J. gen. Virol. (1988), **69**, 1509–1516. *Printed in Great Britain*

Key words: *plum pox virus/potyvirus/capsid protein gene*

Nucleotide Sequence of the Capsid Protein Gene of Plum Pox Potyvirus

By M. RAVELONANDRO,* C. VARVERI, R. DELBOS AND J. DUNEZ

Station de Pathologie Végétale, INRA, Domaine de la Grande Ferrade, BP 131, 33140 Pont-de-la-Maye, France

(*Accepted 5 April 1988*)

SUMMARY

The sequence of 1232 nucleotides from the 3′ region of the plum pox virus (PPV) RNA has been determined. This sequence contains one long open reading frame (ORF). The sequence of the 17 amino-terminal amino acids of the PPV capsid protein was determined chemically. An identical amino acid sequence was found in the translation product of the large ORF, starting at amino acid −330, which is alanine. Our data suggest that the PPV capsid protein, like that of other potyviruses, is a product of the maturation of a large polyprotein. The putative cleavage site is at a glutamine–alanine dipeptide. The capsid protein gene consists of 990 nucleotides and corresponds to a region coding for 330 amino acids which have a combined calculated M_r of 36593. The adjacent 3′ untranslated region is of 215 nucleotides and ends in a polyadenylate tract. PPV capsid protein molecules are thus larger than those of the other potyviruses that have been characterized. The amino acid sequence of PPV coat protein is 47 to 60% homologous to other potyvirus coat proteins and most of the variations are in the amino-terminal region.

INTRODUCTION

Plum pox potyvirus (PPV) causes large yield losses in stone-fruit trees. PPV particles are flexuous rods which contain an infectious RNA molecule with an M_r of about $3{\cdot}5 \times 10^6$ (Kerlan & Dunez, 1976). Potyvirus RNAs are translated into polyprotein precursors that are proteolytically processed into eight translation products (Hellmann *et al.*, 1983; Allison *et al.*, 1985; Dougherty *et al.*, 1985; Yeh & Gonsalves, 1985; Domier *et al.*, 1987). These are, from the amino terminus, an M_r 30000 to 64000 (30K to 64K) protein, a 48K to 56K helper component, a 40K to 50K protein, a 68K to 70K cylindrical inclusion protein, VPg, the 49K and the 54K to 58K nuclear inclusion proteins, and the 30K to 36K capsid protein at the 3′ end as in for example tobacco etch virus (TEV) (Allison *et al.*, 1985), tobacco vein mottling virus (TVMV) (Domier *et al.*, 1986), pepper mottle virus (PeMV) (Dougherty *et al.*, 1985) and Johnson grass mosaic virus (JGMV) (Gough *et al.*, 1987).

In this paper we report the nucleotide sequence of the PPV capsid protein gene and the amino acid sequence deduced for the PPV capsid protein. This protein has many regions of amino acid sequence in common with those of TEV, PeMV, potato virus Y (PVY) (Shukla *et al.*, 1986), JGMV (Shukla *et al.*, 1987) and TVMV.

METHODS

Virus and viral RNA. Seedlings of *Pisum sativum* cv. Express généreux were infected with PPV strain D and grown for 10 to 15 days (Kerlan *et al.*, 1975). Virus was purified from these plants and RNA was isolated from virus particles as described by Varveri *et al.* (1987). The purity of the RNA was confirmed by denaturing agarose gel electrophoresis (Lehrach *et al.*, 1977).

Synthesis and cloning of cDNA. Single-stranded (ss) cDNA was synthesized from 10 μg of PPV RNA primed with 1 μg of oligo(dT)$_{12-18}$ using reverse transcriptase (Life Sciences, St. Petersburg, Fla., U.S.A.). RNA–cDNA hybrids were incubated with RNase H (Bethesda Research Laboratories, BRL) and DNA polymerase I (Boehringer) (Gubler & Hoffman, 1983). Double-stranded (ds) cDNA was then treated with S1 nuclease (Boehringer) and fractionated by centrifugation in an SW50 Beckman rotor at 33000 r.p.m. for 17 h at 15 °C in a sucrose gradient (5 to 20%) in 50 mM-Tris–HCl pH 8, 1 mM-sodium EDTA and 800 mM-NaCl (Ravelonandro, 1985). The ds cDNA ends were made blunt by treating them first with T4 DNA polymerase (Boehringer) and then with the large fragment of DNA polymerase I (Amersham). Ds cDNA was ligated into the *Sma*I site of Bluescribe M13+ plasmid (Vector Cloning Systems) by T4 DNA ligase (Boehringer). Competent *Escherichia coli* DH5α cells (BRL) were transformed with the hybrid plasmids and grown on ampicillin plates in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactoside. Some clones from the cDNA library generated were screened, by *in situ* colony hybridization, with the nick-translated insert from pPPV9 obtained by Varveri *et al.* (1987).

