The nucleotide sequence of potato mop-top virus RNA 2: a novel type of genome organization for a furovirus

Karen P. Scott, Satoshi Kashiwazaki, Brian Reavy* and Bryan D. Harrison

1 Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA and 2 Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, U.K.

Particles of isolate T of potato mop-top furovirus (PMTV) contain three RNA species (6.5, 3.0 and 2.5 kb). Hybridization tests with cloned cDNA probes showed that none of these species was derived from another. RNA 2 (2962 nt), which was sequenced, has non-coding regions of 368 nt and 285 nt at the 5' end and 3' end, respectively. Near the 5' terminus, nucleotides 46 to 110 are able to form a stem-loop structure, the stem of which has 23 bp with only one mismatch and one unpaired nucleotide. From the 5' end, the four open reading frames encode proteins of 51K, 13K, 21K and 8K. The first three of these have sequence similarity to the triple-gene-block proteins of other viruses, particularly barley stripe mosaic hordeivirus. The 51K protein contains a putative NTP-binding motif and the 13K and 21K proteins each contain two hydrophobic regions separated by a hydrophilic region. The 8K protein is rich in cysteine. PMTV differs from other furoviruses in having a tripartite genome. Its RNA 2 differs in gene content from the RNA 2 of soil-borne wheat mosaic virus, which lacks a triple gene block, and from that of beet necrotic yellow vein virus, which has a coat protein gene and read-through domain to the 5' side of its triple gene block. The gene arrangement in PMTV is therefore novel for a furovirus.

Introduction

Potato mop-top virus (PMTV) infects potato crops in many countries, causing brown arcs in the flesh of tubers, sporadic stunting of shoots and yellow blotching of lower leaves, depending on the potato cultivar. The virus was first identified by Calvert & Harrison (1966) and is transmitted by the plasmodiophoromycete ‘fungus’ Spongospora subterranea (Jones & Harrison, 1969). It has fragile rod-shaped particles with two predominant lengths, 100 to 150 nm and 250 to 300 nm (Harrison & Jones, 1970); and it was placed in the furovirus group on the basis of its fungal transmission and particle morphology (Francki et al., 1991). However, PMTV is reported to have distant serological relationships with tobacco mosaic tobamovirus (Kassanis et al., 1972) as well as with soil-borne wheat mosaic furovirus (SBWMV), but not with beet necrotic yellow vein furovirus (BNYVV; Randles et al., 1976).

Previous work shows that viruses assigned to the furovirus group are of more than one type. Thus although BNYVV is considered to be a furovirus, its genome is organized differently to that of SBWMV. With both viruses, two genome segments (RNA 1 and RNA 2) are needed to infect plants and cause disease (Shirako & Brakke, 1984; Kuszala et al., 1986; Tamada et al., 1989). However, BNYVV RNA 2 contains five open reading frames (ORFs), which include a triple gene block that encodes proteins with roles in cell-to-cell movement of the virus (Bouzoubaa et al., 1986; Gilmer et al., 1992), whereas SBWMV RNA 2 has only two ORFs and its putative movement protein is encoded by a single ORF in RNA 1 (Shirako & Wilson, 1993). The RNA 2 of peanut clump furovirus (PCV) has the same gene content as that of BNYVV (Manohar et al., 1993). Studies on PMTV particles showed that those of strain T contain three RNA species, of 6.5 kb, 3.2 kb and 2.5 kb, respectively (Kallender et al., 1990; Scott et al., 1990). The work described here was initiated to determine the relationships among these three RNA species, to identify PMTV genes and to gain further information on the extent of diversity of genome organization among furo-
viruses. In this paper we report the nucleotide sequence of PMTV RNA 2 and show that its gene content differs from that of RNA 2 of either BNYVV or SBWMV.

Methods

Virus purification. Particles of the T isolate of PMTV (Harrison & Jones, 1970) were purified from leaves of mechanically inoculated Nicotiana debneyi and N. benthamiana plants, using the procedure described by Randles et al. (1976), as modified by Torrance et al. (1993).

cDNA synthesis and cloning. PMTV RNA was obtained by suspending virus particles in TE buffer (10 mM-Tris-HCl, pH 8, 1 mM-EDTA) containing 0.05 M-NaCl, 0.5 % SDS and 0.1 mg/ml proteinase K, and incubating the mixture at 37 °C for 2 h, followed by phenol extraction and precipitation with ethanol. The cDNA was synthesized as described by Gubler & Hoffman (1983) using a commercial kit (Pharmacia). Synthesis of cDNA was primed with random hexamer oligonucleotides and the product was methylated using EcoRI methylase, ligated to EcoRI linkers, digested with EcoRI and ligated into λ ZAPII arms with complementary EcoRI ends (Stratagene) to give library 1. A second cDNA synthesis was performed using oligo(dT) as a primer after polyadenylation of the RNA. The products were cloned into pBluescript vector (Stratagene) previously digested with Smal and dephosphorylated using alkaline phosphatase, to give library 2.

