The nucleotide sequence of apple mosaic virus coat protein gene has no similarity with other Bromoviridae coat protein genes

Rudaina H. Alrefai, Patrick J. Shiel, Leslie L. Domier, Cleora J. D’Arcy, Philip H. Berger and Schuyler S. Korban

1 University of Illinois, Department of Horticulture, Plant and Animal Biotechnology, Urbana, Illinois 61801, 2 University of Idaho, Department of Plant, Soil and Entomological Sciences, Moscow, Idaho 83843, 3 USDA/ARS/MWA, Crop Protection Research Unit, Turner Hall, Urbana, Illinois 61801 and 4 University of Illinois, Department of Plant Pathology, Turner Hall, Urbana, Illinois 61801, U.S.A.

A double-stranded cDNA was synthesized from in vitro polyadenylated apple mosaic virus (ApMV) RNA 3 using oligo(dT) and sequence-specific primers, and was cloned into plasmid vectors. A set of overlapping cDNA clones was used to determine the nucleotide sequence of RNA 4. ApMV RNA 4 was found to contain an open reading frame (ORF) of 666 nucleotides, which was flanked by 5’ and 3’ non-translated sequences of 55 and 264 nucleotides, respectively. The ORF encoding the coat protein was identified by comparing the predicted amino acid sequence with that obtained from direct protein microsequencing of the native viral coat protein. The ORF encodes a protein with an M, of 25056. The nucleotide sequence of the ApMV coat protein gene showed no similarity to those of alfalfa mosaic virus, tobacco streak virus (TSV), brome mosaic virus or cucumber mosaic virus. The predicted amino acid sequence of the amino-terminal region of the ApMV coat protein is basic, rich in cysteine residues and may contain a zinc finger motif similar to that found in TSV.

Apple mosaic virus (ApMV) is a member of subgroup III of the ilarvirus group (Francki et al., 1991). It occurs as isometric or quasi-isometric labile particles and often produces ringspot symptoms on hosts. ApMV has been purified from inoculated cucumber seedlings and separated into two bands on 10 to 40 % sucrose gradients (DeSequeira, 1967). Ilarviruses are members of the Bromoviridae family and have tripartite ssRNA genomes. Their coat proteins are translated from subgenomic mRNAs (van Vloten-Doting et al., 1981). Ilarvirus tripartite genomes consist of two monocistronic RNAs (RNAs 1 and 2) and a dicistronic RNA 3, from which the subgenomic RNA 4, coding for the coat protein, is synthesized (van Vloten-Doting & Jaspars, 1977). This project was undertaken to clone and sequence a cDNA copy of ApMV RNA 4 and to characterize the coat protein of the virus.

Apple mosaic virus, isolated from apples in Idaho, U.S.A., and detected by ELISA using ApMV antiserum obtained from the ATCC, was transferred and isolated from infected cucumber cotyledons (Cucumis sativus L. cv. Lemon; DeSequeira, 1967) by the acidification method of Ong & Mink (1989). The virus was further purified in 10 to 40% sucrose density gradients (DeSequeira, 1967) and all virus-containing fractions (top and bottom) were collected separately. All purification steps were carried out at 4 °C within a single day; purified virus was stored at −80°C to maintain infectivity. RNAs were extracted from purified virions of the top and bottom fractions of the sucrose gradient with phenol (Stenger et al., 1987).

Analysis of ApMV RNA using 1·2% agarose gel electrophoresis revealed that the lower fractions present in the sucrose density gradients contained virions with RNAs 1 and 2. These RNAs had estimated sizes of 3·25 and 3·0 kb respectively on the gel (Fig. 1). RNAs 1 and 2 of alfalfa mosaic virus (AlMV) are monocistronic with sizes of 3·6 kb (Cornelissen et al., 1983 a) and 2·6 kb (Cornelissen et al., 1983 b) respectively. The upper fraction in the sucrose gradients contained virions with two smaller RNA species: RNAs 3 and 4 of approximately 2·0 kb and 0·95 kb respectively (Fig. 1). The sizes of RNAs 3 and 4 have been reported to be 2·1 and 0·9 kb in AlMV (Barker et al., 1983) and 2·2 and 1·1 kb for tobacco streak virus (TSV; Cornelissen et al., 1984). RNA 3 of both TSV and AlMV are bicistronic, with the second open reading frame (ORF) being translated from the subgenomic RNA 4 (Barker et al., 1983; Cornelissen et al., 1984).

Since RNA 4 is contained within RNA 3, RNA 3 was further purified by preparative gel electrophoresis on
1.2% low-melting temperature agarose (NuSieve; FMC) gels (Sambrook et al., 1989). Purified ApMV RNA 3 was polyadenylated using Escherichia coli poly(A) polymerase (BRL; Devos et al., 1976) and was used as a template to synthesize first-strand DNA using an oligo(dT)–BamHI primer [5' AGAGAGAGACATAGTGGATCC(T)₁₈ 3']; Krug & Berger, 1987]. A poly(dG) tail was added to the 3' end of first-strand cDNA and the second strand was synthesized by PCR using the XEC primer [5' CGACTGGTCTAGAATT(C)₁₈ 3']; Cornelissen et al., 1984). The nucleotide sequence of ApMV RNA 4 has no apparent similarity to the 3' region of TSV or AIMV RNA 4, except for the presence of several inverted repeats. It also contains one inverted palindromic repeat at position 911. Although there is virtually no similarity in the primary structures of the 3' terminal regions of AIMV and TSV RNAs, they all have stable hairpin structures flanked by the repeating tetranucleotide sequence AUGC (Koper-Zwarthoff & Bol, 1980). These structures may be responsible for the mutual recognition of AIMV and TSV RNAs by each others' coat proteins (Cornelissen et al., 1984). The nucleotide sequence of RNA 4 of another strain of ApMV, PV-32 from ATCC, has been determined (Sánchez-Navarro & Pallás, 1994; GenBank accession number U03857). The nucleotide sequence of ApMV RNA 4 of the Idaho strain showed little homology to that of PV-32. In addition, the 3' non-translated region of the ApMV RNA 4 of the Idaho strain was found to be 100 nucleotides longer than that of PV-32.

