Volume 213, number 2, 438-442 FEB 04540 March 1987

# Nucleotide sequence of the open reading frames adjacent to the coat protein cistron in potato virus X genome

S.Yu. Morozov, L.I. Lukasheva, B.K. Chernov*, K.G. Skryabin* and J.G. Atabekov

*Department of Virology, Moscow State University, Moscow 119899* and **Institute of Molecular Biology, Academy of Sciences of the USSR, Moscow 117984, USSR*

Received 29 December 1986

The cloned cDNA copies corresponding to 1300 nucleotides adjacent to the 3'-terminal poly(A) tract of the potato virus X (PVX) genome have been sequenced. The amino acid sequences of three open reading frames were deduced from the nucleotide sequence. Two putative small nonstructural polypeptides corresponding to the open reading frames adjacent to the coat protein cistron possess some properties of membrane-associated proteins. Direct sequence homology and common structural peculiarities exist between the PVX small proteins and the putative small nonstructural proteins encoded by RNA 2 of hordeiviruses and furoviruses

cDNA; Cloning; Amino acid sequence; Sequence homology; Membrane-bound protein; (Plant virus)

## 1. INTRODUCTION

Potato virus X (PVX), a type member of the potexvirus group, has an infectious monopartite single-stranded RNA genome of ~2.1 × $10^6$ Da [1,2]. The genomic RNA of PVX has a cap structure at the 5'-terminus [3] and is polyadenylated at the 3'-end [4]. The length of the poly(A) tract in PVX RNA varies over a wide range with a maximum of about 250 nucleotides [4].

The presence of the 5'-terminal 'cap' and the 3'-terminal poly(A)-tail is typical for several potexviruses [4–7].

At least three proteins seem to be encoded by the genomes of potexviruses. The 3'-terminal regions of the genomic RNAs encode the coat protein [8–10], which is translated from a subgenomic RNA [7,11]. Two major large nonstructural polypeptides of 160–200 and 140–155 kDa are produced upon in vitro translation of the genomic RNAs of potexviruses [6,7,10,12]. These large products as well as some intermediate minor proteins are probably the members of a nested set of polypeptides.

This paper reports on the nucleotide sequence, organization and interviral homologies of the putative genes immediately proximal to the PVX coat protein cistron, of which the sequence has been previously published [8].

Correspondence address: S.Yu. Morozov, Dept of Virology, Moscow State University, Moscow 119899, USSR

## 2. MATERIALS AND METHODS

### 2.1. *RNAs, enzymes and nucleotides*

Virion PVX RNA was kindly provided by Dr V.K. Novikov. DNA polymerase I from *E. coli*, AMV reverse transcriptase, restriction endonucleases, bacteriophage $T_4$ DNA ligase and kinase were from Boehringer-Mannheim. RNase H was kindly provided by Dr N.P. Rodionova. $^{32}$P-labeled nucleotides were from Amersham. *Hin*dIII and *Bam*HI linkers were synthesized as in [13].



*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/87/$3.50 

### 2.2. *Synthesis and cloning of double-stranded cDNA*

PVX RNA was copied into cDNA and cloned as in [14–16]. The resulting cDNA clones were screened by colony hybridization [17] with oligo(dT)-primed, $^{32}$P-labeled, single-stranded cDNA. Plasmid DNA was isolated from colonies [16] hybridized with the probe. The cDNA inserts were sized by electrophoresis in a 0.8% agarose gel and characterized by restriction enzyme analysis.

### 2.3. *DNA sequencing*

The restriction fragments of cDNA inserts were sequenced by the method of Maxam and Gilbert [18] after 3′-labeling with DNA polymerase 1 (Klenow fragment).

## 3. RESULTS AND DISCUSSION

### 3.1. *Determination of the nucleotide sequence and identification of the open reading frames*

Two overlapping cDNA clones were selected for sequence analysis on the basis of restriction enzyme mapping and cross-hybridization (not shown). One of these clones, pX-17/10, started at position 445 from poly(A) and extended to the 5′-terminus by 1500 bp. A second cDNA clone, pX-1/6, started at position 725 and extended to the 5′-terminus by 1700 bp.

