Partial Nucleotide Sequence of Potato Virus M RNA Shows Similarities to Potexviruses in Gene Arrangement and the Encoded Amino Acid Sequences

By V. V. RUPASOV,1† S. YU. MOROZOV,2 K. V. KANYUKA1 AND S. K. ZAVRIEV1*

1All-Union Research Institute of Agricultural Biotechnology, Moscow 127253 and 2Department of Virology, Moscow State University, Moscow 119899, U.S.S.R.

(Accepted 28 March 1989)

SUMMARY

The nucleotide sequence of the 3'-proximal 2630 nucleotides of potato virus M (PVM) genomic RNA was determined. The sequenced region contained five long open reading frames (ORFs). The ORF nearest to the 3'-terminal poly(A) tail corresponds to a polypeptide of Mr 10848. This ORF is preceded by one which encodes a protein of Mr 33906 (34K) which has an amino acid sequence that is very similar in its carboxy-terminal part to that of the coat proteins of some potexviruses. Three other ORFs encoding polypeptides of Mr 24615, 11893 and 6739 are present in the region 5' to the 34K protein gene. There is extensive similarity between these proteins and the corresponding proteins encoded by the conserved triple gene block in the RNA of potexviruses.

INTRODUCTION

Potato virus M (PVM), a carlavirus, has filamentous particles comprising multiple copies of an Mr 35700 coat protein and a single-stranded infective RNA of Mr 2400000 (Proll et al., 1981; Wetter & Milne, 1981; Tavantzis, 1984). The 3' end of PVM RNA is polyadenylated (Tavantzis, 1984). In vitro translation of PVM RNA yields a large polypeptide with an Mr of approx. 190K but little or no coat protein is synthesized during translation of PVM RNA (Szybiak & Legocki, 1981). One explanation for this is that PVM resembles tobacco mosaic virus (Hunter et al., 1976) in using a subgenomic mRNA that is not encapsidated for coat protein synthesis.

In this study we cloned and sequenced cDNA to the 3'-terminal region of the PVM genome in order to improve our understanding of the genome structure of carlaviruses. The cistrons for five putative PVM-encoded proteins have been localized and the predicted amino acid sequences have been compared with those of proteins of other plant viruses.

METHODS

PVM RNA extraction. PVM (Russian wild strain) was purified from infected tomato plants as described by Proll et al. (1981). RNA was isolated from purified virus preparations using the phenol-SDS method (Proll et al., 1981) with minor modifications.

cDNA cloning and sequencing. Double-stranded cDNA was synthesized using reverse transcriptase from avian myeloblastosis virus (Boehringer-Mannheim) and oligo(dT)12-18 (Pharmacia) as the primer for the first strand synthesis, and DNA polymerase I from Escherichia coli (Boehringer-Mannheim) and ribonuclease H (Promega Biotec) for the second strand synthesis (Gubler & Hoffman, 1983). The double-stranded cDNA was ligated to a Smal-cut plasmid pUC19 or phagemid pTZ19 (Pharmacia) which were used to transform the competent E. coli strain XL-1B (Stratagene Gene Cloning Systems) as described by Maniatis et al. (1982). Ampicillin-resistant transformants were screened for PVM-specific inserts by colony hybridization (Grunstein & Hogness, 1975) with32P-labelled, oligo(dT)-primed cDNA. An alternative probe for clones containing the 3'-terminal sequences was PVM RNA labelled at the 3' end by ligation to32Pcytidine-5',3'-bisphosphate (pCP) and T4 RNA ligase (Englund et al., 1980) and fragmented by mild alkaline hydrolysis (50 mM-Na2CO3, 65 °C for 0000-8804 © 1989 SGM
The sequence of 2630 nucleotides of PVM RNA was deduced from 19 partly overlapping cDNA clones. These clones were selected on the basis of restriction enzyme mapping and cross-hybridization (data not shown). The sequencing strategy for PVM cDNA inserts cloned into recombinant plasmids is shown in Fig. 1. Two cDNA inserts were found to contain the virus-coded 40 to 60 bp poly(dA) sequence and to extend for 290 and 682 nucleotides toward the 5' terminus of the PVM genome. Their 3'-terminal location was confirmed by the strong hybridization of these clones with fragmented [5'-32P]pCp-labelled virion RNA (data not shown).

The sequence of 2630 bases of the 3'-terminal part of PVM genomic RNA and the deduced polypeptide sequences are shown in Fig. 2. We have sequenced more than 97% of the cDNA in both directions, and more than 50% of the sequence was determined by examining two or three different clones covering each particular region. In total, six nucleotide variations (five transitions) were found between independent clones (see legend to Fig. 2).