DNA sequencing. Inserted DNA was isolated and then cut with suitable restriction enzymes. DNA fragments were separated in a 1% low melting point agarose gel (BRL) and then excised and eluted. The fragments were subcloned into M13 mp18 and M13 mp19 and sequenced by the dideoxy chain-termination method (Sanger *et al.*, 1977). The region between the nucleotides at positions 609 and 950 was also sequenced by primer extension from the oligonucleotide primer 5′ d(CCCCATCCATCATCACCC) 3′ (synthesized by the Laboratoire de Génétique, Université de Bordeaux II, France) on a recombinant DNA plasmid template as described by Chen & Seeburg (1985). The whole sequence was confirmed by the chemical method (Maxam & Gilbert, 1980). DNA fragments were treated with calf intestinal alkaline phosphatase (Boehringer), 5′-phosphorylated using T4 polynucleotide kinase (BRL), treated with a second restriction enzyme and, after purification from a 4% non-denaturing polyacrylamide gel (slab gel, 14 × 16 × 0·14 cm), were analysed by the chemical method. Some labelled DNA fragments were treated with alkali and, after purification on denaturing polyacrylamide gel, the resulting ssDNA fragments were sequenced. Thin wedge gels (30 × 40 × 0·035 cm and 20 × 80 × 0·035 cm) were used to separate the reaction products.

Characterization and amino-terminal sequencing of the capsid protein. SDS–PAGE of PPV capsid protein was carried out in a 12% polyacrylamide gel (Laemmli, 1970). Sequencing of the amino-terminal amino acids was performed on 100 μg of intact virus particles by sequential Edman degradation in a gas-phase Applied Biosystems 470A apparatus. The yield of this sequence determination was estimated to be 20%.

Computer-assisted analysis of the predicted amino acid sequences. Nucleotide sequence data were tested with UWGCG programs (Devereux *et al.*, 1984). The MAP program was used for analysing the restriction mapping of the insert and the different reading frames.

RESULTS

Molecular weight of PPV capsid protein and sequencing of the amino-terminal amino acids

SDS–PAGE (Fig. 1) shows that the PPV capsid protein consists of one protein species with an estimated M_r of 36K. Two minor components of M_r 28K and 26K were probably degradation products of the coat protein. Amino-terminal amino acids of PPV capsid were determined by chemical degradation of the intact virus in a gas-phase protein sequencer and yielded the following amino-terminal sequence: ADEREDEEEVDAXKPIV, where X corresponds to an undetermined amino acid.

Isolation and characterization of cDNA clones

Using the cDNA cloning conditions of Gubler & Hoffman (1983), we obtained clones ranging in size from 0·5 to 3·5 kb. Those which corresponded to the 3′ end of the genome were selected by colony hybridization. The probe was nick-translated DNA of clone pPPV9, which contains 3′-terminal sequences (Varveri *et al.*, 1987). One of these clones, pSE508, with an insert of 1·5 kb was selected and by restriction mapping was confirmed to overlap the 3′ cDNA contained in pPPV9 (Fig. 2).

DNA sequencing

A large part of the insert cDNA of pSE508 yielded the 1232 base sequence shown in Fig. 3. The sequence shows a large open reading frame (ORF) and a 3′ untranslated region. No other ORF of more than 70 codons (in the plus strand) and 94 codons (in the minus strand) was present. The chemically determined sequence of the 17 amino-terminal amino acids of the coat

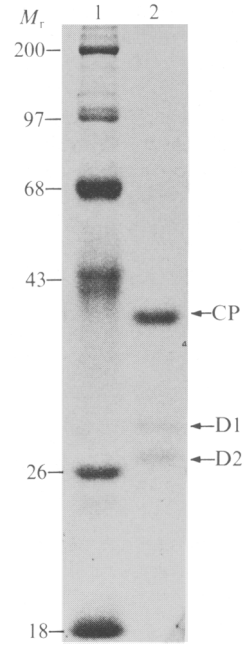


Fig. 1. SDS–PAGE of PPV coat protein. Lane 1, M_r ($\times 10^{-3}$) markers; lane 2, PPV coat protein. CP indicates the M_r 36K coat protein and D1 and D2 are putative degradation products of it.