The DNA sequence equivalent to the 3′-terminal 1300 nucleotides of the PVX genomic RNA and the deduced polypeptide sequences are shown in fig.1. The sequence of 860 nucleotides preceding the poly(A) tract at the 3′-end of the PVX genomic RNA was determined previously using independent cDNA clones [8].

The sequenced region of the PVX RNA contains three open reading frames (ORFs) shown in fig.1. The first open reading frame (ORF1) commences at ATG in position 1286 and continues to position 945, corresponding to a polypeptide of 114 amino acids. The second reading frame (ORF2) initiates within ORF1 at nucleotide 1009 and terminates at nucleotide 800; ORF2 corresponds to a tentative protein of 70 amino acids. The third open reading frame (ORF3) starts from the ATG codon at position 786 and terminates at position 76, encoding a protein of 237 amino acids, which corresponds to the capsid protein of PVX [8]. This conclusion is

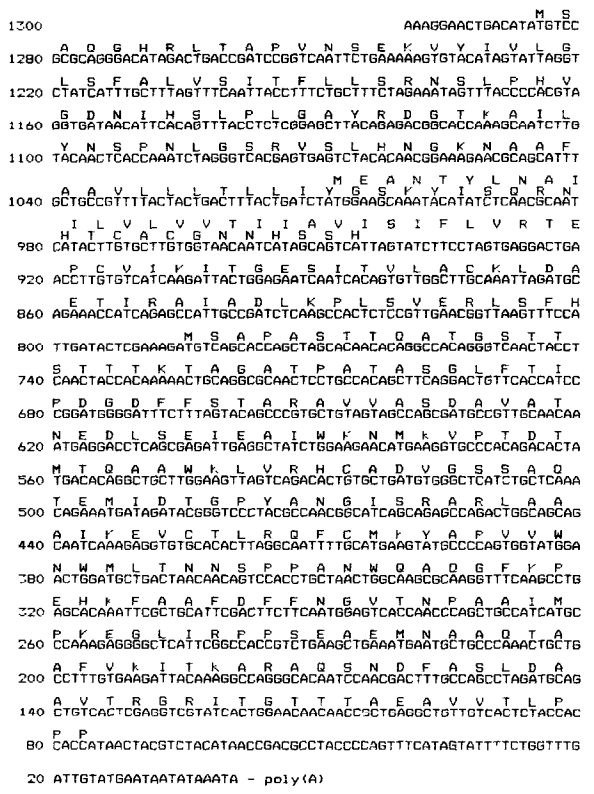


Fig.1. Sequence of 1300 nucleotides from the 3′-end of the PVX genome. The DNA sequence is shown as the equivalent of the viral RNA strand. Nucleotide residues are numbered from the first base in the vicinity of the poly(A) tract (indicated on the left). The three largest polypeptide sequences deduced from this sequence are shown using the single-letter amino acid code and numbered from the initiating methionine.

in accordance with data obtained upon comparison of the predicted amino acid sequence of the ORF3 translation product with that of the papaya mosaic virus (PMV) coat protein determined by direct sequencing [19]. Moreover, we have found unexpectedly that subsets of the PVX ORF3 translation product can be aligned in its central part with the coat proteins of the potyviruses. The sequences of the aligned segments are given in fig.2.

### 3.2. *Analysis of the putative noncoat genes*

PVX ORF1 and ORF2 have not as yet been shown experimentally to correspond to the actual virus-specific nonstructural proteins. Nevertheless,

