Sequence analysis of the parsnip yellow fleck virus polyprotein: evidence of affinities with picornaviruses

A. D. Turnbull-Ross, M. A. Mayo, B. Reavy* and A. F. Murant

Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, U.K.

The 9.9 kb monopartite ssRNA genome of parsnip yellow fleck virus (PYFV) encodes a polyprotein from which the functional proteins are assumed to arise by proteolytic cleavage. The 22.5K, 26K and 31K particle proteins were mapped in the polyprotein by determining their N-terminal amino acid sequences, and were found to begin at amino acid positions 395, 589 and 811, respectively. There could be polypeptide(s) of up to 43K on the N-terminal side of the particle protein sequences. A region within the 26K particle protein has sequence similarity to the VP3 particle protein of picornaviruses. Three other regions in the PYFV polyprotein have sequence similarity to regions thought to have RNA polymerase, NTP-binding and protease functions in the polyproteins of picornaviruses, comoviruses and nepoviruses. Despite these similarities in sequence and in genome organization to viruses in the picorna-like supergroup, PYFV is distinct from all other plant and animal viruses described. This justifies placing it in a separate plant virus genus for which the name ‘sequivirus’ has been proposed.

Introduction

Parsnip yellow fleck virus (PYFV; Murant & Goold, 1968) is a plant virus with isometric particles approximately 30 nm in diameter, and is dependent on a helper virus, anthuriscus yellows virus, for transmission by aphids in a semi-persistent manner. PYFV particles have three coat protein species and one ssRNA species of 9.9 kb (Hemida & Murant, 1989), properties in which it resembles the picornaviruses of vertebrates (Murant et al., 1987). Because of its unusual biological and particle properties PYFV has been made the type member of a new plant virus group, provisionally named the parsnip yellow fleck virus group (Murant, 1988, 1991). The genus name ‘sequivirus’ has now been proposed (Murant, 1993).

PYFV isolates fall into two serotypes known as the parsnip and anthuriscus serotypes. The complete nucleotide sequence of isolate P-121 (parsnip serotype) has been determined (Turnbull-Ross et al., 1992) and shown to have a single large open reading frame which could encode a polyprotein of 3027 amino acids. Protease and polymerase motifs were identified in the C-terminal half of the polyprotein, and screening with PYFV polyclonal antiserum indicated that the particle proteins are in the N-terminal half. The PYFV genome arrangement is therefore similar to that of animal picornaviruses. In this paper we report the N-terminal sequences of the PYFV particle proteins and map their positions in the polyprotein. We also present a sequence comparison of the PYFV polyprotein with those of picornaviruses, comoviruses, nepoviruses and potyviruses.

Methods

Determination of the N-terminal sequences of the capsid proteins. PYFV isolate P-121 was grown in spinach plants and purified from infected leaves as described previously (Hemida & Murant, 1989). The purified particles were then centrifuged through a solution of CsCl in 6mm-sodium phosphate pH 7.0 (starting density 1.1 g/ml) at 40000 rpm (149000 g) for 16 h in a Beckman SW50.1 rotor, and the pellet was then resuspended in 6 mm-sodium phosphate pH 7.0. The particle proteins were then separated in a 12% (w/v) polyacrylamide gel (Laemmli, 1970) and electro-blotted onto ProBlott membrane (Applied Biosystems). The proteins were isolated and individually sequenced employing an Applied Biosystems 470/120 Gas Phase Sequencer using the 03CPTh program as modified by Speicher (1989).

Computer analysis. The PYFV polyprotein sequence, derived from the nucleic acid sequence of the RNA (Turnbull-Ross et al., 1992), was used to search the OWL composite database (Akrigg et al., 1992). Multiple alignments were performed with PILEUP (Devereux et al., 1984) and CLUSTAL V (Higgins et al., 1992). Tentative phylogenetic trees were constructed by neighbour-joining (Saitou & Nei, 1987) and sampled by bootstrapping (Felsenstein, 1985). The alignments used for generating the trees contained at least one potyvirus sequence to act as an outlying virus.
Results

Location of particle proteins in the PYFV polypeptide

Particles of PYFV isolate P-121 contain three distinct proteins of 22.5K, 26K and 31K (Hemida & Murant, 1989). The proteins of purified PYFV particles were separated and subjected to Edman degradation and sequencing. The 10 amino-terminal residues determined for each protein were PSFSQTL(P or L)L1 (22.5K protein), ATLPSAVGDN (26K protein) and ASPVLSQVDF (31K protein). These sequences (with P at position 8 for the 22.5K protein) occur in the polyprotein, starting at positions 395 (22.5K protein), 589 (26K protein) and 811 (31K protein). The calculated Mr values of the 22.5K and 26K proteins are therefore 22564 (194 amino acids) and 23695 (222 amino acids), respectively.