protein corresponds exactly with the translation product predicted for this ORF from an alanine (GCU) codon at position −1208 from the polyadenylate tract. Comparison with the different coat proteins of potyviruses (Fig. 4) revealed that the presumed C terminus of PPV coat protein corresponds with those of other potyvirus coat proteins. This ORF encodes a protein of 330 amino acids with a calculated M_r of 36593, a value which is in agreement with the estimated M_r of the protein obtained by SDS–PAGE (Fig. 1).

The 3′ untranslated region contains 215 bases, excluding the polyadenylate tract which ranges from 50 to 70 adenosine residues in length (data not shown). Like those of other potyvirus RNAs, e.g. TEV (Allison *et al.*, 1985), PeMV (Dougherty *et al.*, 1985) and TVMV (Domier *et al.*, 1986), the 3′ untranslated region of PPV RNA is relatively A + U-rich (58%).

DISCUSSION

The 3′ untranslated region of PPV RNA does not show significant sequence homology with those of other potyviruses and no polyadenylation signals were found. The length of these regions of potyvirus RNA differs greatly, from 475 nucleotides for JGMV (Gough *et al.*, 1987) to 183 nucleotides for TEV (Allison *et al.*, 1985).

The coding nucleotide sequence and amino acid sequence presented in this paper enable location of the PPV capsid protein gene. As in other sequenced potyvirus RNAs, the PPV capsid protein gene is at the 3′ end of the coding sequence. This protein starts with an alanine codon (GCU) and terminates with a UAG codon. We deduce that this reading frame is flanked at its 5′ portion by a coding region, probably the nuclear inclusion gene, as shown for TEV (Dougherty & Hiebert, 1980). The putative site of protease cleavage is glutamine–alanine which is also that used in the proteolytic processing of the PeMV polyprotein (Dougherty *et al.*, 1985) and TVMV (Domier *et al.*, 1986). Cleavage at a dipeptide 'glutamine–X' seems to be a feature common to several viruses such as cowpea mosaic virus (van Wezenbeek *et al.*, 1983), TVMV (Domier *et al.*, 1986), TEV (Allison *et al.*, 1985) and poliovirus (Kitamura *et al.*, 1981).