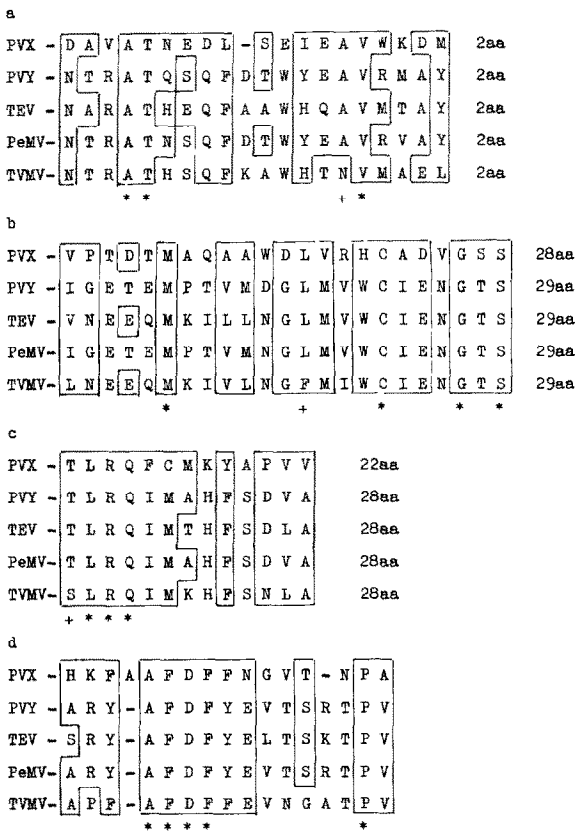


Fig.2. Homology of amino acid sequences between the PVX ORF3 translation product and the coat proteins of potyviruses. Four homologous sequence blocks are presented. The amino acid sequences were taken from PVX (50 amino acids from the N-terminus), potato virus Y (PVY) (81 amino acids from the N-terminus) [20], tobacco etch virus (TEV) (78 amino acids from the N-terminus) [21,22], pepper mottle virus (PeMV) (81 amino acids from the N-terminus) [21] and tobacco vein mottling virus [23]. Numerals show the numbers of amino acid residues which could not be aligned. Asterisks show the 15 invariant amino acid residues. Boxes exhibit the amino acid sequence similarity found in the PVX protein and the members of potyvirus coat proteins. Amino acid sequence similarity is the same as described in [22].

comparative analysis of ORF1 and ORF2 with the respective genome region of another potexvirus – potato aucuba mosaic virus (PAMV), support the idea of the functional significance of the polypeptides deduced from ORF1 and ORF2. Thus, an amino acid sequence homology of about 30% was revealed between the PVX ORF2 tentative product and the corresponding 73-amino-acid-long polypeptide deduced from the nucleotide sequence of PAMV RNA reported in [9] (fig.3). The amino acid sequence homology between the ORF1 product and its PAMV counterpart is about 40% (not shown).

It is interesting that the primary structures of both the small putative PVX-specific nonstructural polypeptides contain continuous sequences of uncharged amino acids, which resemble the essential features of the protein membrane-spanning sequences [24,25]. All such regions which comprise at least 16 amino acids are given in fig.4 together with a schematic representation of their localization.

Comparative analysis of the nucleotide sequences of the extended 3′-terminal regions of RNAs of potexviruses (PVX and PAMV) with those reported in the literature for RNA2a of barley stripe mosaic virus (BSMV) [26] and beet necrotic yellow vein virus (BNYVV) [27] revealed the principal similarity in the organization of ORFs coding for the tentative membrane-bound noncoat proteins.

Fig.4a shows that proteins p14 and p17 of BSMV, p13 and p17 proteins of BNYVV as well as PVX proteins encoded by ORF1 and ORF2 contain hydrophobic amino acid sequences. The length and the relative positions of the coding regions of PVX, BSMV and BNYVV putative 'membrane-bound' proteins are quite similar. Moreover, significant local sequence homologies exist between PVX ORF1-coded protein, its analog in potato aucuba mosaic virus genome, protein p14 coded by BSMV RNA 2a and protein p13 coded by BNYVV RNA2 (fig.4B). The region of

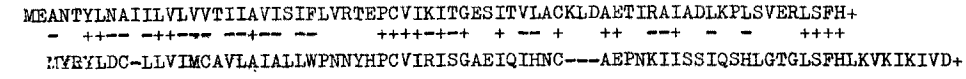


Fig.3. Alignment of PVX ORF2 translation product (top line) with its counterpart of PAMV (bottom line). Conserved (+) and similar (−) amino acids are denoted as described in [22].

