The Complete Nucleotide Sequence of Potato Virus X and Its Homologies at the Amino Acid Level with Various Plus-stranded RNA Viruses

By MARIANNE J. HUISMAN,1 HUUB J. M. LINTHORST,2 JOHN F. BOL2 AND BEN J. C. CORNELISSEN1*

1 MOGEN International N. V., Einsteinweg 97, 2333 CB Leiden and 2 State University of Leiden, Department of Biochemistry, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands

(Received 5 May 1988)

SUMMARY

Double-stranded cDNA of potato virus X (PVX) genomic RNA has been cloned and sequenced. The sequence [6435 nucleotides excluding the poly(A) tract] revealed five open reading frames (ORFs) which were numbered one to five starting at the 5' terminus of the RNA. They encoded proteins of \( M_r \), 165588 (166K), 24622 (25K), 12324 (12K), 7595 (8K) and 25080 (coat protein), respectively. ORFs 1 and 2 were in-phase coding regions. The ORF 1 product contained domains of homology with the tobacco mosaic virus 126K and 183K products. The ORF 2 and 3 products showed homologies with the barley stripe mosaic virus 58K and 14K proteins, the beet necrotic yellow vein virus 42K and 13K products and the white clover mosaic virus 26K and 13K products, respectively. The significance of these homologies with respect to putative functions of the PVX-encoded proteins are discussed.

INTRODUCTION

Potato virus X (PVX) is the type member of the potexvirus group. The flexuous rod-shaped virions (Stols et al., 1970) contain a single-stranded RNA of \( 2.1 \times 10^6 \, M_r \) (Bercks, 1970). At its 5' terminus the RNA contains a m\(^7\)GpppG cap structure (Sonenberg et al., 1978) and at its 3' end a poly(A) tract (Morozov et al., 1983). In a rabbit reticulocyte cell-free system the genomic RNA is translated into two predominant polypeptides with estimated \( M_r \) of 145K and 180K. It has been suggested that the larger protein arises by readthrough of a 'leaky' termination codon (Wodnar-Filipowicz et al., 1980). PVX virions contain solely the genomic RNA. However, in infected tissue two major subgenomic RNAs of 0-9 kb and 2-1 kb are present as well as four minor ones of 1-4, 1-8, 3-0 and 3-6 kb. All these subgenomic RNAs are T-coterminal (Dolja et al., 1987). It has been suggested that in vivo the coat protein is translated from the 0-9 kb subgenomic RNA.

The nucleotide sequence of the genomic RNA of PVX has been partly determined; the 5'-terminal 80 bases as well as 1300 nucleotides of the 3' end have been elucidated (Morozov et al., 1983, 1987). Morozov et al. (1987) analysed the amino acid sequence of the proteins encoded by the three open reading frames (ORFs) at the 3' end of the genome. They observed first of all that the coat protein sequence of PVX showed homologies with coat protein sequences of other potexviruses and even potyviruses. Secondly, they observed homologies between the PVX 12K protein and the barley stripe mosaic virus (BSMV) 14K and the beet necrotic yellow vein virus (BNYVV) 13K proteins. And thirdly, they observed homologies between the 8K protein and the sequence of a similar protein encoded by potato aucuba mosaic virus (another potexvirus). In this paper we report the complete nucleotide sequence and the organization of the PVX genome. The deduced amino acid sequences of the PVX-encoded proteins are compared with those of other viruses with capped plus-stranded RNA genomes. The possible functions of the PVX-encoded proteins are discussed.
**METHODS**

Enzymes were purchased from Bethesda Research Laboratories and Amersham, and chemicals from Sigma. Radiochemicals were from Amersham. Computer programs used were from the Genetics Computer Group of the University of Wisconsin, Madison, Wis., U.S.A.

**Virus purification and isolation of viral RNA.** PVX (strain X3) was isolated from *Nicotiana tabacum* cv. Samsun NN plants 2 weeks post-inoculation. Infected material was homogenized in 0.1 M-Tris–citric acid pH 9-0, 0.25% 2-mercaptoethanol. The supernatant was treated with 0.75 volumes of chloroform/n-butanol (1:1 v/v). The virus was precipitated from the water phase by addition of 0.2 volumes of 30% (w/v) polyethylene glycol. After centrifugation the pellet was resuspended in 0.1 M-Tris–citric acid pH 9-0. This suspension was ultracentrifuged twice to purify the preparation. The virus was stored in aliquots at a concentration of 26 mg/ml in Tris-citric acid pH 9-0 at -80 °C. Purified virions were phenol-extracted twice in 2% SDS, 1 x SSC (150 mM-NaH₂PO₄, 15 mM-sodium citrate pH 7-0). The RNA was ethanol-precipitated twice, dissolved in H₂O and stored at -20 °C.

