The nucleotide sequence of potato virus A genomic RNA and its sequence similarities with other potyviruses

Ülo Puurand, Kristiina Mäkinen, Lars Paulin and Mart Saarma

1 Institute of Biotechnology, University of Helsinki, Karvaamokuja 3A, P.O. Box 45, FIN-00014, Helsinki, Finland
2 Institute of Chemical Physics and Biophysics, Estonian Academy of Sciences, Akadeemia 23, EE0026, Tallinn, Estonia

The complete nucleotide sequence of potato virus A (PVA) was obtained from six independent cDNA clones. The RNA genome of PVA is 9565 nucleotides long and contains one open reading frame (ORF) of 9177 bases encoding a large polyprotein of 3059 amino acids with a calculated M_r of 340K. Seven potential proteinase NIa, one HC-pro and one P1 proteinase recognition sites were found in PVA polyprotein by searching for cleavage site consensus sequences amongst the potyvirus group. The non-coding region preceding the ORF is 161 nucleotides long. The termination codon is followed by a 227-nucleotide sequence. Overall nucleotide sequence identity compared with several completely sequenced potyvirus genomes is between 53 and 58 %, with overall amino acid sequence identity between 65 and 71 %. When the putative amino acid sequences of individual proteins of PVA were compared with the corresponding proteins of other potyviruses, P1 and P3 appeared the least conserved (34 to 53 %) whereas the other proteins were in most cases from 63 to 80 % identical to each other.

The complete nucleotide sequence of seven different potyviruses has been documented: tobacco etch virus (TEV) (Allison et al., 1986), tobacco vein mottling virus (TVMV) (Domier et al., 1986), plum pox virus (PPV) (Maiss et al., 1989; Lain et al., 1989), the necrotic strain of potato virus Y (PVY") (Robaglia et al., 1989), pepper mottle virus (PepMoV) (Vance et al., 1992), turnip mosaic potyvirus (TuMV) (Nicolas & Laliberté, 1992) and pea seed-borne mosaic virus (Johansen et al., 1991). The sequence analysis of potyviruses and in vitro translation studies of potyvirus genomic RNAs have revealed a single open reading frame (ORF) encoding a large polyprotein. This polyprotein is proteolytically processed into functionally active viral proteins by virus-encoded proteases (Hellmann et al., 1983; Dougherty & Carrington, 1988). Sequence similarities between the potyviruses analysed indicate that their polyproteins are cleaved at similar sites, yielding eight virus-encoded proteins in each. The potyviruses also contain conserved protein motifs, which may have functional importance for all members of the potyvirus group. In this paper we report the complete nucleotide sequence of the PVA RNA genome.

PVA was purified from Nicotiana occidentalis as
Two regions, box 'a' and box 'b' (Fig. 1), that are described by Hammond & Lawson (1988). RNA was isolated as described by Puurand et al. (1992). Two types of cDNAs were synthesized, one with oligo(dT)15 primers and the other with random primers. Both cDNAs were cloned in the λ gt11 vector. At least 1000 clones were immunoscreened using commercial alkaline phosphatase-conjugated rabbit polyclonal antibodies to PVA (Boehringer Mannheim). Clone A-ASL dT0.1 was selected for further sequence analysis. The subsequent clones were found by hybridization with the 5' end DNA fragment of the previous clone. The four independent clones that were chosen were subcloned into plasmid vectors and nested unidirectional deletions with exonuclease III (Promega) were generated from both strands. The sequences were determined from both strands manually using the dideoxynucleotide chain termination method (United States Biochemical) or automatically (A.L.F. DNA Sequencer, Pharmacia), or by both methods. The 5' terminal sequence of PVA was determined using PCR. The first strand of cDNA was synthesized using oligodeoxynucleotide TTAACCTCC-AGGATCTTGA-A. A poly(A) tail was added to the ssDNA with terminal transferase. The DNA product was amplified by PCR using the same primer that was used for the synthesis of the first strand of cDNA and (dT)25dN. The product was directly sequenced and after cloning six plasmid clones were sequenced. The 5' terminal nucleotide was detected by primer extension. Clone A-ASL dT0.1 did not contain the whole 3' region. A 0.5-35 kb fragment from the 3' end of A-ASL dT0.1 was used to screen the oligo(dT) primed λ library. Two positive clones were subcloned and sequenced. These clones extended to the poly(A) tail.

The similarity studies of the PVA sequence were performed using the Genetic Computer Group’s programs (Devereux et al., 1984). Determinations of similarity were made between the sequence of PVA and those of TVMV, PVV, PVY, TEV, PepMoV and TuMV. Nucleotide and amino acid sequence alignments were made using the GAP program.

The complete nucleotide sequence of the PVA RNA is shown in Fig. 1. The genome consists of 9565 nucleotides followed by a poly(A) tail. Computer analysis identified a single long ORF beginning at position 162 and ending with a UAA termination codon at position 9336 of the viral positive-sense strand, followed by a 227-nucleotide 3' untranslated region (UTR) (Fig. 1). The overall base composition of PVA RNA is 32.7% A, 20.1% C, 22.5% G and 24.7% U, and is very similar to that of other potyviruses.