Library screening. To assign selected clones from library 1 to their respective RNA species, purified PMTV RNA or total RNA harvested from PMTV-infected leaves (Verwoerd et al., 1989) was separated on denaturing formaldehyde gels and the resulting Northern blots (Sambrook et al., 1989) were probed with radioactively labelled cloned cDNA (Feinberg & Vogelstein, 1984). Clones in library 2 were screened by dot blot hybridization using the PMTV RNA 2-specific insert from clone Tb2 (see Results) as a radioactive probe.

Nucleotide sequencing. Sequences were determined by dideoxynucleotide chain termination (Sanger et al., 1977). Specific smaller fragments of clone Tb2 were obtained by restriction enzyme digestion and isolated from low melting-point agarose gels for subcloning into M13 vectors with complementary cohesive ends and subsequent sequencing using single-stranded templates (Sambrook et al., 1989). Alternatively, subclones of clones pPMT2 and pPMT4, obtained by exonuclease III-generated nested deletions (Henikoff, 1984), were sequenced using double-stranded templates. The sequence of nucleotides at the 5’ end of the RNA was determined by extension from a PMTV-specific primer. Approximately 1 μg of PMTV RNA and 20 ng of the oligonucleotide S813 (5’ AATCTCCTTCTGTAAGAC 3’, complementary to nucleotides 75 to 92 of RNA 2) were heated at 80 °C for 3 min, and annealed at 40 °C for 30 min. The RNA–DNA hybrid was incubated with 1 U of RNase H (Amersham) at 37 °C for 20 min to digest regions of strong secondary structure (see below), followed by incubation with 30 ng/ml proteinase K and subsequent extraction with phenol. The digested RNA was then sequenced as described by Kashwazaki et al. (1990), using the oligonucleotide 5’ TTGAGAAAAAGACCTCAGGAA 3’, complementary to nucleotides 41 to 60 of RNA 2.

The sequence of nucleotides at the junction between those in the cDNA libraries and those determined by primer extension was obtained using PCR. Two specific primers S207 (5’ GCGATCGCTATTTCA-ACTCTACCTAGGC 3’, representing nucleotides 1 to 22 of RNA 2 and with a BamHI restriction enzyme site added to the 5’ end), and S208 (5’ CTTCCGTGAATACACT 3’, complementary to nucleotides 375 to 391) were used, with viral RNA as the template. The product was harvested from a low melting-point agarose gel, cleaved using the restriction enzymes EcoRI and BamHI, and the resulting fragment was then cloned into an appropriately digested pBluescript KS plasmid. Ten independent clones were sequenced in both directions.

The full sequence was assembled using the University of Wisconsin GCG gel assembly package, and analysed using the MAP, COMPARE, DOT PLOT, GAP and PILEUP programs (Devereux et al., 1984). Hydrophobicity plots were made using the PEPLOT program.

Results

Molecular cloning

Clones isolated from the λ ZAPII library (library 1) contained cDNA inserts of various lengths. The longest of these (Tb2: 2.2 kb) hybridized exclusively to PMTV RNA 2 (Fig. 1) indicating that, of the three RNA species reported by Kallender et al. (1990) and Scott et al. (1990), RNA 2 is not derived from RNA 1 and RNA 3 is not derived from RNA 2. The hybridization pattern of additional clones from the λ ZAPII library (Fig. 1) shows that RNA 3 is not derived from RNA 1 and thus that PMTV has three distinct RNA species. The weak reaction of the C3 probe with RNA 2 appears to suggest that RNA 1 and RNA 2 share a short nucleotide sequence but the probe proved to have only short non-terminal sequence identities to RNA 2.

An attempt to synthesize cDNA using oligo(dT) primers without first polyadenylating the RNA was unsuccessful, indicating that PMTV RNA does not possess a poly(A) sequence. RNA 2-specific clones in

Fig. 1. Northern blots of RNA from PMTV-infected leaf tissue. PMTV RNA 1 is detected with cDNA probe C3 (0.5 kb), RNA 2 with probe Tb2 (2.2 kb) and RNA 3 with probe C1 (0.5 kb).
Fig. 2. Nucleotide sequence of PMTV RNA 2. The deduced amino acid sequence is shown below the nucleotide sequence and termination codons are indicated by asterisks.
library 2 were isolated by hybridization to the radioactively labelled Tb2 insert. The sequence of two of these clones, pPMT2 (about 3 kb) and pPMT4 (about 2.5 kb), overlapped up to the added poly(A) sequence, suggesting that the 3' end of RNA 2 had been reached, although the possibility that there are one or more A residues at the 3' terminus is not excluded.