Coding capacity of PVM RNA

Analysis of the sequenced genome region shows the presence of five open reading frames (ORFs) in the 3' part of PVM RNA (Fig. 1 and 2). The first ORF adjacent to the 5' end of the sequenced region initiates at an AUG codon at position 121 and finishes at nucleotide 783, which corresponds to a polypeptide of 221 amino acids with an M_r of 24165. A second ORF...
Fig. 2. The nucleotide sequence of 2630 nucleotides of the PV genome. Amino acid sequences of the coding regions are shown above the sequence. Variants in the nucleotide sequence are A → G (position 1132), G → A (1958), U → C (1727), C → U (2173), U → G (2227) and G → A (2475). Only the G → A transition at position 1958 causes a corresponding amino acid change (Ala → Thr).

starts from the AUG codon at position 786, terminates at position 1112 and encodes a 109 amino acid protein of $M$, 11893. The third ORF commences at an AUG at positions 1112 to 1114 and continues to position 1300, which corresponds to a polypeptide of 63 amino acids ($M$, 6739). In subsequent discussion the putative products of these ORFs are referred to as the 25K, 12K and 7K proteins.

Several small non-virion proteins with extensive hydrophobic domains encoded by animal RNA viruses have been experimentally identified as integral membrane proteins, e.g. M2 of
V. V. RUPASOV AND OTHERS

1 PVM 34K LAVIKKDAE-TRVCRLYAPAFTWN 205
2 PMV CP LAGIVKASGTS-SLRFPCRYPAPFTWN 113
3 PVX CP LAAAIYVC-TLRQFCMAYAPVTWN 137
4 PAMV CP LAGAGISRFQFCYSVAKIVWN 151
5 WClMV CP LAAIKVKH--TIRQFCTMYFPANIWN 111

1 HMITTHNAFFPADAWAMGQFYEDBFAXDFDY 236
2 LRTDKMAFANWQASGYKPSAKFAAFDFPG 143
3 WMTNNSFFWQAGQFKFEPHKPAFAAFDFPG 168
4 LMKHNFPAWAKIGKEDYKPAFAAFDFSA 181
5 IMLDTKTFPSASWSKLYKEKSKFAAFDFPG 142

1 YVKSNRK* 304
2 QFLPPPQ* 212
3 AVYTLPPP* 238
4 ALMAPPS* 249

Fig. 3. Amino acid sequence similarities between the predicted PVM 34K protein and the carboxy-terminal parts of the coat proteins (CP) of the potexviruses PMV, PVX, PAMV and WCIMV. Amino acids of potexvirus coat proteins in common with those in PVM protein are indicated by boxes. Numbers at the right side show positions of amino acids from the amino terminus. Gaps (−) have been introduced for maximum alignment.

influenza A virus (Lamb et al., 1985), SH protein of simian virus 5 (Hiebert et al., 1988) and NS1 protein of flaviviruses (Rice et al., 1985). Morozov et al. (1987) noted that several non-virion proteins of positive-sense RNA-containing plant viruses (potyviruses and carmoviruses) contain highly hydrophobic segments and might be membrane-bound. Small non-virion proteins of potexviruses, hordeiviruses and furoviruses are also hydrophobic and could interact directly with membranes (Morozov et al., 1987, 1989; Forster et al., 1988).

The hydrophobicity of the PVM proteins was analysed by the method described by Eisenberg et al. (1984). Using their criteria for hydrophobicity, protein–membrane interactions were predicted for the PVM 12K and 7K proteins (data not shown).

The fourth ORF extends from the AUG codon at positions 1325 to 1327 to the termination codon at positions 2237 to 2239. This ORF encodes a protein of 304 amino acids with an M, of 33906 (34K). Two lines of evidence suggest that the latter polypeptide is the PVM coat protein. The M, of the 34K protein is close to that determined by polyacrylamide gel electrophoresis for the PVM coat protein (35700; Tavantzis, 1984). The predicted PVM 34K protein sequence and the sequences of potexvirus coat proteins show several similarities most noticeably in the carboxy-terminal portions (Fig. 3). For example, in the stretch of 120 carboxy-terminal amino acids of the PVM 34K protein there are 55, 48, 45 and 39 amino acids identical to those in the respective coat protein sequences of potato virus X (PVX) (Morozov et al., 1983), papaya mosaic virus (PMV) (Short et al., 1986; AbouHaidar, 1988), potato aucuba mosaic virus (PAMV) (Bundin et al., 1986) and white clover mosaic virus (WCIMV) (Harbison et al., 1988). There are
**ORFs in potato virus M RNA**

Fig. 4. Conserved RNA sequences in the 5'-flanking regions of the PVM 34K protein gene and the 25K protein gene. Initiation codons are underlined. Asterisks above the sequences indicate positions of identical residues. Gaps (-) have been introduced for maximum alignment of identical nucleotides. Numbers at the left show positions of nucleotides according to the RNA sequence from Fig. 2.

Fig. 5. Amino acid sequence similarities between the predicted PVM 12K protein and the proteins potexviruses PVX and WCIMV (Morozov et al., 1987; Harbison et al., 1988). Amino acids identical those in the PVM protein are indicated by boxes.