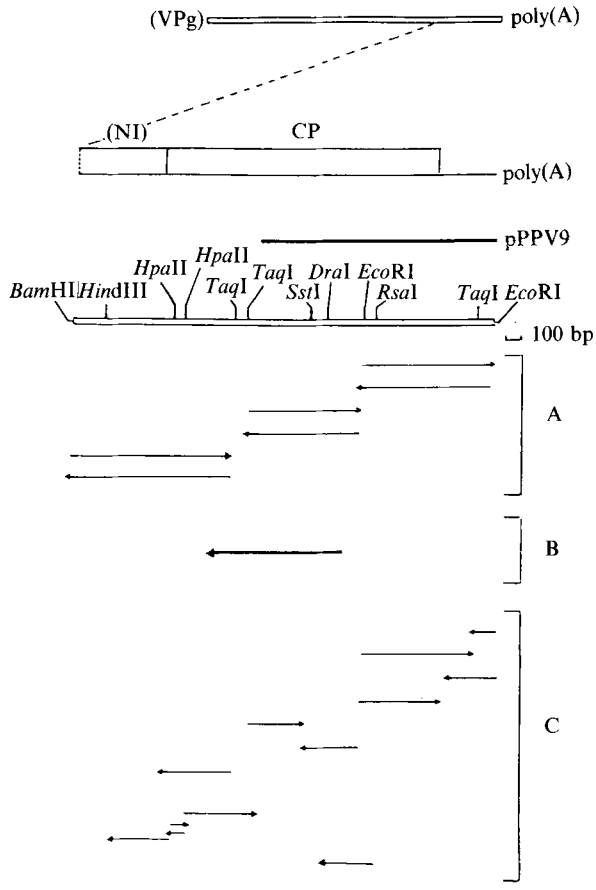
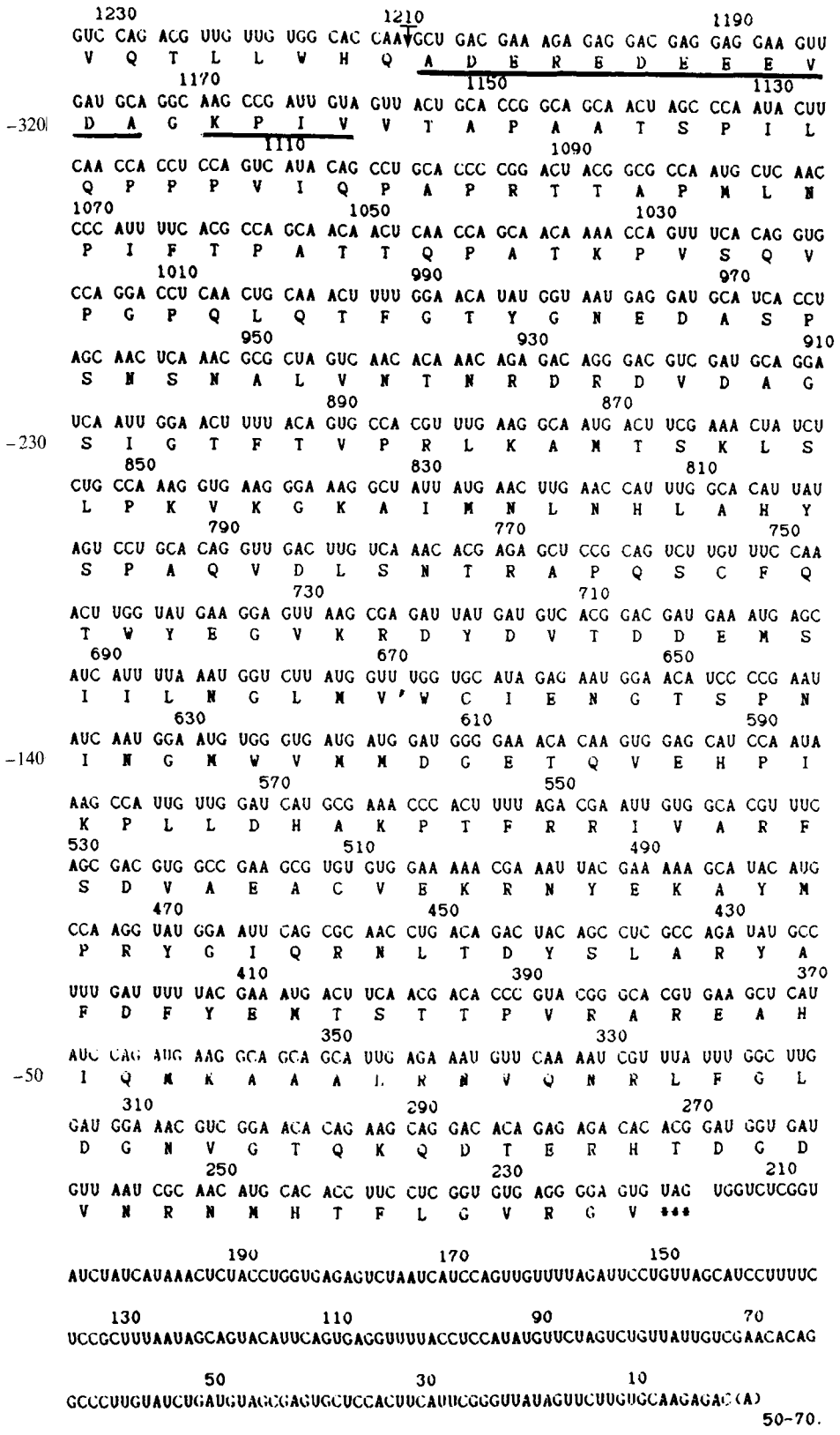