**ApMV coat protein** was extracted from intact virions, dialysed against distilled water and used for amino acid microsequencing. The amino-terminal sequence of the coat protein was determined using an Applied Biosystems Automated Gas Phase Protein Sequencer (Beyreuther et al., 1983) and its Mₑ was determined by electrophoresis on a 10 to 20% SDS–polyacrylamide gel (Laemmli, 1970).

A set of several overlapping clones of RNA 3 was used to deduce the nucleotide sequence of ApMV RNA 4. Nucleotide sequences of at least five clones obtained from in vitro polyadenylated RNA 3 were found to be 3'-coterminal and were therefore assumed to contain the 3' sequence of RNA 4. Based on this assumption, a second set of four cDNA clones from RNA 3 was generated using primer 1. All four clones were completely sequenced on both strands using primers to pBluescript and oligonucleotide primers generated during sequencing. Additional cDNA clones were generated using primer 2 (5' GGCACTGGTCTAGAATT 3') to encompass the nucleotide sequence of ApMV RNA 4. Using the cloning methods mentioned above, one clone from polyadenylated RNA and two clones from native RNA were sequenced. ApMV RNA 4 was found to contain an ORF of 666 nucleotides, beginning at nucleotide 55 and ending at nucleotide 721 (Fig. 2). Inverted palindromic repeats of 18, eight and eight nucleotides were observed at positions 383, 665 and 911. The 3' non-translated region was found to be 264 bases in length, with several inverted repeats (Fig. 2). The transcriptional starting point of AIMV RNA 4 (Barker et al., 1983), brome mosaic virus (BMV; Ahlquist et al., 1981) and cucumber mosaic virus (CMV; Gould & Symons, 1982) is preceded by a C residue, and the first base of the subgenomic RNA is a G nucleotide. The ApMV RNA 4 may also start at G (Fig. 2). The 3' non-translated region of ApMV RNA 4 has no apparent similarity to the 3' region of TSV or AIMV RNA 4, except for the presence of several inverted repeats. It also contains one inverted palindromic repeat at position 911. Although there is virtually no similarity in the primary structures of the 3' terminal regions of AIMV and TSV RNAs, they all have stable hairpin structures flanked by the repeating tetranucleotide sequence AUGC (Koper-Zwarthoff & Bol, 1980). These structures may be responsible for the mutual recognition of AIMV and TSV RNAs by each others' coat proteins (Cornelissen et al., 1984). The nucleotide sequence of RNA 4 of another strain of ApMV, PV-32 from ATCC, has been determined (Sánchez-Navarro & Pallás, 1994; GenBank accession number U03857). The nucleotide sequence of ApMV RNA 4 of the Idaho strain showed little homology to that of PV-32. In addition, the 3' non-translated region of the ApMV RNA 4 of the Idaho strain was found to be 100 nucleotides longer than that of PV-32.

**Fig. 1.** Formaldehyde denaturing (1.2%) agarose gel electrophoresis of ApMV RNAs. Lane 1, RNA Mₛ standards; lane 2, RNA of ApMV from top and bottom fractions of sucrose density gradients; lane 3, RNA of AIMV from top and bottom fractions of sucrose density gradients; lane 4, RNA from the satellite virus of St Augustine Decline (strain N, 824 nucleotides); lane 5, RNA from potato leafroll virus (South Idaho isolate).
Fig. 2. The complete nucleotide sequence of ApMV RNA 4 and the predicted amino acid sequence corresponding to the single ORF. Amino acids at the amino terminus of the coat protein that were determined by direct microsequencing of ApMV are shown in bold. The inverted palindromic repeats are underlined. The transcriptional start nucleotide (g) is double-underlined.

Microsequence analysis of the coat protein (Fig. 2) of ApMV agreed with that deduced from the nucleotide sequence. The amino terminus of the coat protein is basic, with three residues of lysine, arginine and histidine, and five cysteine residues that can be arranged in a zinc finger configuration (Cys-X$_2$-Cys-X$_3$-His-X$_3$-Cys) similar to TSV (Sehnke et al., 1989). An even distribution of basic and acidic amino acids was observed in the carboxy terminus. The coat protein ORFs of CMV, BMV and TSV have been reported to be 711 (Gould & Symons, 1982), 579 (Ahlquist et al., 1981) and 714 nucleotides in length (Cornelissen et al., 1984) respectively. No significant similarity was seen when either the nucleotide sequence of RNA 4 or the predicted amino acid sequence of the coat protein of ApMV was compared to those of TSV, AlMV, BMV or CMV. This is in concordance with earlier reports showing that coat proteins of TSV and AlMV were not serologically related and that their predicted amino acid sequences were not homologous (Cornelissen et al., 1984; van Vloten-Doting, 1975).

The ApMV coat protein gene will be used in constructs for transfer into ApMV-susceptible apple genotypes. Studies on gene expression will elucidate the feasibility of this system in providing coat protein-mediated resistance for apple genotypes against ApMV and possibly other viruses.

This project was supported in part by the nursery industry through contributions to the Horticultural Research Institute (a non-profit organization devoted to nursery industry progress through research), a grant from the University of Illinois Campus Research Board and the University of Illinois Agricultural Experiment Station project number 65-325.

References


(Received 27 January 1994; Accepted 26 April 1994)