**cDNA synthesis and cloning in λgt10.** cDNA synthesis and cloning was performed precisely as described in the instructions for use of the Amersham kits for cDNA synthesis and λgt10 cloning, respectively. For first strand cDNA synthesis an oligo(dT) primer was extended on PVX RNA by the use of reverse transcriptase. For the second strand synthesis RNase H, *Escherichia coli* DNA polymerase I and T4 DNA polymerase were used (Gubler & Hoffman, 1983). The double-stranded cDNA was treated with EcoRI methylase, provided with EcoRI linkers and ligated to the EcoRI-digested λgt10 arms. The DNA was packaged in vitro to obtain phage particles and recombinant phages were plated on *E. coli* NM514.

**Isolation of λgt10 clones.** The cDNA library was screened as described (Davis et al., 1986). Hybridization was carried out at 55 °C with an oligonucleotide primer (corresponding to bases 6004 to 6024) and at 65 °C with nick-translated recombinant plasmids. The hybridization solution used contained 5 x SSC, 5 x Denhardt's solution, 10% dextran sulphate, 0.1% SDS and 200 μg/ml denatured calf thymus DNA (Davis et al., 1986). Filters were washed in 0.1 x SSC, 0.1% SDS at the same temperature as was used for the hybridization. Phage DNA was obtained essentially as described (Davis et al., 1980). For further analysis the phage inserts were subcloned into pUC9. Subclones of PVX cDNA fragments in pUC9 were transformed to *E. coli* DH5α according to the method described by Dagert & Ehrlich (1979). Plasmid DNA was isolated as described by Holmes & Quigley (1981).

**cDNA sequencing.** cDNA sequencing was done by the dideoxy method (Sanger et al., 1977) using [α-³²P]dATP (Biggin et al., 1983). The cDNA was subcloned into M13 tg130/131 vectors (Kieny et al., 1983). Overlapping restriction fragments were subcloned and sequenced to cover the sites that were used for subcloning. Most of the DNA was sequenced in both orientations.

**RNA sequencing.** Direct sequencing of PVX RNA was used to sequence across the internal EcoRI sites in the genome, across some of the sites that were used for subcloning into M13 vectors, for parts of the genome devoid of useful restriction sites and to obtain the sequence of the 5' terminus. The procedure used was as follows. PVX RNA was annealed to a suitable primer (in a 100-fold molar excess) in a buffer containing 50 mM-Tris–HCl pH 8.3, 10 mM-MgCl₂, 40 mM-KCl by heating the mixture at 95 °C for 5 min and a subsequent incubation at 40 °C for 30 min. Dithiothreitol to a concentration of 10 mM, 2.5 units of reverse transcriptase and 10 μCi [α-³²P]dATP (sp. act. 600 mCi/mmol) were added. This mixture was then added to the termination mixtures. The final concentrations for dCTP, dGTP and dTTP were 62.5 μM and for dATP 12.5 μM. Each termination mixture contained one of the dideoxy NTPs ddCTP, ddGTP or ddTTP at a final concentration of 18.75 μM; in the case of ddATP the final concentration was 6.25 μM. The samples were incubated for 30 min at 40 °C and were chased with 0.3 volumes of 0.5 mM of all dNTPs.

**RESULTS**

**Nucleotide sequence of PVX genomic RNA**

Double-stranded cDNA of PVX was synthesized and cloned into λgt10. Inserts of selected clones were subcloned as EcoRI restriction fragments into pUC9. These subclones were used for further analysis. The PVX cDNA clones pX51, pX25 and pX633 appeared to span most of the PVX genome (Fig. 1). The nucleotide sequence of these three clones was determined by use of the dideoxy sequencing method on restriction fragments subcloned into M13 phages. The 29 bases at the extreme 5' end were not included in clone pX633 and were determined by direct dideoxy sequencing of the RNA by extension of a synthetic oligonucleotide complementary to bases 76 to 94. The complete nucleotide sequence of PVX RNA is shown in Fig. 2. Comparison of this sequence with the partial sequences reported before, i.e. the 5'-terminal 80 nucleotides and the 3'-terminal 1300 nucleotides, reveals minor differences (Fig. 2). At position 67/68 we found a GA sequence whereas Morozov et al. (1983) noted an AG sequence at that position.
Nucleotide sequence of PVX  

Fig. 1. Schematic representation of the organization of the PVX RNA genome and the localization of the cDNA clones used to determine the nucleotide sequence. The 5' end of the PVX RNA contains an m7GpppG cap structure (↑) and the 3' end a poly(A) tail (A0). CP, Coat protein.