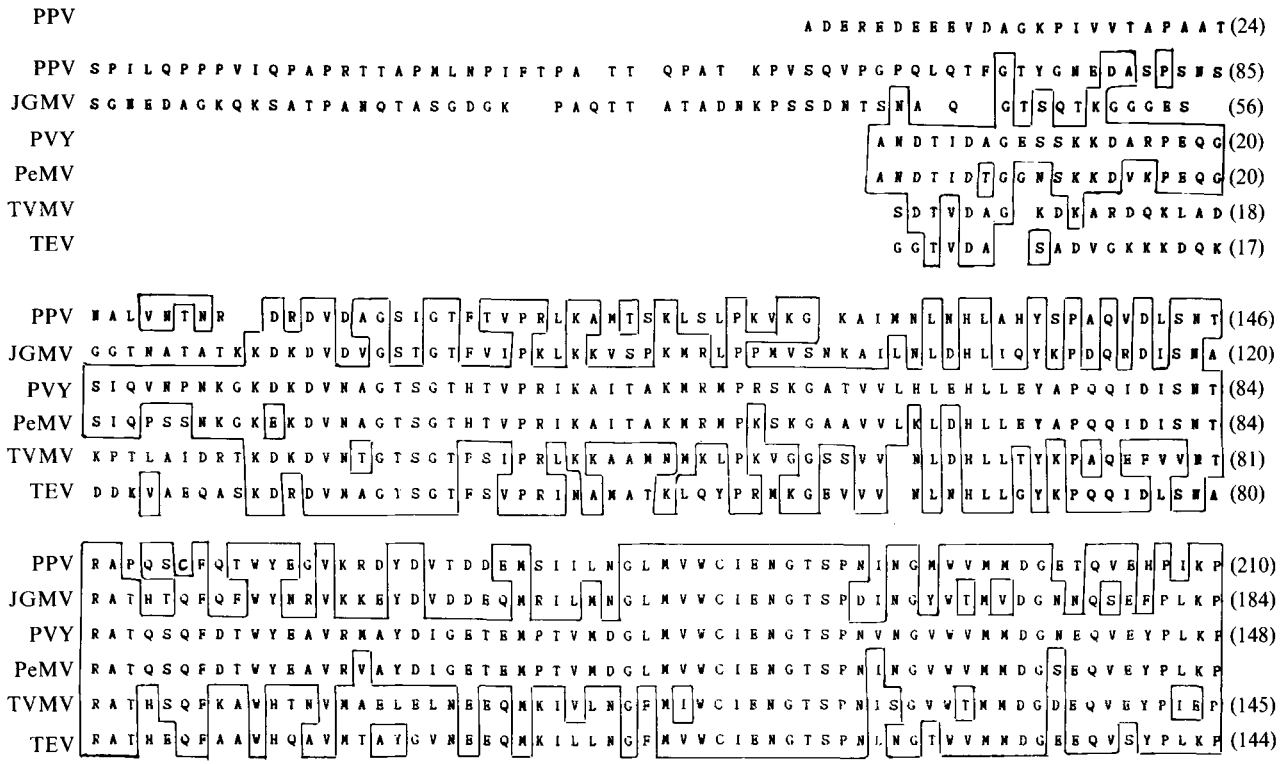
Fig. 2. Diagram of PPV RNA (top), the deduced location of the coding regions of the coat protein (CP) and nuclear inclusion protein (NI) genes, the location of pPPV9 sequences, the restriction enzyme map of cDNA insert in pSE508 and the sequencing strategy used: A, dideoxy sequencing of restriction fragments; B, primer extension; C, chemical sequencing. Arrows on the map indicate the polarity of the sequenced DNA strand and the length of the sequence information in nucleotides.

Analyses of amino-terminal amino acids showed that, like an isolate (NAT) of TEV (Allison *et al.*, 1985), PeMV (Dougherty *et al.*, 1985) and PVY (Shukla *et al.*, 1986), the coat protein of PPV does not have a blocked N terminus. The capsid of PPV is thus unlike those of JGMV and TVMV, both of which do have a coat protein with a blocked N terminus (Shukla *et al.*, 1987; Domier *et al.*, 1986).

Comparison of the PPV coat protein with those of other potyviruses suggests that the PPV polypeptide, with 330 amino acids, is longer than those (so far characterized) of the other potyviruses (263 to 303 amino acids). Alignment of amino acid sequences has shown that PPV coat protein has extensive homologies with other potyvirus coat proteins. The extents of homology are 60% (TEV), 59% (TVMV), 53% (PeMV), 52% (PVY) and 47% (JGMV). The amino-terminal regions are the least similar parts of the proteins (Fig. 4).

Fig. 3. Nucleotide sequence of the 1232 bases of the 3′ region of PPV RNA, showing the deduced amino acid sequence of the largest ORF. Nucleotides are numbered from the polyadenylate tract; amino acids are numbered to the left from the stop codon, which is indicated by three asterisks. The positions of underlined amino acids have been confirmed by Edman degradation. The glutamine–alanine cleavage site indicated by a vertical arrow represents the beginning of the PPV coat protein gene.