Comparison of the PYFV polyprotein with other viral proteins

The PYFV polyprotein sequence was used to search the OWL composite database (Akrigg et al., 1988) for similarity to other viral sequences. Four regions of statistically significant sequence similarity were detected. Three of these (a part of the particle protein region, an NTP-binding motif and an RNA polymerase motif) were illustrated earlier (Turnbull-Ross et al., 1992) as dot matrix comparisons of the PYFV polyprotein with those of cowpea mosaic comovirus (CPMV; Lomonossoff & Shanks, 1983; Van Wezenbeek et al., 1983), tomato black ring nepovirus (TBRV; Meyer et al., 1986; Greif et al., 1988), human rhinovirus 14 (HRV14; Stanway et al., 1984) and hepatitis A virus (HAV; Najarian et al., 1985). The fourth region was similar in sequence to viral cysteine proteases (see below).

The PYFV 26K particle protein

The portion of the PYFV 26K particle protein sequence between amino acids 653 and 798 was similar to sequences in the VP3 proteins of viruses in the cardiovirus and rhinovirus genera of the Picornaviridae (Fig. 1). The greatest sequence identity was 31.9% with VP3 of encephalomyocarditis virus (EMCV; Bae et al., 1989). The sequence similarity was greatest to the picornaviral dfl, eft, gfl and Ifl regions of VP3 (Rossmann et al., 1985).

A putative NTP-binding domain

Sequences in the PYFV polyprotein between amino acids 1455 and 1595 resembled an NTP-binding domain (Hodgman, 1988; Gorbalenya et al., 1988; Gorbalenya & Koonin, 1989). Fig. 2 shows a multiple alignment of this region with the putative NTP-binding domains of poliovirus (PV; Nomoto et al., 1982), HAV, foot-and-mouth disease virus (FMDV; Robertson et al., 1985), CPMV and TBRV. Of these domains the most closely related to the PYFV sequence is that of the 2C protein from FMDV which has 33.8% sequence identity. The other domains have sequence identities with the PYFV sequence of 32.8% (PV), 30.9% (HAV), 29.9% (CPMV) and 29.3% (TBRV). Comparison with the tobacco etch virus (TEV; Allison et al., 1986) NTP-binding protein showed only 17.7% sequence identity. Prominent
Sequence analysis of PYFV polyprotein

---

**Fig. 2.** Multiple alignment of conserved regions of the putative NTP-binding proteins of PV (Nomoto et al., 1982), HAV (Najarian et al., 1985), FMDV (Robertson et al., 1985), PYFV, CPMV (Van Wezenbeek et al., 1983) and TBRV (Greif et al., 1988). The letters below the sequences indicate the conserved (con) residues. Residues conserved in at least four of the viruses are in reverse contrast. Numbers on the left refer to the amino acid position in the polyprotein.

**Fig. 3.** Alignment of viral cysteine protease sequences of PV, HAV, PYFV, CPMV and TBRV. The figures above the sequences indicate the number of amino acids from the PV protease N terminus. Residues identical to those in the PV protease sequence are in reverse contrast. The letters below the sequences indicate the residues that are conserved in all five viruses. Residues reported to be involved in substrate binding are marked with an asterisk. Numbers on the left refer to the amino acid position in each polyprotein.

---

features of the alignment were the well conserved GXGKS and DD motifs (Gorbalenya et al., 1988, 1989a). A GYX_2Q motif and a more disperse MAXLX_2KX_2FXSXTN motif, both previously described by Argos et al. (1984), were also identified.

A putative protease domain

A region (amino acids 2028 to 2178) of the PYFV polyprotein was similar in sequence to picornaviral 3C proteases (Palmenberg, 1990), the CPMV B-RNA-encoded 24K protein (Verver et al., 1987), the TBRV RNA-1-encoded 23K protein (Greif et al., 1988) and the TEV Nla protein (Dougherty & Parks, 1991). The alignment (Fig. 3), based on that of Gorbalenya et al. (1989b), allowed us to identify candidates in PYFV for the conserved histidine (PV protease position 40) and cysteine (PV protease position 147) residues of the catalytic triad. The alignment around the conserved aspartate/glutamate residue (PV protease position 71) was less convincing. We have placed PYFV glutamate 2066 at the conserved site but aspartate 2068 could equally well be the third residue of the catalytic triad. Two of the amino acids believed to be involved in substrate recognition (PV protease positions 159 and 161; Bazan & Fletterick, 1988) differed in PYFV from the consensus found in other viruses. A cysteine residue occurred at position 159 in place of glycine and a leucine residue occurred at position 161 in place of histidine; a leucine for histidine replacement has previously been observed in TBRV RNA-1 polyprotein (Greif et al., 1988).