Since the first AUG codon is found at base 84 it is unlikely that this difference represents a significant deviation. The differences in the 1300 nucleotides at the 3' terminus as compared with the sequence reported by Morozov et al. (1987) include an insertion of a GGA triplet coding for a glycine residue. The relevance of this amino acid insertion is discussed in the paragraph on homologies.

Coding sequences on the PVX genome

The PVX genomic RNA contains five ORFs (Fig. 1). These are preceded by an 84 base leader sequence. The first ORF (ORF 1) extends from the AUG at base 85 until the UAA stop codon at base 4453 and encodes a protein of Mf 165588 (166K). This cistron comprises about two-thirds of the coding capacity of the virus. In phase with this first ORF a second one is found, which starts at position 4486, 11 triplets downstream of the ORF 1 UAA stop, and ends at the UAG codon at position 5164. ORF 2 thus encodes a protein with an Mf of 24622 (25K). Seventeen bases upstream of the UAG stop codon of ORF 2 the translational start of a third ORF is found, which thus partly overlaps ORF 2. The stop signal (UAG) for ORF 3 resides at base 5492, and hence the ORF codes for a protein of Mf 12323 (12K). This 12K cistron shows not only an overlap with ORF 2 at its 5' terminus but also an overlap with ORF 4 at its 3' terminus (Fig. 1). This 3'-terminal overlap is 68 bases in length. The ORF 4 translational start and stop signals are located at positions 5427 and 5637, respectively; this ORF encodes an Mf 7594 (8K) product. The last ORF on the PVX genome starts at an AUG codon 13 bases downstream of the UGA stop codon of ORF 4. This last ORF encodes the PVX coat protein (Mf 25080). The translational stop (UAA) at position 6361 is followed by an untranslated region of 76 nucleotides and the poly(A) tract.

Homologies found with other plus-stranded RNA viruses

Recently, the complete nucleotide sequence of another potexvirus, white clover mosaic virus (WCIMV), has been elucidated (Forster et al., 1988). The organization of the WCIMV genome is identical to that of PVX. Comparison of the amino acid sequence of the PVX-encoded products with those of WCIMV shows the highest homology between the ORF 1-encoded proteins, i.e. 66% by the computer program GAP (data not shown). Comparison of the primary structure of the PVX ORF 1 product with the 126K and 183K proteins of tobacco mosaic virus (TMV) shows that two out of three domains of homology that are found in the comparisons of these TMV-encoded proteins with those of other viruses (Haseloff et al., 1984; Cornelissen & Bol, 1984; Ahlquist et al., 1985) are also present in the 166K protein of PVX (Fig. 3). The C-terminal domain of homology in the TMV 183K readthrough protein is characterized by the sequence motif S/TG---T---NS/T [18 to 37 amino acids (aa)]GDD. This motif is conserved in virus-encoded proteins putatively involved in RNA replication of most animal and plant plus-stranded RNA viruses (Kamer & Argos, 1984; Cornelissen & Van Vloten-Doting, 1988). The C-terminal portion of the PVX 166K protein contains the sequence TG ---T---NT(22 aa)GDD.
Fig. 2. The complete nucleotide sequence of the genomic RNA of PVX; the amino acid sequence
The nucleotide sequence of PVX derived therefrom is written above. The nucleotide differences between this sequence and the one published by Morozov et al. (1983, 1987) are underneath, and the effects on the amino acid sequence are given in parentheses above the deduced amino acid sequence.
and hence it fits the conserved motif. The middle part of the 166K protein contains a domain of homology with the C-terminal part of the TMV 126K protein; the homologous domain starts with the G--G-GKS sequence which has been found to be conserved as G--G-GKS/T in various ATP- and GTP-binding proteins (Zimmern, 1987, and references therein). Interestingly, the G--G-GKS/T sequence motif is also found in the N-terminal part of the 25K protein (Fig. 4). This protein bears considerable homology to the C-terminal half of the BNYVV 42K protein (Bouzoubaa et al., 1986), to the C-terminal two-thirds of the 58K protein of BSMV (Gustafson & Armour, 1986) and to the WCIMV 26K protein (Forster et al., 1988). The start of this extensive region of homology is marked by the G--G-GKS/T sequence motif (Fig. 4). The ORFs of BNYVV, BSMV, WCIMV and PVX coding for the 13K, 14K, 13K and 12K proteins, respectively, show appreciable homology at the aa level, which is centred around the sequence motif GD(7 to 8 aa)GG-YBDGS/TB (B stands for K or R) (Gustafson & Armour, 1986; Bouzoubaa et al., 1987). This homology has been observed before (Morozov et al., 1987); however, because our sequence has an extra GGA triplet the extent of the homology has considerably widened, since in our sequence of the PVX 12K protein, both the middle glycine residues of the sequence motif are present (Fig. 5). The extra GGA codon in our PVX 12K cistron may reflect a difference between the PVX isolates used by Morozov et al. (1987) and ours. If so, the question arises as to the biological meaning of conserved aa sequences such as are found in the sequence motif observed in the PVX 12K homologous proteins mentioned above. Unfortunately, no evidence is presently available about whether or not these smaller ORFs are expressed in vivo.

