The Complete Nucleotide Sequence of Plum Pox Virus RNA

By E. MAISS,1* U. TIMPE,1 A. BRISSKE,1 W. JELKMANN,1 R. CASPER,1 G. HIMMLER,2 D. MATTANOVICH2 and H. W. D. KATINGER2

1Biologische Bundesanstalt für Land- und Forstwirtschaft, Institut für Viruskrankheiten der Pflanzen, Messeweg 11/12, D-3300 Braunschweig, F.R.G and
2Universität für Bodenkultur, Institut für Angewandte Mikrobiologie, Peter Jordan-Strasse 82, A-1190 Wien, Austria

(Accepted 18 November 1988)

SUMMARY

The complete nucleotide sequence of the RNA of an aphid non-transmissible plum pox virus (PPV-NAT) isolate has been determined from five overlapping cDNA clones. cDNA prepared by primer extension was used to determine the 5' terminus. The assembled RNA is 9741 nucleotides in length, excluding a 3' terminal poly(A) sequence. One large open reading frame starts at nucleotide positions 36 to 38 and is terminated with an UAG codon at positions 9522 to 9524. The putative start codon is located at positions 147 to 149. The encoded polyprotein has a predicted Mr of 353.8K. Comparison of cistrons from tobacco vein mottling virus and tobacco etch virus with those predicted for PPV-NAT indicated a similar genome organization. A highly conserved sequence of 12 nucleotides was found in the 5' non-coding region of these three potyviruses. The potential polyadenylation signal from yeast (UAUGU) was found in the 3' non-coding region of PPV-NAT and several other members of the potyvirus group.

INTRODUCTION

Plum pox virus (PPV), a member of the potyvirus group, causes heavy yield losses in plum, peach and apricot all over Europe. Potyviruses contain a monopartite, plus-sense, single-stranded RNA genome of approx. 10 kb (Hollings & Brunt, 1981). Several investigations have demonstrated that post-translational proteolytic processing of polyproteins is the mechanism of potyviral genome expression (Dougherty et al., 1985; Allison et al., 1986; Carrington & Dougherty, 1987; Hellmann et al., 1988; Chang et al., 1988). The positions of functional virus proteins in the polyprotein have been proposed for tobacco vein mottling virus (TVMV) (Domier et al., 1986), helper component protein (HC), cylindrical inclusion protein (CI), two nuclear inclusion proteins (NIa and NI0 and the coat protein (CP). A sixth peptide, the genome-linked protein (VPg) is also encoded by the viral RNA (Shahabuddin et al., 1988). In addition it is thought that another virus-encoded protein could mediate the transport of the virion or the RNA in the plant (Domier et al., 1987).

We have determined the nucleotide sequence of the RNA of an aphid non-transmissible isolate of PPV (PPV-NAT) to analyse its genetic organization and mode of genome expression. This should enable us to choose suitable regions to establish molecular cross-protection against PPV in transformed plants.

METHODS

Virus and viral RNA preparation. PPV-NAT was propagated in Nicotiana clevelandii. The virus was purified according to Lisa et al. (1981), and the RNA was extracted as described previously (Maiss et al., 1988).

cDNA synthesis and cloning. cDNA of the RNA was synthesized, dC-tailed and cloned into dG-tailed pBR322 as described previously (Maiss et al., 1988). Five pPPV-NAT clones (274, 31, 47, 232 and 65) were chosen for
sequencing. Clones pPPV-NAT274 and pPPV-NAT65 were found to extend furthest in the 5' and 3' direction of the RNA, respectively. Large restriction fragments of the chosen clones were subcloned into the polylinker region of pT7T3 18U and pT7T3 19U (Pharmacia-LKB) in order to sequence both strands of the cDNA fragments. The plasmids were unidirectionally deleted according to Henikoff (1984).

**Nucleotide sequencing.** Sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977) of both cDNA strands with the universal primer 5' d(GTAAAACGACGGCCAGT) 3' or the specific T3 promoter primer 5' d(ATTAACCCTCACTAAAG) 3' was done after preparation of ssDNA templates with the helper phage M13 K07 (Pharmacia-LKB).