1230 1210 1190
GUC CAG ACG UUG UUG UGG CAC CAA GCU GAC GAA AGA GAG GAC GAG GAG GAA GUU
V Q T L L W H Q A D E R E D E E E V
1170 1150 1130
GAU GCA GGC AAG CCG AUU GUA GUU ACU GCA CCG GCA GCA ACU AGC CCA AUA CUU
-320 D A G K P I V V T A P A A T S P I L
1110 1090
CAA CCA CCU CCA GUC AUA CAG CCU GCA CCC CGG ACU ACG GCG CCA AUG CUC AAC
Q P P P V I Q P A P R T T A P M L N
1070 1050 1030
CCC AUU UUC ACG CCA GCA ACA ACU CAA CCA GCA ACA AAA CCA GUU UCA CAG GUG
P I F T P A T T Q P A T K P V S Q V
1010 990 970
CCA GGA CCU CAA CUG CAA ACU UUU GGA ACA UAU GGU AAU GAG GAU GCA UCA CCU
P G P Q L Q T F G T Y G N E D A S P
950 930 910
AGC AAC UCA AAC GCG CUA GUC AAC ACA AAC AGA GAC AGG GAC GUC GAU GCA GGA
S N S N A L V N T N R D R D V D A G
890 870
UCA AUU GGA ACU UUU ACA GUG CCA CGU UUG AAG GCA AUG ACU UCG AAA CUA UCU
-230 S I G T F T V P R L K A M T S K L S
850 830 810
CUG CCA AAG GUG AAG GGA AAG GCU AUU AUG AAC UUG AAC CAU UUG GCA CAU UAU
L P K V K G K A I M N L N H L A H Y
790 770 750
AGU CCU GCA CAG GUU GAC UUG UCA AAC ACG AGA GCU CCG CAG UCU UGU UUC CAA
S P A Q V D L S N T R A P Q S C F Q
730 710
ACU UGG UAU GAA GGA GUU AAG CGA GAU UAU GAU GUC ACG GAC GAU GAA AUG AGC
T W Y E G V K R D Y D V T D D E M S
690 670 650
AUC AUU UUA AAU GGU CUU AUG GUU UGG UGC AUA GAG AAU GGA ACA UCC CCG AAU
I I L N G L M V W C I E N G T S P N
630 610 590
AUC AAU GGA AUG UGG GUG AUG AUG GAU GGG GAA ACA CAA GUG GAG CAU CCA AUA
-140 I N G M W V M M D G E T Q V E H P I
570 550
AAG CCA UUG UUG GAU CAU GCG AAA CCC ACU UUU AGA CGA AUU GUG GCA CGU UUC
K P L L D H A K P T F R R I V A R F
530 510 490
AGC GAC GUG GCC GAA GCG UGU GUG GAA AAA CGA AAU UAC GAA AAA GCA UAC AUG
S D V A E A C V E K R N Y E K A Y M
470 450 430
CCA AGG UAU GGA AUU CAG CGC AAC CUG ACA GAC UAC AGC CUC GCC AGA UAU GCC
P R Y G I Q R N L T D Y S L A R Y A
410 390 370
UUU GAU UUU UAC GAA AUG ACU UCA ACG ACA CCC GUA CGG GCA CGU GAA GCU CAU
F D F Y E M T S T T P V R A R E A H
350 330
AUC CAG AUG AAG GCA GCA GCA UUG AGA AAU GUU CAA AAU CGU UUA UUU GGC UUG
-50 I Q M K A A A L R N V Q N R L F G L
310 290 270
GAU GGA AAC GUC GGA ACA CAG AAG CAG GAC ACA GAG AGA CAC ACG GAU GGU GAU
D G N V G T Q K Q D T E R H T D G D
250 230 210
GUU AAU CGC AAC AUG CAC ACC UUC CUC GGU GUG AGG GGA GUG UAG UGGUCUCGGU
V N R N M H T F L G V R G V ***
190 170 150
AUCUAUCAUAAACUCUACCUGGUGAGAGUCUAAUCAUCCAGUUGUUUUAGAUUCCUGUUAGCAUCCUUUUC
130 110 90 70
UCCGCUUUAAUAGCAGUACAUUCAGUGAGGUUUUACCUCCAUAUGUUCUAGUCUGUUAUUGUCGAACACAG
50 30 10
GCCCUUGUAUCUGAUGUAGCGAGUGCUCCACUUCAUUCGGGUUAUAGUUCUUGUGCAAGAGAC(A)
50-70.


