus that infects plants belonging to the family Araceae. It causes severe mosaic with yield losses of up to 60% in edible aroids and reduced productivity in many ornamental aroids (Zettler & Hartman, 1987). The viral genome is a single-stranded RNA of M$_r$ 3 x 10$^6$ to 3.5 x 10$^6$ (Zettler et al., 1978). The coat protein (CP) of DMV is serologically related to those of other potyviruses (Abo El-Nil et al., 1977; Li et al., 1992). However, no molecular information is available on any of the genes encoded by DMV. We cloned and sequenced the 3'-terminal 1475 nucleotides of the DMV genome which includes the CP gene and the 3' non-coding region (3'NCR) to learn more about this important virus and its molecular relationships to other potyviruses.

Crude viral nucleic acids were prepared from leaf tissue infected with the Florida DMV isolate LA (Pappu et al., 1993b). Reverse transcription (RT) and PCR amplification (Saiki et al., 1985) reactions (RT–PCR) were performed as previously described (Pappu et al., 1993b). The amplified fragment was cloned into Smal-cut pUC118 and sequenced as previously described (Pappu et al., 1993b). For Western blots, overnight cultures of Escherichia coli strain DH5α containing the recombinant plasmid with the CP gene were used. The bacterial cells were pelleted, and lysed by the boiling method (Maniatis et al., 1982). The total protein in the supernatant was precipitated using a one-half volume of 7.5 M-ammonium acetate. The protein in the pellet was dissociated by boiling in 25 μl of 2x cracking buffer (0.125 M-Tris–HCl pH 6.8, 4% SDS, 20% glycerol and 10% 2-mercaptoethanol) prior to electrophoresis. Infected leaf tissue extracts were prepared as described by Li et al. (1991). Protein samples were electrophoresed on a 12% SDS–PAGE gel as previously described (Li et al., 1991) and were electroblotted to a nitrocellulose membrane (Stratagene). The coat protein was detected with polyclonal antisera (1:500 dilution) raised against the recombinant plasmid with the CP gene were used. The bacterial cells were pelleted, and lysed by the boiling method (Maniatis et al., 1982). The total protein in the supernatant was precipitated using a one-half volume of 7.5 M-ammonium acetate. The protein in the pellet was dissociated by boiling in 25 μl of 2x cracking buffer (0.125 M-Tris–HCl pH 6.8, 4% SDS, 20% glycerol and 10% 2-mercaptoethanol) prior to electrophoresis. Infected leaf tissue extracts were prepared as described by Li et al. (1991). Protein samples were electrophoresed on a 12% SDS–PAGE gel as previously described (Li et al., 1991) and were electroblotted to a nitrocellulose membrane (Stratagene). The coat protein was detected with polyclonal antisera (1:500 dilution) raised against the DMV-FL isolate (Abo El-Nil et al., 1977). The oligonucleotide (CN156) that was used as the upstream, genome sense primer contained the sequence 5' ATG AT(ACT) GA(AG) (GT)C(ACGT) TGG GG Y and was derived from the Nib gene region. Oligo(dT) (Pappu et al., 1993b) primers with a degenerate 3' end were used as downstream, anti-genome sense primers. RT–PCR using the above primer pair resulted in a fragment of approximately 1.5 kb.

Sequence analysis (Rhoads & Roufa, 1985; Devereux et al., 1984) revealed an uninterrupted reading frame of 408 amino acid residues starting with the MIEAWG...
sequence and ending at a stop codon 251 nucleotides from the 3' end of the clone (Fig. 1). Protease cleavage sites in potyvirus polyprotein sequences for the production of the CP are between Q/A, Q/S or Q/G residues or, rarely, a Q/V site as reported for lettuce etch virus (TEV) and tobacco vein mottling virus (TVMV). The first protease site results in a CP 329 amino acids long. Thus the proposed DMV CP is the second largest reported (after TVMV), further indicating that the protease digestion is more likely to occur at site 1. Protease digestion at this site agrees with that seen with tobacco mosaic virus. The presence of the sequence DEVVL upstream of the glutamine at position 79 also correlates with the protease site context observed for other potyviruses (Yu et al., 1989). The sequence following site 1 also agrees with that seen with tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV), further indicating that the protease digestion is more likely to occur at site 1. Protease digestion at this site results in a CP 329 amino acids long. Thus the proposed DMV CP is the second largest reported (after potato virus, PPV, strain D; Ravelonandro et al., 1988) of aphid-transmitted potyviral CPs.

The N-terminal region of the DMV CP included an unusual stretch of 52 amino acids that were threonine- and asparagine-rich and was repetitive. A sequence of 10 amino acid residues (GNNTNTNTNT) was repeated three times in tandem within this TN-rich sequence (Fig. 1). Several potential N-glycosylation sites were revealed within the CP gene coding sequence as identified by the GCG program MOTIFS (Devereux et al., 1984). They are shown double underlined in Fig. 1. Most of these N-glycosylation sites were clustered within the TN-rich sequence. A group of six proline residues was found at the end of the TN repeat sequence (Fig. 1).

The putative CP was expressed in E. coli to ensure that the proposed gene did in fact encode the genuine viral capsid protein. The start codon was introduced at position 79 of the polyprotein by mutating the glutamine codon (CAG) to the start codon ATG, by means of PCR using the oligonucleotides CN178 (5' GAATTCACC-ATGGCTAGCAGACACAGTTG 3') and CN179 (5' AGATCTTTACTGTGGAGATGCCAC 3'). These sequences were clustered within the TN-rich sequence. A group of six proline residues was found at the end of the TN repeat sequence (Fig. 1).

