Nucleotide Sequence of Potato Virus Y (N Strain) Genomic RNA

By CHRISTOPHE ROBAGLIA, MYLIÈNE DURAND-TARDIF, MAURICE TRONCHET, GENEVIÈVE BOUDAZIN, SUZANNE ASTIER-MANIFACIER and FRANCINE CASSE-DELBART

Laboratoire de Biologie Cellulaire and Station de Pathologie Végétale, INRA, Centre de Versailles, 78026, Versailles Cedex, France

(Accepted 28 November 1988)

SUMMARY

The complete nucleotide sequence of the genomic RNA of the potyvirus potato virus Y strain N (PVYn) was obtained from cloned cDNAs. This sequence is 9704 nucleotides long and can encode a polyprotein of 3063 amino acids. The positions of the cleavage sites at the N terminus of the capsid and cytoplasmic inclusion proteins have been determined. Other putative protein cleavage sites have been deduced by searching for consensus sequences and by analogy with the polyprotein of the tobacco vein mottling virus and of the tobacco etch virus. Comparison of the PVY polyprotein sequence with that of other potyvirus polyproteins shows similarities in genome organization and a high level of identity along most of the polyprotein, except for the putative proteins flanking the helper component. A search for specific protein motifs has revealed the existence of a potential metal-binding site at the putative N terminus of the helper component in potyviruses. The possible functions of this structure are discussed.

INTRODUCTION

Potato virus Y (PVY) is the type member of the potyvirus group which forms the largest and economically the most important of the plant virus groups recognized by the International Committee on Taxonomy of Viruses (Hollings & Brunt, 1981). The particles of these viruses are flexuous, rod-shaped and contain a molecule of single-stranded RNA about 10 kb long with a genome-linked protein (VPg) at the 5' end and a poly(A) stretch at the 3' end (Hari, 1981; Hari et al., 1979). The nucleotide sequence of two potyvirus RNAs, those of tobacco etch virus (TEV) (Allison et al., 1986) and tobacco vein mottling virus (TVMV) (Domier et al., 1986) have recently been determined. This has confirmed that these genomes are translated into a single large polyprotein which is subsequently processed by proteolytic cleavage into smaller, functional proteins, as was previously deduced from in vitro translation experiments (Yeh & Gonzalves, 1985; Hellman et al., 1983; Vance & Beachy, 1984).

A partial genetic map of these viruses has been established and cistrons have been assigned to certain virus-encoded proteins; these are, from 5' to 3', the helper component (HC), the cytoplasmic inclusion protein (CI), the two nuclear inclusion proteins and the capsid protein. It was shown that one of the nuclear inclusion proteins (N1a) is a protease (Carrington & Dougherty, 1987a; Hellman et al., 1988), and protein sequence similarities with picornaviruses and comoviruses have suggested that the other nuclear inclusion (N1b) is involved in RNA replication (Allison et al., 1986; Domier et al., 1987).

In this article we report the cloning and the determination of the complete nucleotide sequence of the PVY (N strain; PVYn) RNA and the analysis of its deduced amino acid sequence.

METHODS

Virus purification and RNA isolation. Virus purification was as described by Dougherty & Hiebert (1980) with some modifications: the extraction buffer was supplemented with 3% urea, and the last step was a 10 to 40%
sucrose gradient in 20 mM-HEPES pH 7.5. The final pellet was resuspended in sterile water. For RNA isolation, the purified virus was diluted with an equal volume of extraction buffer (200 mM-ammonium carbonate, 2 mM-EDTA, 2% SDS, pH 9) and proteinase K was added at 15 µg/ml of virus. After 15 min at room temperature, the mixture was centrifuged in a log linear sucrose gradient (Brakke & Van Pelt, 1970) in twofold diluted extraction buffer, at 39000 r.p.m. for 240 min in an SW41 Beckman rotor. The RNA fraction was precipitated with 2.5 volumes of cold ethanol after addition of sodium acetate (100 mM final concentration), and resuspended in sterile water.