The PVX ORF 4 product (the 8K protein) shows homologies to similar proteins encoded by other potexviruses, notably WCIMV and potato aucuba mosaic virus (Forster et al., 1988; Morozov et al., 1987). The PVX coat protein has been demonstrated to have extensive aa sequence homologies with some potexvirus and potyvirus coat proteins (Sawyer et al., 1987; Morozov et al., 1983, 1987; Short et al., 1986; Forster et al., 1988).

DISCUSSION

Five ORFs are found in the sequence of the genomic RNA of PVX. However, upon addition of PVX genomic RNA to an in vitro translation system the major products are just two large polypeptides (Wodnar-Filipowicz et al., 1980). This has been suggested to be the result of translational readthrough of a leaky stop codon. At the end of the 5'-proximal ORF which encodes the 166K protein an ochre stop codon was revealed (Fig. 2). Readthrough of this stop codon would lead to the production of an M, 191 480 (191K) protein. Translational readthrough of the ochre stop codon would lead to the production of both the 166K and the 191K product. These two proteins could very well correspond to the 145K and 180K polypeptides found in translation studies in vitro (Wodnar-Filipowicz et al., 1980). However, three considerations seem to argue against translational readthrough in vivo. Firstly, to date, translational readthrough of an ochre stop codon has never been observed. Secondly, the translational start for the 25K protein is located 1971 bases upstream of the poly(A) tract, so the observation of a major subgenomic RNA of 2.1 kb strongly suggests that this RNA is the subgenomic messenger for the 25K ORF 2 product (Dolja et al., 1987). The third argument against readthrough originates from comparison of the PVX genome organization to that of WCIMV (Forster et al., 1988) where the amber stop codon at the end of the 147K protein is followed, out of phase, by the second ORF coding for a 26K protein.

Upon translation in vitro of the PVX RNA the predominantly detected products are two large polypeptides. None of the other ORFs seems to be translated from the genomic RNA. Another way to express cistrons is by producing subgenomic RNAs. Such RNAs have been detected for other potexviruses e.g. daphne virus X (Guilford & Forster, 1986), narcissus mosaic virus (NMV; Short & Davies, 1983), WCIMV (Forster et al., 1987) and clover yellow mosaic virus (Bendena et al., 1987). For NMV and the M isolate of WCIMV the subgenomic RNA for the coat protein is encapsidated, whereas for other potexviruses this has not been ascertained. As was pointed out before, the ORF 2 product might well be translated from the 2.1 kb major subgenomic RNA, whereas the coat protein might be translated from the other major subgenomic RNA of 0.9 kb.
Nucleotide sequence of PVX

Fig. 3. Schematic representation of regions of homology between the C-terminal domains of the TMV 126K and 183K proteins, and the two domains of the PVX 166K protein and the WCIMV 147K protein. Black boxes represent homologies.

Fig. 4. Parts (a) and (b) show alignments of two stretches of amino acid sequences from the PVX 25K (ORF 2) product with the 26K protein of WCIMV and the C-terminal parts of the BNYVV 42K and the BSMV 58K products. The homologous amino acid residues are boxed.