**Direct RNA sequencing and primer extension.** RNA sequence analysis was performed by cDNA synthesis with reverse transcriptase (Boehringer) and a synthetic oligodeoxynucleotide primer 5' d(GATAGCTTGAGATTGAT) 3' by a modified method of Kück et al. (1987). Approximately 2 µg of PPV-NAT RNA was coprecipitated with 5' end-labelled oligonucleotide primer and redissolved in 10 mM-Tris-HCl pH 7.5, 1 mM-EDTA, 0.5% SDS. After incubation for 30 min at 35 °C, the solution was kept at 15 °C for 60 min. After ethanol precipitation the pellet was dissolved in 8 mM-Tris-HCl pH 8.5, 8 mM-MgCl₂, 32 mM-NaCl, 5 mM-dithiothreitol, 80 µM-dATP, dCTP, dGTP and -dTTP, 128 µM of the respective dideoxynucleoside triphosphate, 5 units RNasin and 15 units reverse transcriptase. The reaction was carried out in a final volume of 5 µl at 45 °C for 30 min, followed by addition of 1 µl deoxynucleotide mixture (each 0.5 mM) and incubation at 45 °C for 20 min.

To determine the first nucleotide of the RNA, the same end-labelled primer was used to generate a primer-extended cDNA. This cDNA was fractionated on an 8% sequencing gel, eluted and sequenced as described by Maxam & Gilbert (1980).

**Sequence analysis.** Sequence data were assembled and analysed using the computer program GENMON (distributor: GBF, Braunschweig) and the program of Staden (1984).

**RESULTS AND DISCUSSION**

**Sequencing of PPV-NAT RNA and cDNA clones and determination of 5'- and 3'-terminal sequences**

The PPV-NAT cDNA inserts from an overlapping set of five recombinant DNA molecules were chosen for nucleotide sequence analysis. These cDNA inserts represent approximately 99.9% of the PPV-NAT genome. The overlapping regions ranged from 150 to 500 nucleotides. The relationship of the clones to the PPV-NAT genome and the sequencing strategy are shown in Fig. 1.

The sequence of the 5'-terminal region of PPV-NAT RNA that is not contained within pPPV-NAT274 (Maiss et al., 1988) was determined by direct RNA sequencing using reverse transcriptase and a synthetic oligodeoxynucleotide primer. This provided evidence for only 12 additional nucleotides beyond the 5' terminus of pPPV-NAT274, suggesting that the 5' terminus of the RNA had been reached. Verification of the first nucleotide was done by chemical sequencing of a primer-extended cDNA. Nevertheless, VPg covalently linked to the 5' terminus of the RNA of potyviruses or portions of this protein could interfere with reverse transcriptase and hinder the determination of the exact 5' sequence.

Two sequenced inserts at the 3' terminus, in pPPV-NAT309 (Maiss et al., 1988) and pPPV-NAT65, were found to include terminal poly(A) segments of 14 or approx. 40 residues, respectively. Lain et al. (1988) showed that PPV RNA contains a 3' poly(A) tail between 15 and 500 residues long. Therefore it is evident that the 3' terminus is covered by both clones.

**Primary structure of PPV-NAT RNA**

The complete nucleotide sequence of PPV-NAT RNA contains 9741 nucleotides excluding the 3' terminal poly(A) sequence (Fig. 2). PPV-NAT RNA is longer than those of the two other fully sequenced potyviruses, TMV with 9471 nucleotides (Domier et al., 1986) and tobacco etch virus (TEV) with 9495 nucleotides (Allison et al., 1986). Base composition of the PPV-NAT RNA revealed a high adenine content (31.3%) followed by uracil (25.2%), guanine (23.0%) and cytosine (20.5%). This composition is similar to those of TMV and TEV.
Coding regions

Computer translation of the RNA and its complement in all six possible reading frames revealed a single large open reading frame (ORF). A termination codon (UAG) occurs at positions 9522 to 9524 (Fig. 3). Two putative initiation codons were found in the ORF at nucleotides 36 to 38 and 147 to 149. But only the second AUG codon with the preceding sequence AGUCAAG is in reasonable agreement with the consensus ribosomal recognition sequence AGCCACC noted by Staden (1984). Additionally, this AUG is surrounded by A in position −3 and U in position +4. As demonstrated by Kozak (1981) functional initiation sites have either an A in position −3 and a G, A or a pyrimidine in position +4 or a G in position −3 and a purine in position +4. Therefore it is concluded that translation starts at position 147. Eight other potential initiation sequences with start codons occur downstream in the sequence. Three of these, having AUG codons at positions 3618 to 3620, 6054 to 6056 and 6666 to 6668, are favourable for encoding proteins of 1969, 1156 and 952 amino acid residues, respectively.