Sequencing of the N-terminal part of the capsid protein and of the CI protein. Sequencing of the capsid protein was performed on virus purified according to Ross (1967), and resuspended in water. The CI protein was purified according to Hiebert & MacDonald (1973) and resuspended in water. Sequencing was performed by the Edman degradation technique on an Applied Biosystems automated protein sequencer. A large amount (50 µg) of the cytoplasmic inclusion protein was analysed in order to overcome the blocking of the N-terminal serine.

cDNA cloning and sequencing. Most of the sequence information was obtained from a shotgun cloning procedure (M. Durand-Tardif & C. Robaglia, unpublished data). Briefly, double-stranded cDNA was synthesized according to Gubler & Hoffman (1983), using random hexamers (Pharmacia) as first strand primers. After treatment with S1 nuclease, repair with T4 DNA polymerase and sizing on a low melting temperature agarose gel, the cDNA was ligated to Smal-cut, dephosphorylated Bluescript vector (Stratagene) and recombinant plasmids were introduced into Escherichia coli NM522 (Pharmacia). Recombinant clones were picked at random and sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977) using either Klenow polymerase or T7 DNA polymerase (Sequenase, US Biochemicals).

Other clones were generated by inserting double-stranded cDNA into pTZ19 (Pharmacia) by dG/dC tailing (Maniatis et al., 1982). Among these, two clones, pY139 and pY118, were partially sequenced using a deletion strategy (Lin et al., 1985), and one clone, pY117, was partially sequenced by subcloning restriction fragments into Bluescript. These clones were selected by hybridization with a mixture of synthetic 20-mer nucleotides GC(A/G)AATGATAC(A/T)AT(C/T/A)GA(T/C)GC deduced to correspond to sequence coding for the amino acid sequence of the N terminus of the capsid protein.

Single-stranded DNA was generated by superinfection with the helper phage M13K07 (Pharmacia). Plasmid DNA was also used as template (Chen & Seeburg, 1985).

All the primers were synthesized on an Applied Biosystems 381A DNA synthesizer using β-cyanoethylphosphoramidites; they were deprotected, ethanol-precipitated and used without further purification.

5' End sequence determination. Viral RNA was annealed to a 16-mer primer (GCCATTGAGGATCTGA, positions 173 to 189) labelled with [α-32P]ATP and extended with reverse transcriptase in the presence of dideoxynucleotides as described by Geliebter (1987). The sequence was also determined according to Bassel-Duby et al. (1986) using the same unlabelled primer and [α-35S]ATP as label in the reaction.

Computer analysis. All the computer analyses were performed using different programs available from the Centre Interuniversitaire de Traitement de l’Information (CITI 2, Faculté de médecine, Paris). A shotgun sequencing program was used for the assembly of the complete sequence and the search for similarities was performed using the Kanehisa algorithm (Goad & Kanehisa, 1982).

RESULTS

Sequence of PVYn genomic RNA

The sequence of about 150 clones from a randomly primed cDNA library combined with sequence information from certain clones previously constructed by oligo(dT) priming allowed us to obtain nearly 90% of the sequence. The remainder was obtained by sequencing either plasmid templates or the purified RNA with specific oligonucleotides as primers. The sequence was determined from DNA in both orientations (an average of 4-5 determinations on at least two independent clones was achieved for each base) except for the 150 nucleotides at the 5' end which were not represented in any clone and therefore were directly sequenced on the viral RNA.

We do not know whether the extreme 5' end of the genome has been reached since the presence of the small genome-linked protein (VPg) at the 5' end may prematurely stop the reverse transcription reaction.

The nucleotide sequence of PVYn RNA is shown in Fig. 1. The genome is 9704 nucleotides long followed by a poly(A) tail. A 5' non-coding region of 185 bases precedes the unique long open reading frame which can encode a polyprotein of 3063 amino acids (Fig. 1) and is followed by a 3' non-coding region of 331 bases. The two other reading frames on the positive-strand viral
RNA are frequently punctuated by termination codons and therefore are not likely to code for proteins of significant size. The first AUG codon at position 185 is likely to be the initiator codon since it is in a context (UCAAUGGC) similar to the consensus sequence for translation initiation in plants (AACAAUGGC) as described by Lütcke et al. (1987), especially if one considers that the adenine at position -3 is not as strictly conserved in plant genes as it is in animal genes (Kozak, 1986). The overall base composition is 31.1% A, 18.7% C, 23.4% G, 26.8% U, whereas that of the 185 nucleotide leader sequence is 41% A, 25.4% C, 7% G and 26.5% U; this very low G content seems to be a common feature of plant viral 5' leader sequences (Gallie et al., 1987).

**Assignment of polyprotein cleavage sites**

Shukla et al. (1986) have sequenced the entire PVY (strain D) capsid protein, and we have confirmed that the N terminus is the same for the capsid protein of the N strain (A/G,NDTIDAGGSNKKDKADQEQG); this allows us to map the N terminus of the capsid protein at position 2796. We have also determined the amino acid sequence of the N terminus of the CI protein (SLDDVI) and therefore we can map it at position 1157. As the C-terminal halves of the polyproteins encoded by potyvirus genomes are highly conserved (see below), it is possible to postulate that the junction between the small (NIa) and the large (NIb) nuclear inclusion proteins is at position 2275 on the PVY polyprotein, and the possible cleavage site at the C terminus of the cytoplasmic inclusion protein is at position 1791.