The 5' leader sequence of PVA RNA is 161 nucleotides long and has a base composition significantly different to that of the total PVA sequence (49.7% A, 22.4% C, 21.1% U and 6.8% G). This very low G content seems to be a common feature of plant viral 5' leader sequences (Gallie et al., 1987). Both 5' and 3' UTRs of PVA are 36 to 58% similar to the analogous regions of other potyviruses that have been sequenced. In our previous report, we noted that we had not obtained the poly(A) tract of PVA RNA, and therefore the sequence of the 3' UTR region was incomplete (Puurand et al., 1992). Here we have obtained several cDNA clones containing an additional sequence of 22 nucleotides followed by the dT sequence, which we believe represents the poly(A) tail. Two regions, box 'a' and box 'b' (Fig. 1), that are conserved in the 5' region of all potyviruses can also be...
Short communication

Metal-binding motif

HC-pro putative active site

NTP-binding motif

VPg

Polymerase motif

<table>
<thead>
<tr>
<th>P1 33</th>
<th>HC-pro 52</th>
<th>P3 40</th>
<th>6K1</th>
<th>CI 64</th>
<th>6K2</th>
<th>Nla 55</th>
<th>Nlb 59</th>
<th>CP 30</th>
</tr>
</thead>
</table>

Fig. 2. Proposed map of the PVA polyprotein showing the putative cleavage sites and the resulting individual viral polypeptides with their calculated Mrs. Specific amino acid sequence motifs along the proposed polyprotein map are indicated. The putative P1/HC-pro, HC-pro/P3 and seven (A to F and V) Nla protease cleavage sites are indicated by open arrows.

found in the PVA 5' region. However, in other potyviruses box 'a' is located much nearer to the 5' end of the sequence than it is in the PVA sequence. Unlike other potyviruses box 'a' is also downstream from box 'b' in the PVA sequence.

The long ORF of PVA, starting at nucleotide 162, encodes a predicted polypeptide of 3059 amino acid residues (about 340K) (Fig. 1). Mature potyviral proteins are expressed by proteolytic processing of the polyprotein by the virus-encoded proteases. Putative protease cleavage sites in the PVA polyprotein were identified by comparison with consensus cleavage sequence motifs. There are seven putative Nla protease cleavage sites in the PVA polyprotein (Fig. 1 and 2). Each of these occurs at the Nla protease recognition sequence (Dougherty et al., 1988) between a glutamine (Q) and either a serine (S), glycine (G) or an alanine (A). The consensus cleavage site of PVA Nla protease – V–F(T)–Q–S/G/A – most resembles that of TVMV. The putative HC-pro protease cleavage site is located between G796 and G798 of the PVA polyprotein (Fig. 1) cleaving the peptide bond between the proteins HC-pro and P3. Comparison of the putative HC-pro protease recognition sites in potyviral polyproteins reveals that this sequence is highly conserved in all potyviruses that have been sequenced. The recognized sequence in PVA, – H–Y–R–V–G795–G796–T–S – is identical to the PVY8 and PepMoV HC-pro protease recognition sequence for the first six amino acids, but differs in the last two (Vance et al., 1992). The putative cleavage site between P1 and HC-pro of the PVA polyprotein is located at position Y797–S799 (Fig. 1) and is part of the putative protease recognition site H–Y–S, also found in PPV and TEV.

In addition to the putative polyprotein cleavage sites, we searched for common amino acid sequence motifs between PVA RNA-encoded proteins and those of the other potyvirus proteins that had been sequenced. The overall similarity of the PVA polyprotein was lowest with PepMoV (64.7%) and highest with TVMV (70.9%). When individual proteins were compared, it became apparent that P1 and P3 are the least conserved proteins among the potyviruses, whereas all other proteins have a high degree of similarity. Alignment of the coat protein sequence of PVA with that of several other potyviruses has been reported (Puurand et al., 1992). Taking into account that the N-terminal region of the coat protein is virus-specific in potyviruses (Shukla & Ward, 1989), the 'core' region of the coat protein is clearly the most highly conserved region of potyvirus polyproteins. CI, HC-pro, 6K1, Nla (pro) and Nlb proteins also show a high degree of similarity among corresponding potyvirus proteins. The first part of the conserved sequence consensus motif [S(T)GXXXTXXXNS(T) (18 to 37 amino acids, aa) GDD], which is conserved in a variety of both animal and plant positive-stranded RNA virus RNA-dependent polymerases (Kamer & Argos, 1984), is present in the PVA deduced protein Nlb in a position similar to that of other potyviruses (Fig. 1). A nucleotide-binding motif (GAVGSGKST), located near the N terminus of the putative CI protein, has previously been correlated with the helicase-like proteins (Hodgman, 1988). This sequence motif is strictly conserved also in PVA and located at the same position as that found in other potyviruses (Fig. 1 and 2). The sequences at the N-terminal region of the potyvirus polyprotein are less conserved among the potyviruses than those at the C terminus. Three conserved motifs have been found in the N-terminal part of TEV, TVMV, TuMV, PepMoV, PVY8, PPV, and now in PVA. First a 'zinc-finger'-like metal-binding motif (C-8aa-C-13aa-C-4aa-C-2aa-C) was found, which was first described in PVY8 by Robaglia et al. (1989) (Fig. 2). A second motif (C–72aa–H) is located in the HC-pro protein and is also strictly conserved among the potyviruses (Fig. 2). The second motif has been shown to be required for the protease activity of HC-pro, and is hence important for polyprotein processing (Oh & Carrington, 1989). The finding of common amino acid sequence motifs at the
analogous positions of all the potyviruses that have been sequenced indicates their functional importance.

We are indebted to Dr Frank Rabenstein, Dr Lilian Järvekülg, Dr Jari Valkonen, Dr Christophe Roos, Päivi Laamanen and Ulla Kalliomäki for their help. This work was supported by grants from the Finnish Ministry of Agriculture and Forestry and by grants to K.M. from the Kemira Foundation.

References


(Received 2 July 1993; Accepted 20 September 1993)