None of the clones in either library extended to the extreme 5' end of RNA 2, and the sequence for this region was therefore obtained by primer extension reactions. First attempts at primer extension sequencing using primers S813 (complementary to nucleotides 75 to 92) or primer $2E5 (complementary to nucleotides 41 to 60) failed to give an unambiguous sequence for the 5' end of the RNA although a strong stop signal corresponding to the 5' end of the RNA was observed. The sequence that was determined indicated the presence of a potentially strong hairpin-loop structure near the 5' end of the RNA that may have interfered with sequencing attempts. The sequence of the extreme 5' region was therefore obtained using primer S2E5, after the hairpin-loop had been disrupted by the annealing primer S813 to the RNA and subsequent digestion with RNase H. When this was done, a few bands representing nucleotides adjacent to primer S2E5 were weak, so these nucleotides were determined by analysis of additional cDNA clones obtained using PCR with primers S208 and S207.

**Nucleotide sequence.** The complete sequence of PMTV RNA 2 is shown in Fig. 2. Of the 2962 nucleotides, 2927 were within cDNA clones and were sequenced in both directions. Non-coding regions of 368 nt and 285 nt are present at the 5' and 3' ends, respectively. The sequence contains four ORFs encoding proteins of 51K, 13K, 21K and 8K (Fig. 3).

The deduced amino acid sequences of these proteins were compared with all the sequences in the EMBL and SWISSPROT databanks. The first three proteins resemble the triple-gene-block proteins found in members of several genera of plant viruses. These proteins were more similar to those of barley stripe mosaic virus (BSMV), a hordeivirus, than to those of BNYVV, as indicated by DIAGON plots (not shown).
The amino acid sequence of each triple-gene-block protein was compared with the sequences of the corresponding triple-gene-block proteins of a range of other viruses including carlaviruses, furoviruses, hordeivirus, carlaviruses, and furoviruses. The first ORF in PMTV RNA 2 extends from nucleotides 369 to 1760 and encodes a protein of M, 51175. Although the N-terminal part of this protein is not well conserved with that of comparable proteins of other viruses with triple gene blocks, considerable sequence similarity occurs in the middle of the sequence and towards the C terminus. The conserved dNTP binding site motif (Walker et al., 1982; G/A X X X X G K S/T; residues 215 to 222 in the PMTV protein) marks the first of several blocks of amino acid sequence which are strongly conserved among these different plant viruses, as shown in Fig. 4. Residues conserved in three or more sequences are recorded in the consensus sequence.

13K protein

The protein product (M, 13108) of the second ORF in the PMTV triple gene block (nucleotides 1747 to 2106) again had most similarity to the comparable BSMV protein although considerable sequence conservation was also found with the equivalent triple-gene-block protein of the other viruses analysed (Fig. 5). This protein contains no motifs with well-defined functions.

21K protein

The amino acid sequence of the third protein (M, 20972) encoded by PMTV RNA 2 (nucleotides 1961 to 2533) was compared in the same way, but the sequence of this protein seems the least well-conserved among the viruses possessing a triple gene block. The only substantial similarities found were between the equivalent PMTV and BSMV proteins (29% identity) and these are illustrated in a GAP alignment (Fig. 6). The corresponding protein of PCV has 28% sequence identity to the PMTV protein but over a much smaller distance because its size is only 9K (Manohar et al., 1993).

8K protein

The protein (M, 7969) encoded by the fourth ORF in PMTV RNA 2 (nucleotides 2461 to 2667) did not have significant sequence identities to any other sequence in the database. It does, however, contain an unusually large proportion of cysteine residues (8 out of 69) that may have some bearing on its function. Proteins involved in nucleic acid binding are often associated with zinc atoms, which are tetrahedrally liganded to cysteine residues (Sehnke et al., 1989). However, other plant virus proteins with the potential for Zn-binding have two amino acids separating each pair of cysteine residues that form the tetrahedral liganding site [e.g. potato virus M]...
Fig. 7. Predicted stem-loop structure near the 5' end of PMTV RNA 2 showing base pairing. The nucleotides are numbered according to their positions in the full RNA 2 sequence. The positions of primers S2E5 and S813 are indicated.

Rupasov et al. (1989); and BNYVV, Morozov et al. (1989), whereas in PMTV only one amino acid separates the pair of cysteines occurring at two points in the sequence.

5'-terminal secondary structure of RNA

An untranslated nucleotide sequence near the 5' terminus of PMTV RNA 2 has the potential to undergo extensive base-pairing to form a complex secondary structure [nucleotides 45 to 345 have a free energy of $-339.8$ kJ/mol ($-80.9$ kcal/mol)]. The principal stem-loop (nucleotides 46 to 110) in this region has a free energy of $-115.9$ kJ/mol ($-27.6$ kcal/mol). The stem contains 23 bp with only one mismatch and one unpaired base (Fig. 7). The loop contains 16 bases, which are mostly unpaired. The existence and stability of this feature probably explain the failure of cDNA clones to reach the 5' end of the RNA and the failure of primer extension and RNA sequencing to give satisfactory results until the secondary structure of the RNA was destroyed.