The probability that this latter cleavage site is used is strengthened by the fact that its surrounding sequence is strictly identical to that found at the verified junction between the NIb nuclear inclusion and the capsid proteins (Fig. 2) a situation that also occurs in the TVMV polyprotein. In the TEV polyprotein the homologous site has been shown to be the substrate for autocatalytic cleavage by the M, 49000 (49K) protease (Carrington & Dougherty, 1987b). Examination of the sequences downstream from this region reveals a highly conserved sequence corresponding to a possible cleavage site at position 1843 on the PVY polyprotein. These two sites define a small protein of 52 amino acids which is the best candidate for the VPg. The hydropathy profile of this region is very similar in the proteins of the three potyviruses (Fig. 3).

Domier et al. (1986) have proposed a consensus sequence for these cleavage sites on the TVMV polyprotein: V(R or K)FQ/(G, S or A). Allison et al. (1986) have determined the N-terminal sequence of the TEV capsid protein and Carrington & Dougherty (1987a, b) have determined the N-terminal sequence of in vitro synthesized TEV 49K protease and have proposed a consensus cleavage site for this enzyme: E--Y-Q/(G or S). These sites on the PVY polyprotein bear the consensus sequence V-(H or E)Q/(G, S or A) (Fig. 2).

The situation is less clear towards the N-terminal half of the polyprotein. The presence of a 28K and a 42K protein on either side of the HC has been postulated in the TVMV polyprotein. The corresponding TEV proteins would be respectively 34K and 42K (Domier et al., 1987). By analogy with the TVMV protein sequence, the junction between the putative extreme 5' protein and the HC of PVY could be mapped to position 275 on the PVY polyprotein; however, comparison with the TEV protein sequence, in which this site is mapped at position 264 (Domier et al., 1987), implies that the highly conserved sequence among the three viruses, FIVRG (starting at position 256 on PVY polyprotein), will be part of the upstream protein in PVY and TVMV, and part of the downstream HC in TEV (Fig. 4). A similar situation occurs at the possible C terminus of the HC; examination of the PVY protein sequence reveals it to be at position 824. If this site is utilized, it would predict that the conserved sequence which lies from position 775 to position 810 on PVY will be part of the HC in PVY and part of the putative 42K protein in TEV and TVMV.

The positions of the putative cleavage sites and their comparison with those of TVMV and TEV are shown in Fig. 2. Taken together, these data predict that seven proteins of 31K, 62K, 38K, 71K, 56K, 55K, 60K and 30K could be encoded by the PVY genome. These predictions agree with the experimental Mr determinations since estimates were 66K for the CI protein and 34K for the capsid protein (Hiebert & MacDonald, 1973) and more recently, Thornbury et al. (1985) have reported that the HC could be a protein of 75K.
AAP1AAACACGAGCAAAUGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUGCAUGUAIACUUAGGUG
Fig. 1. Nucleotide sequence of PVY\textsubscript{n} RNA and deduced amino acid sequence of the polyprotein. The positions of the potential or experimentally determined cleavage sites are underlined.
<table>
<thead>
<tr>
<th>Protein</th>
<th>PVY</th>
<th>TVMV</th>
<th>TEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>28K-34K/HC</td>
<td>ARSRVTQ</td>
<td>ARARFSQ</td>
<td>KNERVDQ</td>
</tr>
<tr>
<td></td>
<td>S1NLSM</td>
<td>S1TLRA</td>
<td>SKLFYG</td>
</tr>
<tr>
<td>HC/42K</td>
<td>LWINEKQ</td>
<td>AVKALTQ</td>
<td>NRDVYTQ</td>
</tr>
<tr>
<td></td>
<td>SLAMIA</td>
<td>AFIRPD</td>
<td>GAIEML</td>
</tr>
<tr>
<td>42K/CI</td>
<td>DYEVRHQ</td>
<td>ENNVRFQ</td>
<td>REIYQ</td>
</tr>
<tr>
<td></td>
<td>SDLIVD</td>
<td>SDLTV</td>
<td>SDLDDYV</td>
</tr>
<tr>
<td>CI/VPG</td>
<td>LQFVHHQ</td>
<td>LEAVRFQ</td>
<td>LETYQ</td>
</tr>
<tr>
<td></td>
<td>AAATSLA</td>
<td>SESEMA</td>
<td>SDSEVA</td>
</tr>
<tr>
<td>VPg/Protease (N1a)</td>
<td>VETVSHQ</td>
<td>HEPVRFQ</td>
<td>NEPVYFQ</td>
</tr>
<tr>
<td></td>
<td>GKNKSK</td>
<td>GKSRRR</td>
<td>GKNKQK</td>
</tr>
<tr>
<td>Protease (N1a)/Polymerase (N1b)</td>
<td>HDVVVEQ</td>
<td>DDLVRTQ</td>
<td>NELVYSG</td>
</tr>
<tr>
<td></td>
<td>AKHSAW</td>
<td>GEKRK</td>
<td>GEKRK</td>
</tr>
<tr>
<td>Polymerase (N1b)/Capsid</td>
<td>SYEVBHQ</td>
<td>RETVRFQ</td>
<td>TENVYFQ</td>
</tr>
<tr>
<td></td>
<td>ANDTID</td>
<td>SDTVAG</td>
<td>SGTVDA</td>
</tr>
</tbody>
</table>