Fig. 5. Alignment of the amino acid sequence of the PVX 12K (ORF 3) product with the published PVX sequence (Morozov et al., 1987) (MPVX 12K) and the BSMV 14K, the BNYVV 13K and the WCIMV 13K products. Amino acid residues homologous to the PVX sequences are boxed.
It is possible that the two small virus-encoded products of 12K and 8K are translated from the 1.8 and 1.4 kb minor subgenomic RNAs, respectively.

Sequence comparisons have offered the opportunity to evaluate newly elucidated sequences, and, in particular, to address the question of which functions the ORF products perform in the viral replication cycle. The homologies between the PVX 166K and the TMV 126K/183K proteins suggest a common function. The C-terminal domains homologous to those of the TMV 126K and 183K proteins are present in products encoded by other plus-stranded RNA viruses (Cornelissen & Bol, 1984; Haseloff et al., 1984; Ahlquist et al., 1985). Comparison of the aa sequence of the PVX 166K protein with the TMV 126K and 183K proteins shows that the 166K protein contains both these two blocks of homology. In this respect the PVX 166K protein resembles the 237K protein of BNYVV (Bouzoubaa et al., 1987) which also contains both these blocks of homology on one protein. The C-terminal domain of the TMV 183K protein is thought to be of functional importance to the viral replicase (Kamer & Argos, 1984). Therefore, it is conceivable that the PVX 166K protein is involved in viral replication. The comparison of sequences gives us reason to believe that the PVX 25K, WCIMV 26K, BNYVV 42K and BSMV 58K proteins probably have some common function, although we do not know which function. The same holds for the comparison of the PVX 12K, WCIMV 13K, BNYVV 13K and BSMV 14K proteins. The homologies observed between BNYVV, BSMV and the potexvirus-encoded proteins suggest a closer evolutionary relationship between these viruses than hitherto assumed.

One of the PVX-encoded proteins performs a function in cell-to-cell spread of the virus. For TMV this transportation function was elegantly shown to reside on the 30K protein (Meshi et al., 1987). LS1, a mutant of TMV is temperature-sensitive (ts) in cell-to-cell spread of the virus (Nishigushi et al., 1978, 1980). The mutation responsible for the ts phenotype appeared to be a single aa change in the 30K protein (Ohno et al., 1983; Mashi et al., 1987). Taliansky et al. (1982) have shown that at the restrictive temperature systemic spread of LS1 could be observed provided the leaves had been preinoculated with PVX. This suggests that the 30K protein function can be provided in trans by a transportation function located on the genome of PVX. Proteins encoded by the PVX genome that might fulfil a function in viral spread are the 25K, 12K and 8K proteins, because these proteins have not as yet been definitely assigned functions.

Initiation of encapsidation from the 5' end onwards would explain why the PVX subgenomic RNAs are not encapsidated. The coat proteins of PVX, WCIMV and another potexvirus papaya mosaic virus (PMV) show considerable aa sequence homologies (Short et al., 1986; Sawyer et al., 1987; Harbison et al., 1988). For PMV it has been reported that the 5'-terminal sequences of the genomic RNA can be folded into a secondary structure that could interact with the PMV coat protein in such a way that in vitro assembly into virions could be observed (Lok & AbouHaidar, 1986). It can be envisaged that PVX and WCIMV RNAs are encapsidated in a similar fashion. WCIMV RNA has been predicted to fold into a secondary structure resembling the one described for PMV (Forster et al., 1988). Unlike the predictions for PMV and WCIMV the computer program FOLD predicts an unpaired 5' end up to nucleotide 32 for the 5'-terminal 150 nucleotides of the PVX genomic RNA (data not shown). However, since the computer-generated secondary structures still lack experimental or phylogenetic support, it is premature to speculate about the biological meaning of the differences found in the 5'-terminal secondary structures.

To make sure that the clones described represent a viable PVX molecule we are currently placing the full length cDNA behind the T7 promoter by the use of a synthetic oligonucleotide containing the T7 promoter sequence (Janda et al., 1987). This will enable us to produce RNA transcripts in vitro which can be tested for infectivity.

Ms Anne-Marie Krebbers is gratefully acknowledged for her endurance in preparing the manuscript. This work was sponsored in part by STW with financial aid from The Netherlands Organization for Advancement of Pure Research (ZWO).
REFERENCES


(Received 7 March 1988)