The product of translation starting at nucleotide position 147 contains 3125 amino acid residues (M_r 353.8K). Mature potyviral proteins are processed from a precursor polyprotein by a virus-encoded protease (Hellmann et al., 1988; Carrington & Dougherty, 1987). It has been proposed that cowpea mosaic virus (CPMV) polyproteins are cleaved at selected Gln–Gly, Gln–Met and Gln–Ser sites (van Wezenbeek et al., 1983). Only Gln–Ser and Gln–Gly sites were used to cleave the polyprotein of TEV by the viral protease (Dougherty et al., 1988). In addition to these sites a Gln–Ala site is used for protease cleavage in the TVMV polyprotein (Domier et al., 1986). The positions of all Gln–Ser, Gln–Gly and Gln–Ala sites in the PPV-NAT genome are shown in Fig. 4. Using the proposed locations of cleavage sites and the size of the predicted mature proteins of TVMV and TEV, a suggested cistron map is shown (Table 1; Fig. 4).

Comparison of amino acid sequences from the 5′ termini of the coding regions of TVMV, TEV and PPV-NAT revealed only a few homologies. A part of the putative 34.3K protein of PPV-NAT was found to be slightly homologous to the 30K protein of tobacco mosaic virus (TMV; Meshi et al., 1982). Deom et al. (1987) demonstrated that expression of the 30K protein of TMV in transgenic tobacco plants could play a role in cell-to-cell spread during virus infection. This suggests that the 34.3K protein could have a similar function.

No potential protease cleavage site in the predicted translation product of PPV-NAT was found between the HC and the 42K protein as shown for TVMV (Domier et al., 1986). Using the
next possible cleavage site the HC of PPV-NAT has a length of 613 amino acids and an M, of 69-4K, and is larger than the putative HCs of the other potyviruses. In this case the PPV-NAT protein corresponding to the TVMV 42K protein is only 29K. Alignment of the 484 amino acid residues of TVMV HC with the 484 amino acids of PPV-NAT, starting with the first potential HC amino acid residue (numbers 302 to 785) displayed 51-4% identity. Addition of the remaining 129 amino acids to the 29K protein and comparison of the resulting product with the 42K protein of TVMV revealed 24.5% homology. Introducing a shift of three amino acid residues in TVMV, to get an optimal alignment, revealed more than 30% identity. Therefore it may be possible that a mutation at the end of the HC sequence has changed the amino acid
Fig. 4. Proposed protease cleavage sites in the PPV-NAT polyprotein. The location of potential cleavage sites at the dipeptides Gln-Ala, Gln-Gly and Gln-Ser in the polyprotein is shown by bars. The proposed cleavage sites for the mature proteins are indicated by solid triangles. The cistron map of PPV-NAT RNA is at the bottom. The hypothetical mutated cleavage site (see text) between HC and a proposed 42K protein is indicated by an open triangle.

Table 1. Positions of the proposed mature proteins in the polyprotein of PPV-NAT

<table>
<thead>
<tr>
<th>PPV-NAT protein</th>
<th>Amino acid position</th>
<th>Cleavage site</th>
</tr>
</thead>
<tbody>
<tr>
<td>34K</td>
<td>1-301</td>
<td>Gln-Ser</td>
</tr>
<tr>
<td>HC*</td>
<td>302-914</td>
<td>Gln-Ala</td>
</tr>
<tr>
<td>29K*</td>
<td>915-1168</td>
<td>Gln-Ser</td>
</tr>
<tr>
<td>CI</td>
<td>1169-1856</td>
<td>Gln-Gly</td>
</tr>
<tr>
<td>NI_a</td>
<td>1857-2292</td>
<td>Gln-Ser</td>
</tr>
<tr>
<td>NI_b</td>
<td>2293-2810</td>
<td>Gln-Ala</td>
</tr>
<tr>
<td>CP</td>
<td>2811-3125</td>
<td></td>
</tr>
</tbody>
</table>

* The HC of PPV-NAT seems not to be cleaved properly from the polyprotein. Therefore the following protein, in TVMV known as 42K, appears to be only 29K in PPV-NAT (see text).
three potyvirus proteins was the similarity of the sequence GXXGXGKS, where X represents any amino acid residue.

