Nucleotide Sequence of Papaya Mosaic Virus RNA

By T. L. SIT, M. G. ABOUHAIDAR* AND S. HOLY

Department of Botany, University of Toronto, Toronto, Ontario M5S 3B2, Canada

(Accepted 26 May 1989)

SUMMARY

The RNA genome of papaya mosaic virus is 6656 nucleotides long [excluding the poly(A) tail] with six open reading frames (ORFs) more than 200 nucleotides long. The four nearest the 5' end each overlap with adjacent ORFs and could code for proteins with \( M_r \) 176307, 26248, 11949 and 7224 (ORFs 1 to 4). The fifth ORF produces the capsid protein of \( M_r \) 23043 and the sixth ORF, located completely within ORF 1, could code for a protein with \( M_r \) 14113. The translation products of ORFs 1 to 3 show strong similarity with those of other potexviruses but the ORF 4 protein has only limited similarity with the other potexvirus ORF 4 proteins of 7K to 11K.

INTRODUCTION

Particles of the potexvirus papaya mosaic virus (PMV) contain a single-stranded positive sense RNA of \( M_r \) \( 2.2 \times 10^6 \) (Koenig, 1971). An m7GpppG cap structure was found at the 5' terminus (AbouHaidar & Bancroft, 1978) and the 3' end has a poly(A) tail (AbouHaidar, 1988). In vitro translation of genomic PMV RNA produced three polypeptides of estimated \( M_r \) 155K, 73K and 22K (Bendena et al., 1985). The 22K product was shown to be the capsid protein by immunoprecipitation with antibodies raised to PMV. Two 3' coterminal subgenomic RNAs of 1.0 and 2.2 kb have been detected in PMV-infected tissue (Mackie et al., 1988). Immunoprecipitation of the in vitro translation products of the purified smaller subgenomic RNA showed that it codes for the PMV capsid protein but the product of the 2.2 kb subgenomic RNA remains unknown (Mackie et al., 1988).

The \( 5' \)-terminal 138 nucleotides of PMV genomic RNA have been directly sequenced by Lok & AbouHaidar (1986), and the 3'-terminal 900 nucleotides were sequenced from cDNA clones to reveal the coding region for the capsid protein (AbouHaidar, 1988).

The complete nucleotide sequences of three other potexviruses have been reported, those of narcissus mosaic virus (NMV) (Zuidema et al., 1989), potato virus X (PVX) (Huisman et al., 1988) and white clover mosaic virus (WC1MV) (Forster et al., 1988). Each of these potexvirus RNAs contains five conserved open reading frames (ORFs) with ORFs 2 to 4 overlapping adjacent ORFs and a possible ORF 6 product encoded within ORF 5 (the capsid protein) of WC1MV and NMV (Zuidema et al., 1989). In this paper, we report the complete nucleotide sequence of PMV RNA and compare the ORF products with those of other potexvirus RNAs.

METHODS

cDNA synthesis. Virus and RNA extraction and purification procedures were those previously described (Erickson & Bancroft, 1978; Lok & AbouHaidar, 1986). First-strand cDNA synthesis was primed with oligo(dT)\(_{12-18}\) or a specific primer (bases 228 to 247 in the PMV sequence). For second-strand synthesis either the RNase H method was used (Gubler & Hoffman, 1983) with GC-tailing to fit the pUC18 PstI site or a specific primer was used which was complementary to the 5'-terminal 17 nucleotides (Lok & AbouHaidar, 1986). By this latter method the cDNA obtains a restrictive endonuclease site at its 5' terminus which is convenient for cloning the cDNA by the method of Schmid et al. (1987).

DNA sequencing. Subclones of the various cDNA clones were generated by restriction digestion and religation of the recombinant plasmid. Double-stranded plasmid DNA was sequenced directly by the dideoxy nucleotide method (Sanger et al., 1977) using \([\text{32P}]\text{dATP}\) (Biggin et al., 1983) and modified bacteriophage T7 DNA polymerase (Sequenase; U.S. Biochemical) (Tabor & Richardson, 1987). Sequencing primers for pUC18 (M13...
universal and reverse primers; Pharmacia) as well as oligonucleotide primers for PMV were used for sequence
determination. Plasmid templates were linearized with \textit{PvuII}, boiled for 3 min with primer and cooled quickly on
ice prior to the sequencing reaction. All sequence was determined using DNA in both orientations.