Possible modes of gene expression in the 3'-terminal region of PVM RNA

As shown above, the localization of coat protein genes is similar in the PVM and potexvirus genomes. On the basis of this resemblance, it can be speculated that carlaviruses and potexviruses have common pathways for expression of their 3'-distal coat protein cistrons.
Fig. 6. Amino acid sequence similarities between the predicted 25K protein of PVM and parts of the corresponding proteins of PVX and WCIMV (Forster et al., 1988; Skryabin et al., 1988b), BSMV (Gustafson & Armour, 1986) and BNYVV (Bouzoubaa et al., 1986). Numbers at top right indicate positions of amino acids from the amino terminus.
ORFs in potato virus M RNA

1 PVM 7K MI YVVLVDS A FC IVL YL IS QG GS 24
2 PVX 8K M E N T LAI I I L VV T II AY S I P V R E 30
3 WCMIV 7K MD P T II I I G Y LL VP IV Y PAK I NT S 26
4 PAMV 8K M Y R Y L D CL - L V I M A V L A I A L W P N Y H 28

Virus-specific subgenomic RNAs were found in potexvirus-infected cells (Dolja et al., 1987; Guilford & Forster, 1986; Bendena et al., 1987; Mackie et al., 1988). The 5' end of the major PVX subgenomic RNA (GAAA box) is located five bases upstream from the initiation codon of the coat protein gene and the 5' end of the PVX 2-1 kb subgenomic RNA (GAAU box) was predicted to be located seven bases upstream from the AUG codon of the 25K protein gene on the basis of the nucleotide sequence alignment of the putative potexvirus mRNA promoters preceding the initiation codons of three potexvirus coat protein genes and two 25K to 26K protein genes (Forster et al., 1988; Skryabin et al., 1988 a, b). As in the case of potexviruses, the alignment of nucleotide sequences 5' to the initiation codons of the PVM 25K protein gene and the 34K protein gene revealed a region of nucleotide sequence homology (Fig. 4). It is significant that a GAAA box is located 12 bases upstream from the AUG codon of the PVM 34K protein gene and a GAAU box is located 8 bases upstream from the initiation codon of the 25K protein gene (Fig. 2 and 4). Possibly, by analogy with potexviruses, these tetranucleotide blocks are the 5' termini of the corresponding PVM subgenomic mRNAs.

Extended nucleotide sequences homologous to the upstream regions of the PVM 25K protein gene and the 34K protein gene were not found in the regions preceding the 12K, 7K and 11K protein genes. However, the junction sequences between the 25K and 12K protein genes (UGAUG), the 12K and 7K protein genes (AUGA) and the 34K and 11K protein genes (AUGA) have closely overlapping initiation and termination codons, and the expression of downstream ORFs could therefore be by the 'relay race' mechanism of translation (Gordon et al., 1988) or translational frameshift (for review see Valle & Morch, 1988).

Protein comparisons with other plant virus-encoded proteins

Comparison of the 25K, 12K and 7K polypeptide sequences with those of other viruses showed significant similarities with several non-virion proteins of potexviruses. In addition the 25K and 12K proteins show similarities, although less extensive, with the non-virion proteins of BSMV and BNYVV. It has been shown previously that the genomes of the potexviruses, BSMV, and BNYVV contain a conserved block of three genes (Morozov et al., 1987, 1989; Forster et al., 1988). The highest homology was seen in the 12K to 14K proteins for all three groups of these plant viruses (Morozov et al., 1987, 1989; Forster et al., 1988). In accordance with this observation, the PVM 12K protein shares 40% identity with the PVX 12K protein (Fig. 5).

The 25K proteins of PVM and PVX were found to have 25% identity (Fig. 6). These proteins are also similar to the 58K protein of BSMV and the 42K protein of BNYVV (Fig. 6). As shown
by Gorbalenya et al. (1988) all these polypeptides contain the consensus sequence of NTP-utilizing DNA helicases including the sequence Gly–X–Gly–Lys–Ser/Thr which is probably involved in the formation of a phosphate-binding pocket (see also Hodgman, 1988; Skryabin et al., 1988a, b).

The PVM 7K protein derived from the 3' end of the triple block of genes did not resemble the corresponding BSMV and BNYVV proteins but was as similar to the 7K to 8K potexvirus proteins as were the 25K to 26K proteins of PVM and potexviruses (see Fig. 6 and 7).

On the basis of these similarities, we suggest that the carlaviruses and potexviruses most probably form a ‘supergroup’ of related plant virus groups. The presence of an additional ORF at the 3' end of the PVM genome agrees with this suggestion. Insertions (or deletions) of a unique gene in RNA virus genomes have been found on comparison of different members of the tobravirus (Angenent et al., 1986) and coronavirus groups (de Groot et al., 1988).

REFERENCES


ORFs in potato virus M RNA


(Received 29 November 1988)