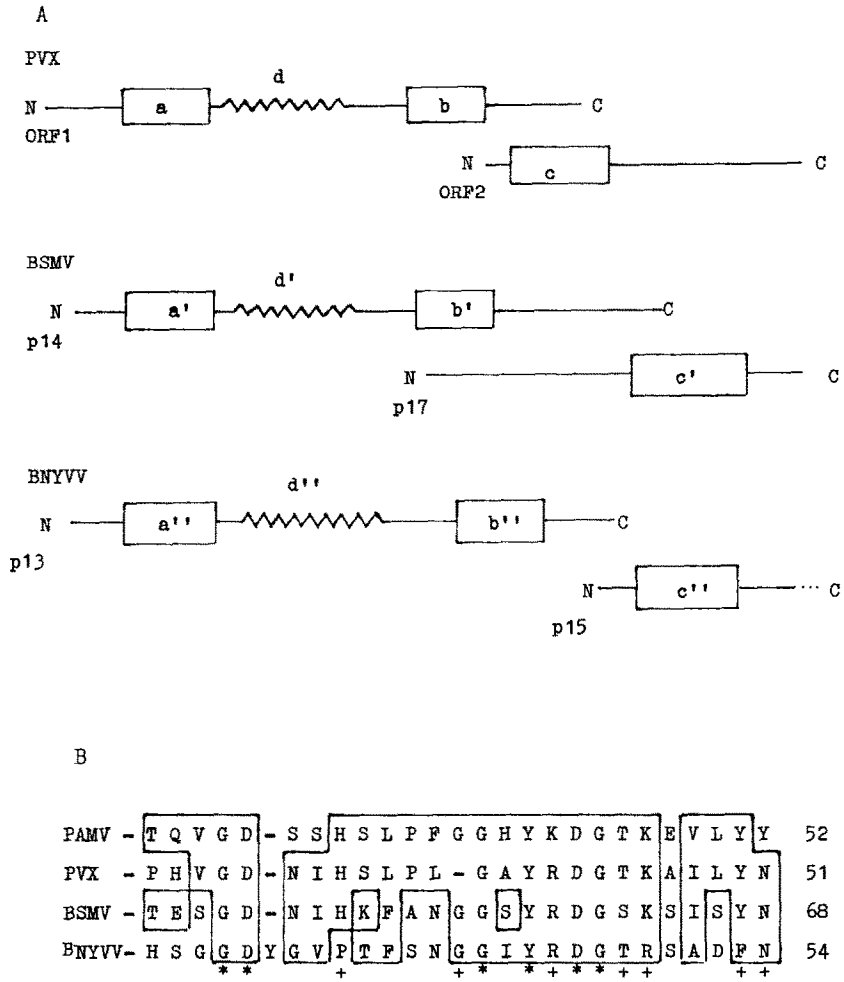


Fig.4. (A) Organization of the sequences of tentative small nonstructural proteins coded by PVX RNA, BSMV RNA 2a [26] and BNYVV RNA 2 [27]. The positions of the helical hydrophobic segments predicted according to [24,25] are indicated by the boxes (a,a′,a′′,b,b′,b′′,c,c′,c′′). Sets of conserved amino acid residues are indicated by zig-zag lines. (B) Alignment of the conserved amino acid sequences of the tentative PVX membrane-bound protein marked in the upper part of the figure with PAMV [9], BSMV [26] and BNYVV [27] products. Positions where two or more sequences consist of the identical and related amino acids are boxed. Amino acid sequence similarity is the same as described in [22]. The position of the C-terminus in each sequence is indicated by the numbers at the end of the sequences. Identical residues in all proteins or in 3 proteins are indicated below the sequences by * and +, respectively.

resemblance is located between amino acids 40 and 64 of the PVX protein (fig.1), 37 and 62 of the BSMV protein [26] and 38 and 64 of the BNYVV protein [27]. There are 15 identical positions between PVX and PAMV and 14 between PVX and BSMV (fig.4). Although the sequence homologies do not suggest a function for the putative membrane-bound proteins, our observations on the high degree of structural conservation between these small proteins of potexviruses, hordeiviruses (BSMV) and furoviruses (BNYVV) suggest that the encoded proteins may be involved in virus multiplication.

Analysis of the amino acid sequences encoded by the genomes of the plus-RNA-containing plant viruses has shown that, in addition to the potexvirus, hordeivirus and furovirus translation products, the long hydrophobic amino acid tracts

are located in the double-read-through domain of the p98 product coded by the carnation mottle virus genome [28] and in the putative protein of 48 kDa coded by the genomes of two potyviruses – tobacco etch virus [22] and tobacco vein mottling virus [23].