PPV
PPV
JGMV
PVY
PeMV
TVMV
TEV
(24)
(85)
(56)
(20)
(20)
(18)
(17)
(146)
(120)
(84)
(84)
(81)
(80)
(210)
(184)
(148)
(145)
(144)

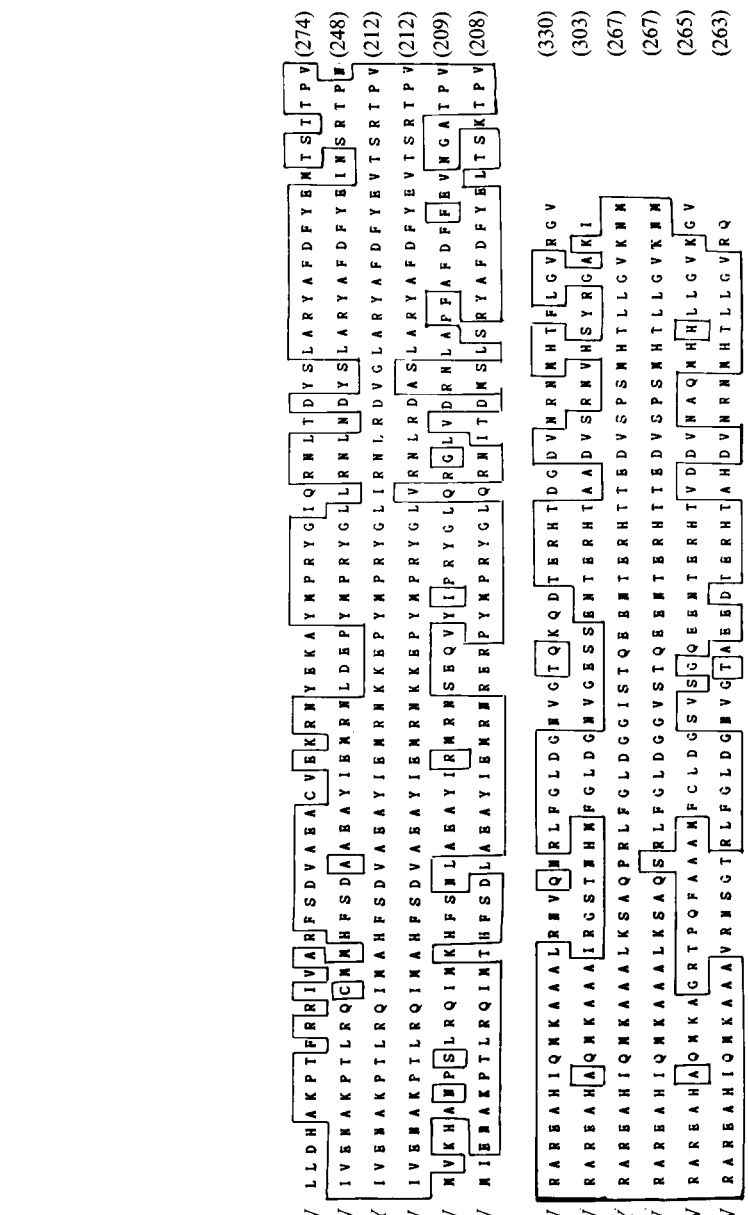


Fig. 4. Alignment of the amino acid sequences of the coat proteins of PPV, JGMV, PVY, PeMV, TVMV and TEV. This comparison was performed on the basis of the PVY coat protein data. The boxed amino acid residues and regions represent homologies in the sequences of the coat proteins of PPV, JGMV, PeMV, TVMV and TEV.

We thank Dr L. Pinck and Dr J. M. Dore for the generous gift of the Bluescribe vector. The amino terminus sequencing was done by Dr J. C. Pernollet and Dr J. C. Huet, from the Laboratoire d'Analyses de Protéines, Versailles, France.

REFERENCES

ALLISON, R. F., SORENSON, J. G., KELLY, M. E., ARMSTRONG, F. B. & DOUGHERTY, W. G. (1985). Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: evidence for the synthesis and processing of a polyprotein in potyvirus genome expression. *Proceedings of the National Academy of Sciences, U.S.A.* **82**, 3969–3972.