The amino acid sequences of the predicted NIₐ region of TVMV and PPV-NAT revealed 39.6% identity. The comparison of TEV and PPV-NAT showed 45.6% homology. This protein is a protease, involved in the cleavage of the polyprotein to mature viral proteins (Hellmann et al., 1988). The cluster of amino acids with Cys and His residues representing the active centre of the enzyme is highly conserved in PPV-NAT, TVMV and TEV (Fig. 5). The amino acids of the putative N terminus of the NIₐ protein of TVMV seem also to be present in the VPg, the protein covalently linked to the 5' terminus of the potyviral genome (Shahabuddin et al., 1988). Scanning of the PPV-NAT polyprotein for homology with the first 21 amino acids of the TVMV VPg showed the most homology (46.7%) with the N-terminal region of the putative PPV-NAT NIₐ protein (Fig. 6). Homologies between the TVMV VPg and the NIₐ of TEV were not found.

The predicted amino acid sequence of the NIₐ region of PPV-NAT revealed the most homologies of all the putative mature potyviral proteins (61.3% to TVMV and 55.0% to TEV). The conserved amino acid sequence predicted for RNA-dependent RNA polymerases (Kamer & Argos, 1984; Hamilton et al., 1987) was found in this region (Fig. 7). It is likely that the NIₐ of PPV-NAT has a similar function.
The amino acid sequence of CP of PPV has been predicted for four isolates (Lain et al., 1988; Ravelonandro et al., 1988; Maiss et al., 1989). Comparison of these proteins is shown in Fig. 8. The CP of the PPV-NAT isolate has a deletion of 15 amino acids near the N terminus. All PPV CPs are larger than those of other potyviruses like sugarcane mosaic virus (SCMV; Gough et al., 1987), pepper mottle virus (PeMV; Dougherty et al., 1985), TVMV (Domier et al., 1986), TEV (Allison et al., 1986) and potato virus Y (Shukla et al., 1986). The larger CP is reflected in the extended nucleotide sequence of PPV-NAT compared to TVMV and TEV.

Non-coding regions

The A/U-rich 5' leader sequence of the PPV-NAT genome is 146 bases long. Comparison with the leader sequences of TVMV and TEV revealed an identical region of 12 nucleotides (Fig. 9), starting at position 14 in PPV-NAT and TEV and at position 13 in TVMV.

The 3' non-coding region of PPV-NAT is 220 nucleotides in length without the poly(A) tail. No identical sequences were found in this region of the potyviruses PPV-NAT, TVMV, TEV, PeMV and SCMV. A potential eukaryotic polyadenylation signal was found in the TVMV RNA only. The sequence UAUGU, located approx. 70 to 90 nucleotides upstream of the
poly(A) tail in TVMV, TEV, PeMV and PPV-NAT may also represent a polyadenylation signal as demonstrated for yeasts, where this region is located 30 to 40 nucleotides in front of the poly(A) tail (Zaret & Sherman, 1982). This sequence was also found in the 3' non-coding region of tomato black ring virus RNA-1 (Greif et al., 1988) and RNA-2 (Meyer et al., 1986), and CPMV (Van Wezenbeek et al., 1983; Lomonossoff & Shanks, 1983); these viruses also contain plus-sense, monocistronic polyadenylated RNAs.

From virus structure similarities and the extent of protein sequence homology it is evident that functional similarities exist between TVMV, TEV and PPV-NAT. However, further analysis of the putative transport proteins, which are only slightly homologous and may be involved in specific virus-host interactions, and the 42K protein is necessary to understand the functions of these viral products.

Isolation of the PPV coat protein gene and expression of the corresponding protein in transgenic plants may be a tool for establishing molecular cross-protection against the virus.

REFERENCES


(Received 5 September 1988)