Fig. 2. Comparison of the predicted cleavage sites of PVY, TVMV and TEV polyproteins.
Comparison of PVY nucleotide sequence

943

Hydrophobic

(a)

(b)

(c)

Hydrophilic

Fig. 3. Comparison of hydropathy profiles (Kyte & Doolittle, 1982) of potyvirus VPg regions. (a) PVY, from 1770 to 1860; (b) TVMV, from 1725 to 1825; (c) TEV, from 1775 to 1865. Window: 10 amino acids.

Fig. 4. Protein sequence alignment of PVY, TVMV and TEV. Numbers are from 170 to 346 and from 729 to 864 on the PVY amino acid sequence. Arrows indicate putative proteolytic cleavage sites. Residues which could be involved in metal binding are boxed.

Comparison of PVY proteins with other potyvirus proteins

The most conserved protein among the three potyviruses PVY, TVMV and TEV is the NiB nuclear inclusion protein which displays 58% identity between PVY and TVMV, and 63% identity between PVY and TEV. This protein shares similarities with picornavirus and comovirus proteins thought to be involved in the replication process (Allison et al., 1986; Domier et al., 1987) and is the best candidate for the core RNA-dependent RNA polymerase of potyviruses.

It is now established that the NiA protein of TEV and TVMV is a protease involved in the cleavage of at least a part of the polyprotein (Carrington & Dougherty, 1987a, b; Hellman et al., 1988). The TEV protein shares 48% identity with its counterpart in PVY which is likely to have
PVYn | AMTIDAGOSNKKDAPKPEQS|OPNPNGKKDKUNAGTSTHTUPR|IKAITSKMAMPTS|GSTULN|LEHLEV
PVYd | E|S|A|V|A|R|H
PeMV | T|H|U|S|E|A|K|A|K

Fig. 5. Alignment of the capsid protein sequences of PVY strain N, PVY strain D and PeMV. Only variable amino acids are indicated.

the same function. A cysteine residue (position 2182) and a histidine residue (position 2197) near the C terminus of the protein may be involved in the active site of a thiol protease (Carrington & Dougherty, 1987b).

The capsid protein of PVY is 54% identical to that of TVMV, 62% identical to that of TEV, and 87% identical to that of pepper mottle virus (PeMV) (Dougherty et al., 1985). The number and the location of the differences between these proteins are in the same order as those observed between the PVYn used in this study and the PVY strain D studied by Shukla et al. (1986). This would suggest that PeMV may be in fact a strain of PVY instead of a different potyvirus species (Fig. 5).

The proposed CI protein of PVY is 56% and 53% identical to that of the corresponding proteins of TEV and TVMV, respectively. However, the proteins encoded by the 5' part of the genomes are less similar (Fig. 6). The HC is 48% identical to that of TVMV and 46% to that of TEV. The putative 38K polypeptide shows about 30% identity with both of its 42K counterparts and the putative 31K protein encoded by the extreme 5' end of the PVY RNA is only 24% identical to the putative 28K protein of TVMV and could not be aligned (with the standard parameters used throughout this study) with the putative 34K protein of TEV. However, decreasing the penalty for insertions allowed an alignment with 20% identity.