Discussion

The results of the nucleic acid hybridization tests with cloned cDNA probes show that the three RNA species (6.5, 3.0 and 2.5 kb) found in PMTV particles have distinct nucleotide sequences and that none of the species is derived from another. The nucleotide sequence of the second largest species, RNA 2, shows it to be slightly smaller than estimated previously by gel electrophoresis. It differs from RNA 2 of two other furoviruses, SBWMV and BNYVV, in containing four ORFs, none of which encodes the virus coat protein. The gene arrangement in PMTV RNA 2 resembles that of the 3'-terminal half of BNYVV RNA 2 but PMTV RNA, unlike BNYVV RNA, seems not to be polyadenylated. PMTV RNA 2 differs greatly from SBWMV RNA 2, which does not contain the triple gene block found in PMTV and BNYVV; indeed the putative movement protein of SBWMV is encoded by RNA 1 (Shirako & Wilson, 1993). The only feature shared by RNA 2 of all three viruses is their 3' ORF, which encodes a cysteine-rich protein of unknown function. These comparisons indicate that the genome organization of PMTV differs from those of SBWMV and BNYVV, which differ from one another.

The three most 5'-terminal ORFs in PMTV RNA 2 encode proteins with characteristics typical of the triple-gene-block proteins which are involved in cell-to-cell movement of BNYVV and members of several other genera of plant viruses that have elongated particles. However, in amino acid sequence, each of the PMTV proteins is more similar to the comparable protein of BSMV, a hordeivirus, than to that of BNYVV, a furovirus. The sequence identities of the PMTV and BSMV proteins are 32% (51K/58K), 53% (13K/14K) and 29% (21K/17K). The comparable values for comparisons of PMTV and BNYVV proteins are 27%, 37% and 17%, respectively. The greater similarities of the PMTV and BSMV proteins than of the PMTV and BNYVV proteins are further evidence of the diversity of viruses included in the furovirus group.

All three of the ORFs that make up the triple gene block must be intact for white clover mosaic potexvirus (Beck et al., 1991) or BNYVV (Gilmer et al., 1992) to move from cell to cell, probably because their products interact with one another, and with viral RNA and coat protein (Goulden et al., 1993) to mediate movement. The same may well be true for PMTV. The hydrophobicity plot for each triple-gene-block protein of PMTV (Fig. 8)
shows that the 13K and 21K proteins each have two hydrophobic regions separated by a hydrophilic one. The hydrophobic regions are of a suitable length for spanning membranes (Eisenberg et al., 1984). The presence of these domains suggests a possible mode of action. The hydrophobic domain on the 13K (or equivalent) protein may bind to membranes and the largest protein presumably binds to the viral nucleic acid. In BNYVV both these proteins were found exclusively in the membrane fraction (Niesbach-Klosgen et al., 1990), although the 42K protein has no membrane binding regions. This suggests that the 42K protein may attach to the membrane via the 13K protein, and that the third protein, which was not detected in BNYVV-infected plants and may be present at lower concentrations, may also be involved in the association with, or release from, membranes of the 42K protein and viral RNA. Interactions of this kind would explain the strong tendency for co-occurrence of the triple block genes in viral genomes and the selective advantage thereby conferred. In the two viruses known to contain only two of these genes, infection or replication is less effective than that of viruses with complete triple gene blocks. For instance, lily X potexvirus lacks the third of the genes and reaches only low concentrations in plant tissue (Memelink et al., 1990). Strawberry mild yellow edge-associated potexvirus appears to lack an initiation codon for the first triple block ORF and is confined to phloem tissue (Jelkmann et al., 1990, 1992). This suggests that the protein normally encoded by this gene is not needed for cell-to-cell movement in phloem tissue, and evidence that different processes are involved in virus movement between phloem cells and between mesophyll cells.

Many plant viral RNA species have stable secondary structures at their 3' ends, which resemble tRNA structures. However, among plant virus RNA species, only PMTV RNA 2 appears to have such a prominent stem–loop structure near its 5' end. By analogy with the 5' untranslated region of BNYVV RNA 3, which has a complex secondary structure that is important for replication (Gilmer et al., 1993), the comparable region of PMTV RNA 2 may play a similar role.

Karen Scott held a Postgraduate Research Studentship of the Scottish Office Agriculture and Fisheries Department (SOAFD), Satoshi Kashiwazaki held a Japan Science and Technology Agency Research Training Fellowship, and the research project was supported by SOAFD. This work benefitted from use of the SEQNET facility.

References


Received 31 May 1994; Accepted 18 August 1994