CHEN, E. Y. & SEEBURG, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4**, 165–170.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

DOMIER, L. L., FRANKLIN, K. M., SHAHABUDDIN, M., HELLMANN, G. M., OVERMEYER, J. H. HIREMATH, S. T., SIAW, M. F. E., LOMONOSOFF, G. P., SHAW, J. G. & RHOADS, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Research* **14**, 5417–5430.

DOMIER, L. L., SHAW, J. G. & RHOADS, R. E. (1987). Potyviral proteins share amino acid sequence homology with picorna-, como- and caulimoviral proteins. *Virology* **158**, 20–27.

DOUGHERTY, W. G. & HIEBERT, E. (1980). Translation of potyvirus RNA in a rabbit reticulocyte lysate: cell-free translation strategy and a genetic map of the potyviral genome. *Virology* **104**, 183–194.

DOUGHERTY, W. G., ALLISON, R. F., PARKS, T. D., JOHNSTON, R. E., FEILD, M. J. & ARMSTRONG, F. B. (1985). Nucleotide sequence at the 3′ terminus of pepper mottle virus genome RNA: evidence for an alternative mode of potyvirus capsid protein organization. *Virology* **146**, 282–291.

GOUGH, K. H., AZAD, A. A., HANNA, P. J. & SHUKLA, D. D. (1987). Nucleotide sequence of the capsid and nuclear inclusion protein genes from the Johnson grass strain of sugarcane mosaic virus RNA. *Journal of General Virology* **68**, 297–304.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

HELLMANN, G. M., THORNBURY, D., HIEBERT, E., SHAW, J. G., PIRONE, T. P. & RHOADS, R. E. (1983). Cell-free translation of tobacco vein mottling virus RNA. II. Immunoprecipitation of products by antisera to cylindrical inclusion, nuclear inclusion and helper component proteins. *Virology* **124**, 434–444.

KERLAN, C. & DUNEZ, J. (1976). Some properties of plum pox virus and its nucleic acid and protein components. *Acta horticulturae* **67**, 185–192.

KERLAN, C., DUNEZ, J. & BELLET, F. (1975). Méthodes simples de purification du virus de la Sharka (plum pox virus). *Annales de Phytopathologie* **7**, 287–297.

KITAMURA, N., SEMLER, B. L., ROTHBERG, P. G., LARSEN, G. R., ADLER, C. J., DORNER, A. J., EMINI, E. A., HANECAK, R., LEE, J. J., VAN DER WERF, S., ANDERSON, C. W. & WIMMER, E. (1981). Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature, London* **291**, 547–553.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 680–685.

LEHRACH, H., DIAMOND, D., WOZNEY, J. M. & BOEDTKER, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* **16**, 4743–4751.

MAXAM, A. M. & GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology* **65**, 499–560.

RAVELONANDRO, M. (1985). *Etude du genome du virus de la mosaïque de la luzerne (AlMV): séquence complète du RNA3. Structure des régions 5′ terminales non codantes des 4 RNA et relation possible avec l'efficacité de traduction.* Thèse de Doctorat, Université de Strasbourg.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SHUKLA, D. D., INGLIS, A. S., McKERN, N. M. & GOUGH, K. H. (1986). Coat protein of potyviruses. 2. Amino acid sequence of the coat protein of potato virus Y. *Virology* **152**, 118–125.

SHUKLA, D. D., GOUGH, K. H. & WARD, C. W. (1987). Coat protein of potyviruses. 3. Comparison of amino acid sequences of the coat proteins of four Australian strains of sugarcane mosaic virus. *Archives of Virology* **96**, 59–74.

VAN WEZENBEEK, P., VERVER, J., HARMSEN, J., VOS, P. & VAN KAMMEN, A. (1983). Primary structure and gene organization of the middle-component RNA of cowpea mosaic virus. *EMBO Journal* **2**, 941–946.

VARVERI, C., RAVELONANDRO, M. & DUNEZ, J. (1987). Construction and use of a cloned cDNA probe for the detection of plum pox virus in plants. *Phytopathology* **77**, 1221–1224.

YEH, S.-D. & GONSALVES, D. (1985). Translation of papaya ringsport virus RNA *in vitro*: detection of a possible polyprotein that is processed for capsid protein, cylindrical-inclusion protein and amorphous inclusion protein. *Virology* **143**, 260–271.

(Received 3 November 1987)