Presence of a potential metal-binding site

Examination of the protein sequence of the putative HC reveals the presence of a cluster of cysteine residues starting at position 309 which is also present at the same location in the analogous proteins of TEV and TVMV; it takes the form Cys-X8-Cys-X13-Cys-X4-Cys-X2-Cys. This arrangement is similar to that of several proteins known to form 'zinc fingers': Cys-X(2-4)-Cys-X(2-15)-Cys/His-X(2-4)-Cys/His. This consensus sequence for metal-binding sites represents those chosen by analogy with the sequence in TFIIIA and has been found in several nucleic acid-binding proteins (Berg, 1986; Evans & Hollenberg, 1988). In potyviruses, this cluster contains several cysteine and histidine residues which could possibly be engaged in such a structure and could lead to different binding conformations (Fig. 4). These sequences are rich in hydrophobic amino acids (Kyte & Doolittle, 1982), and secondary structure prediction (Chou & Fasman, 1978) shows that they are likely to adopt an alpha-helix configuration, all of which are features that have been observed in the 'zinc finger' motifs (Casas-Finet et al., 1988). The suggestion that HC proteins bind to nucleic acids is strengthened by the observation that they can be purified by adsorption to oligo(dT) (Thornbury et al., 1985). However, the site for nonspecific binding to oligo(dT) may be different from the metal-binding site as was shown for the outer capsid protein σ3 of reovirus (Schiff et al., 1988). It remains unclear why HC protein, which is supposed to be the aphid transmission factor, should bind nucleic acid. Thornbury et al. (1985) have shown that the active HC is a dimer, and another possibility could be that this cluster of cysteines allows the dimerization of the molecule by the common fixation of metal atoms. Such a situation has been shown for the tat protein of human immunodeficiency virus
PVY nucleotide sequence

Fig. 6. Dot plot (Staden, 1982) comparison of PVY, TVMV and TEV polyproteins. Window: 11; score: 135.

Fig. 7. Possible map of PVY polyprotein. Specific motifs are indicated. Solid bars indicate cleavage sites for \( \text{N}^\text{a} \) protease. Hatched bars indicate other possible cleavage sites.
(Frankel et al., 1988). However, the possibility remains that the protein which was purified was a maturation intermediate not only of the polypeptide needed for HC activity but also of another viral protein. The HC of cauliflower mosaic virus is a rather short protein of 18K (Armour et al., 1983) which is significantly smaller than the observed 58K for the PVY HC. A weak but significant homology has been observed between the putative C terminus of the TVMV 28K protein (40 amino acids upstream from the cysteine cluster) and the TMV 30K protein which is involved in the cell-to-cell movement of either the virus or viral RNA (Domier et al., 1987). A working hypothesis could be that this potential metal-binding motif would be part of a cell-to-cell movement protein.

Berger & Pirone (1986) have shown that the potyvirus HC is a glycoprotein and potential glycosylation sites (Doolittle, 1986) taking the form NXS/T are found at positions 450 and 744 within the putative helper sequence. A consensus sequence for nucleotide binding (Doolittle, 1986) could be found at the N terminus of the CI protein. This sequence (GAVGSGKST, from positions 1242 to 1250 on the PVY amino acid sequence) is strictly conserved among the three potyviruses. These observations are summarized in Fig. 7.

DISCUSSION

We have determined the complete nucleotide sequence of the PVY (N strain) RNA. This is the third potyvirus RNA entirely sequenced to date. Its genome is 9704 bases long compared with 9015 for TVMV and 9496 for TEV, and it can encode a polyprotein of 3063 amino acids. Comparison of the protein sequence for the three viruses shows that two-thirds of the polyprotein towards the C terminus is highly conserved, including the capsid protein, the RNA-dependent RNA polymerase, the NIa protease, the VPg and the cytoplasmic CI protein. Towards the N terminus, a region of strong homology might be the HC but its boundaries remain unclear. Two proteases are involved in the proteolytic cleavage in animal picornaviruses (Nicklin et al., 1986) and in comoviruses a second viral polypeptide has been shown to act as a cofactor of the protease and to modify its specificity (Verver et al., 1987); it is thus possible that cleavage sites other than those observed to date by protein sequencing could exist within the potyvirus polypeptide, and their characterization will lead to a better understanding of the organization of the potyvirus genome.

We thank our colleagues of INRA Centre de Versailles for helpful discussions; Anne-Lise Haenni for critical reading of the manuscript; Claude Mugnier for kind computer assistance; Josette Albouy, Constance Charbonnier, Chantal Kusiak and Yves Maury for help in sequencing. This work was supported by grant 86 0827 from MRT and AIP 4560 from INRA.

REFERENCES


DOMIER, L. L., FRANKLIN, K. M., SHAHABUDDIN, M., HELLMAN, G. M., OVERMEYER, J. M., HIREMATH, S. T., SIAW, M. F. E.,


(Received 30 August 1988)