The putative CP sequence and deduced amino acid sequence of the 3'-terminal region of the DMV-LA genome. The putative protease site (Q/A) of the CP and the repeat sequences are shown underlined. Residues conserved in potyviral CPs in general, both upstream and downstream of the start of the CP (or protease site) are shown in boldface. The potential N-glycosylation sites are double underlined. The 3'NCR sequence is shown in Pappu et al. (1993b).
Fig. 2. Western blot analysis of the DMV-LA CP expressed in E. coli. Lane 1, protein from E. coli DH5α carrying the pUC118 vector without the CP gene; lane 2, protein preparation from E. coli containing plasmid with the DMV CP gene; lane 3, protein extract from uninfected tissue; lane 4, an extract from DMV-infected tissue. The antiserum did not react with proteins from E. coli without the CP gene (Fig. 2, lane 1) or from uninfected tissue (Fig. 2, lane 3). These results indicate that the protein expressed from protease site 1 is similar in size to the CP expressed in infected plant tissue (Fig. 2). The apparent slower migration of the E. coli-expressed CP may be due to the expression of the CP in bacteria from the start codon of the β-galactosidase gene in the pUC118 vector, which adds 14 amino acids to the amino terminus of the CP.

The possibility that the TN stretch resulted from an insertional event or an artefact of cloning was investigated by sequencing seven independent clones obtained from four different DMV-infected tissue sources. These produced sequences identical to the original clone and eliminated the possibility of cloning artefacts (results not shown). It is possible that DMV-LA acquired this sequence by a recombinational event during evolution.

Comparisons of the TN sequence region with sequences in the nucleic acid and protein sequence databases using the BLAST program (Altschul et al., 1990) revealed significant similarities with the extracellular protein p60 of Listeria monocytogenes. This protein contains 19 repeats of threonine and asparagine (Kohler et al., 1990) and is required for the adherence to and invasion of mouse fibroblast cells (Kuhn & Goebel, 1989). Similarly, proline-rich sequences were found in the 56K readthrough proteins of luteoviruses (Bahner et al., 1990) and the 33K protein of pea enation mosaic virus (Demler & de Zoeten, 1991) that are implicated in the aphid transmission of these viruses. It is suggested for these viruses that the structure-disrupting tendency of the prolines might play a role in making that region of the protein available to some unknown receptors in the aphid. The prolines found at the end of the TN-repeat sequence in DMV may be playing a similar role in the aphid transmission of the virus.

The sequence of the 3′NCR of DMV-LA was recently published (Pappu et al., 1993b). Pairwise sequence comparisons in that study revealed sequence similarities of 34 to 44% between DMV and other potyviruses, indicating DMV to be quite distinct. In this report, similar comparisons of the DMV CP were made with other potyviral CPs. Sequence similarities ranged from 40.86% (wheat streak mosaic virus) to 83.27% (soybean mosaic virus-N, SbMV-N), and are in agreement with the 3′NCR data. Stronger similarities were found with viruses of the passionfruit woodiness virus (PWV) subgroup (Rybicki & Shukla, 1992; Brand et al., 1993).

Multiple sequence alignments and phylogenetic reconstructions were performed using CLUSTAL V (Higgins et al., 1992) and NJTREE (Saitou & Nei, 1987). The sequences of two CPs (223 amino acids long) from distinct isolates of citrus tristeza virus were used as
unrelated outgroups to test the significance of the alignments (Pappu et al., 1993a). The CP phylogeny places DMV in a subgroup cluster with other PWV-like viruses, but as a distinct species with no close affinities to any other virus (Fig. 3).

Previous studies on the DMV CP indicated that it is serologically related to the CPs of blackeye cowpea mosaic virus (BICMV) and TEV but not to those of potato virus Y (PVY) or turnip mosaic virus (TuMV) (Abo El-Nil et al., 1977). The relationship between the CPs of BICMV and DMV can be explained by the fact that both are members of the PWV subgroup (Rybicki & Shukla, 1992). Similarly, the lack of relationship between the CPs of DMV, PVY and TuMV is probably due to the distinct generic relationship between DMV and the other two (Fig. 3; Rybicki & Shukla, 1992; Ward et al., 1992). The cross-reaction between the CPs of DMV and TEV is difficult to explain, as these viruses are as distantly related as DMV is to PVY or TuMV. The apparent relatedness of the two (DMV and TEV) is probably one of the unexpected paired relationships noted elsewhere for this group of viruses (Shukla et al., 1992). Our observations once again support the significance of CP and 3'NCR sequence analysis in assessing the taxonomic relationships among potyviruses.

We thank Dr F. W. Zettler for the gift of the DMV-FL antisera, Drs R. F. Lee and K. S. Derrick for their suggestions on the manuscript, and Dr J. Carrington for helpful discussions. This work was supported in part by a grant from the Caribbean Basin Advisory Group to C.L.N. We are also grateful to the Interdisciplinary Center for Biotechnology Research of the University of Florida and the Information Technology Service of the University of Cape Town for providing computing facilities. Florida Agricultural Experiment Station Journal Series no. R-03467.

References


(Received 21 May 1993; Accepted 3 September 1993)