\textit{Amino acid sequence comparisons.} The FASTP computer program (Lipman & Pearson, 1985) was used to
determine amino acid similarities between polypeptides.

\textbf{RESULTS}

\textit{Cloning and sequence analysis}

Four cDNA inserts (clones TP11, CP21, T21 and H3) spanned most of the PMV genome except for the 5' terminal region (Fig. 1). The complete nucleotide sequence of genomic PMV RNA is shown in Fig. 2. The viral genome is 6656 nucleotides in length excluding the poly(A) tract at the 3' terminus which varies in length from 50 to 125 residues (AbouHaidar, 1988). Comparison of the previously reported 5' sequence of PMV RNA (Lok & AbouHaidar, 1986) with the cDNA sequence obtained shows four differences, at positions 45, 55, 96 and 138, where C residues were found. Replacements were also found at the 3' end of the RNA and are indicated below the sequence. They include an extra CAC triplet at position 6118 producing an extra threonine residue (Fig. 2). This threonine had been reported in the amino acid sequence of the PMV capsid protein by Short \textit{et al.} (1986).

\textit{Coding regions of the PMV genome}

The genomic RNA of PMV contains six ORFs larger than 200 nucleotides (Fig. 1a). ORF 1 encodes a protein of $M_r$, 176307 (176K). The second coding region partially overlaps ORF 1 by 14 nucleotides and encodes a protein of $M_r$, 26248 (26K). ORF 3 overlaps ORF 2 by 43 nucleotides and encodes a protein of $M_r$, 11949 (12K). The fourth overlapping ORF (ORF 4) produces a protein of $M_r$, 7224 (7K). ORF 5 codes for the PMV capsid protein ($M_r$, 23043). All the ORFs overlap others with the exception of ORF 5 which is separated from ORF 4 by a non-coding sequence of 32 nucleotides (Fig. 2).

There is also one other ORF (ORF 6) located completely within ORF 1 in the +1 reading frame and which encodes a protein of $M_r$, 14113 (14K).

Non-coding regions are the 5'-terminal 99 nucleotides, a 32 nucleotide intercistronic region between ORFs 4 and 5 and the 3'-terminal 121 nucleotides. The non-coding regions at the 5' terminus and upstream of the capsid protein gene are rich in A+U as are similar regions in other plant virus genomes (Goel et al., 1982).

The negative (−) sense of the genomic RNA contains two ORFs coding for proteins of $M_r$, 14791 (−ORF 1) and 14448 (−ORF 2) (Fig. 1b). The −ORF 1 product has been previously described by AbouHaidar (1988) as being highly hydrophobic as is the case for the −ORF 2 protein. Several other ORFs (7K to 9K products) are also seen on the negative sense RNA of PMV (Fig. 1b) as well as PVX and WC1MV (data not shown).

\textit{Similarity of PMV proteins to other viral proteins}

The 176K product encoded by ORF 1 of PMV is similar in two regions to analogous proteins found in the other sequenced potexvirus RNAs: NMV 186K, PVX 166K and WC1MV 147K. There are about 400 amino acids at the N termini and about 800 amino acids at the C termini. The differences in size between the various ORF 1 proteins correspond to the different lengths of the middle regions between these two regions (Zuidema \textit{et al.}, 1989). The similarities at the N termini range from 41 to 47% and from 51 to 56% at the C termini when compared with PMV regions (data not shown). Also, the GXGKGV(KST) sequence observed in the other potexvirus ORF 1 products can be seen towards the beginning of the C-terminal homologous region at amino acid positions 822 to 829 (nucleotide positions 2563 to 2586; Fig. 2).

The ORF 6 product (14K), encoded completely within ORF 1, showed no similarity to other potexvirus ORF products.

The GXGKST sequence is present towards the N terminus of the PMV 26K and other potexvirus 25K/26K proteins (Fig. 3). The overall similarities of these proteins compared to that of PMV range from 25 to 34%.
The PMV 12K protein shows about 43% similarity to the NMV 14K, PVX 12K and WC1MV 13K proteins. Most similarity is centred around the region shown in Fig. 4.