## REFERENCES


[1] Koenig, R. and Lesemann, D.-E. (1978) in: CMI/AAB Description of Plant Viruses no.200, 5 pp.

[2] Purcifull, D.E. and Edwardson, J.R. (1981) in: Handbook of Plant Virus Infections and Comparative Diagnosis (Kurstak, E. ed.) pp.627–693, Elsevier/North-Holland, Amsterdam, New York.

[3] Sonenberg, N., Shatkin, A.J., Ricciardi, R.P., Rubin, M. and Goodman, R.M. (1978) Nucleic Acids Res. 5, 2501–2512.

[4] Morozov, S.Yu., Gorbulev, V.G., Novikov, V.K., Agranovsky, A.A., Kozlov, Yu.V., Atabekov, J.G. and Bayev, A.A. (1981) Dokl. Akad. Nauk SSSR 259, 723–725.

[5] Abouhaidar, M.G. (1983) Can. J. Microbiol. 29, 151–156.

[6] Artuykova, E.V., Gorbulev, V.G., Rodionova, N.P., Krylov, A.V. and Atabekov, J.G. (1985) Mol. Biol. 19, 1021–1027.

[7] Guilford, P.J. and Forster, R.L.S. (1986) J. Gen. Virol. 67, 83–90.

[8] Morozov, S.Yu., Zakharyev, V.M., Chernov, B.K., Prasolov, V.S., Kozlov, Yu.V., Atabekov, J.G. and Skryabin, K.G. (1983) Dokl. Akad. Nauk SSSR 271, 211–215.

[9] Bundin, V.S., Vishnyakova, O.A., Zakharyev, V.M., Morozov, S.Yu., Atabekov, J.G. and Skryabin, K.G. (1986) Dokl. Akad. Nauk SSSR 290, 728–733.

[10] Bendena, W.G., Abouhaidar, M. and Mackie, G.A. (1985) Virology 140, 257–268.

[11] Short, M.N. and Davies, J.W. (1983) Biosci. Rep. 3, 837–846.

[12] Wodnar-Filipowicz, A., Skreczkowski, L.J., and Filipowicz, W. (1980) 109, 151–155.

[13] Barone, A.D., Tang, J.-V. and Caruthers, M.H. (1984) Nucleic Acids Res. 12, 4051–4061.

[14] Watson, C.J. and Jackson, J.F. (1985) in: DNA Cloning (Glover, D.M. ed.) vol.1, pp.79–88, IRL Press.

[15] Yanish-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.

[16] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[17] Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961–3964.

[18] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[19] Short, M.N., Turner, D.S., March, J.F., Pappin, D.J.C., Parente, A. and Davies, J.W. (1986) Virology 152, 280–283.

[20] Shukla, D.D., Inglis, A.S., McKern, N.M. and Gough, K.H. (1986) Virology 152, 118–125.

[21] Allison, R.F., Sorenson, J.C., Kelly, M.E., Armstrong, F.B. and Dougherty, W.G. (1985) Proc. Natl. Acad. Sci. USA 82, 3969–3972.

[22] Allison, R.F., Johnston, R.E. and Dougherty, W.G. (1986) Virology 154, 9–20.

[23] Donier, L.L., Franklin, K.M., Shahabuddin, M., Hellman, G.M., Overmeyer, J.H., Hiremath, S.T., Saaw, M.F.E., Lomonosoff, G.P., Shaw, J.G. and Rhoads, R.E. (1986) Nucleic Acids Res. 14, 5417–5430.

[24] Eisenberg, D., Schwarz, E., Komoromy, M. and Wall, R. (1984) J. Mol. Biol. 179, 125–142.

[25] Rao, J.K.M. and Argos, P. (1986) Biochim. Biophys. Acta 869, 197–214.

[26] Gustafson, G. and Armour, S.L. (1986) Nucleic Acids Res. 14, 3895–3909.

[27] Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K. and Jonard, G. (1986) J. Gen. Virol. 67, 1689–1700.

[28] Guilley, H., Carrington, J.C., Balarc, E., Jonard, G., Richards, K. and Morris, T.J. (1985) Nucleic Acids Res. 13, 6663–6677.