There was no significant similarity between the 15K – ORF 1/ – ORF 2 proteins of PMV (Fig. 1b) and the 20K protein of WC1MV encoded on the negative sense RNA (Forster et al., 1988) or the 12K protein from the PVX negative sense RNA (Huisman et al., 1988).

**Nucleotide sequence similarities found in PMV**

Fig. 5 shows the degree of similarity of nucleotide sequences found in front of the AUGs in PMV ORFs 2 to 5. The sequences upstream of the PVX 25K and capsid protein genes are shown for comparison.

When the first 500 nucleotides of the 5' terminus of PMV RNA were compared with those of PVX, 54.4% were identical. The similarity of PMV RNA with the WC1MV and NMV 5' termini was 50 and 48.8% respectively. When the 3' end of the positive sense PMV RNA was compared with the 3' end of the negative sense RNA, only 37% similarity was seen (data not shown).

**DISCUSSION**

The genomic organization of PMV presented in this paper is similar to that of the other potexviruses. There are four sequentially overlapping ORFs on the positive sense of the RNA, coding for 176K, 26K, 12K and 7K proteins and a separated fifth ORF that encodes the capsid protein. There is also an ORF within ORF 1. A similar small ORF was found within ORF 1 of WC1MV RNA (Forster et al., 1988) but there was no significant similarity between these two putative proteins. In contrast, the 10K ORF product coded within the NMV capsid protein gene contains regions homologous to a similar 7K product within the equivalent WC1MV gene (Zuidema et al., 1989).

The first ORF of PMV starts at position 100 and codes for a 176K protein which shares two major domains of similarity with other potexvirus ORF 1 products. Huisman et al. (1988) have shown the similarity between the PVX 166K protein and the tobacco mosaic virus (TMV) 126K/183K proteins which have been considered to be putative components of the TMV replicase (Ishikawa et al., 1986). In addition, the 176K product contains the consensus sequence proposed for RNA polymerases (Argos, 1988) as well as the proposed NTPase–helicase sequences common to other potexvirus ORF 1 products (Skryabin et al., 1988).

*In vitro* translation of genomic PMV RNA produced two large polypeptides of 155K and 73K
Nucleotide sequence of PMV RNA

Fig. 2. The complete nucleotide sequence of PMV genomic RNA. The variable length poly(A) tail at the 3' end is denoted by An. The amino acid sequences of the encoded proteins are located above the respective nucleotide sequence. Nucleotide changes from previously reported PMV sequences (Lok & et al., 1986) are indicated by U.

(Bendena et al., 1985) The 176K ORF 1 product presumably represents the 155K polypeptide and the 73K protein is probably a premature termination product of the in vitro translation of the 176K ORF since there are no other ORFs on the PMV genome coding for proteins greater than 30K in size. Moreover, the first 400 amino acids of ORF 1 show a high degree of similarity to the 25K, 12K or 8K proteins of PVX and the 3' end is denoted by An. The amino acid sequences of the encoded proteins are located above the respective nucleotide sequence. Nucleotide changes from previously reported PMV sequences (Lok & et al., 1986) are indicated by U.

ORF 6 is completely within ORF 1. The presence of an ORF within another has been noted after the AUG at position 100 is the correct initiation codon.
2.30 T. L. SIT, M. G. ABOUHAIDAR AND S. HOLY

Fig. 3. Alignment of amino acid sequences from the PMV 26K (ORF 2) product with 25K/26K products of other potexviruses. Positions of amino acid residues within the respective polypeptides are indicated. Gaps were introduced for best alignment. Homologous amino acid residues are boxed.

Fig. 4. Similarity between the PMV 12K (ORF 3) product and other potexvirus ORF 3 polypeptides. Positions of amino acid residues within the respective polypeptides are indicated except for PAMV. Homologous amino acid residues are boxed.

Fig. 5. Comparison of nucleotide sequences upstream from the AUG start codons of PMV and PVX ORFs. The 5'-terminal sequence of the PVX subgenomic RNA for capsid protein (Dolja et al., 1987) is compared with similar sequences found upstream of the PVX 25K ORF and PMV ORFs 2 to 5. The position from these sequences to their respective initiation codons is indicated. Gaps were introduced for best alignment. Identical nucleotides are boxed.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES


(Received 22 March 1989)