A putative RNA-dependent RNA polymerase

Koonin (1991) proposed eight conserved motifs in RNA-dependent RNA polymerases and these, except motif VIII, were detected in a multiple alignment of the
polyproteins of PYFV and members of the picornavirus, comovirus and nepovirus groups (Fig. 4). An additional motif was identified in this limited alignment, with the consensus (T/S)SXG(Y/F)P (Fig. 4; Argos et al., 1984). For the 280 amino acids compared, the PYFV and CPMV sequences were the most closely related, with 39.7% sequence identity. The identities of the PYFV sequence to those of the other viruses in Fig. 4 were 29.4% (PV), 30.9% (HAV) and 34.1% (TBRV).

Attempts to detect phylogenetic relationships between PYFV and picorna-like viruses

Tentative phylogenetic trees constructed from the results of pair-wise sequence comparisons yielded unconvincing results. For example, although comparisons among RNA polymerase domains suggested that PYFV is related more closely to comoviruses and nepoviruses than to other viruses, bootstrapping tests did not establish this as statistically significant. When the NTP-binding domains were compared, PYFV grouped with the picornaviruses, but when sequences around the domains were included in the comparisons PYFV grouped with the comoviruses and nepoviruses.

Discussion

The finding that the N termini of the PYFV particle proteins map to the N-terminal half of the polyprotein confirms the location previously inferred from polyclonal antibody screening of PYFV cDNA clones expressing PYFV-β-galactosidase fusion proteins (Turnbull-Ross et al., 1992). The most N-terminal PYFV particle protein (22.5K) begins at amino acid 395, and there could therefore be polypeptide(s) of up to 43K on the N-terminal side of the particle proteins (amino acids 1 to 394). This region has no obvious sequence similarity with picornaviral leader peptides, the N-terminal 58K/48K proteins encoded by CPMV M-RNA, the N-terminal 50K and 46K proteins encoded by TBRV RNA-2, the N-terminal 35K and HC-pro proteins of TEV, or the tobacco mosaic virus 30K movement protein (Meshi et al., 1987).

The PYFV 26K particle protein (amino acids 589 to 810) is similar in sequence to the central region of VP3 from animal picornaviruses (Fig. 1). The sequence similarity suggests that the PYFV 26K protein has a three-dimensional structure like that of VP3 of picornaviruses.
The PYFV 31K particle protein begins at amino acid 811 and presumably arises by cleavage at QA. The exact carboxy terminus is unknown but if it also results from cleavage at QA (amino acids 1069 to 1070) it would have an Mr of 28,975. The 31K protein contains an RGD motif (amino acids 870 to 872). RGD motifs are found in many extracellular matrix proteins (Ruoslahti & Pierschbacher, 1987; Ruoslahti, 1988) that interact with integrins (Hynes, 1987), and in VP1 of FMDV (Geysen et al., 1985) where it is involved in cell attachment of the virus (Fox et al., 1989). A conserved RGD is also found in several strains of coxsackievirus A9 (Chang et al., 1992). It is tempting to speculate that the RGD motif in PYFV may be involved in some form of protein–protein interaction, perhaps associated with aphid transmission, viral movement or compartmentalization.

We based our alignment of PYFV protease domain (Fig. 3) on that of Gorbalenya et al. (1989b) which has been corroborated by site-directed mutagenesis of the proteases from PV (Kean et al., 1991; Hämmerle et al., 1991), CPMV (Dessens & Lomonossoff, 1991) and TEV (Dougherty et al., 1989). The N-terminal sequences of the particle proteins indicate three possible protease cleavage sites in the polyprotein, SS/PS, AN/AT and AQ/AS; Q/A is a site cleaved by other viral cysteine proteases (Palmenberg, 1990). However, the possibility cannot be excluded that some or all of the primary cleavage sites are elsewhere and that the observed N termini result from secondary proteolytic degradation.

Comparisons of the amino acid sequences of the proteins of many positive-strand RNA viruses have revealed the existence of conserved domains for NTP-binding proteins and RNA polymerases (Argos et al., 1984; Gorbalenya & Koonin, 1989; Kamer & Argos, 1984; Poch et al., 1989; Koonin, 1991), and on this basis Goldbach (1986, 1987) proposed that most viruses are in one of two supergroups. Our amino acid sequence comparisons (Fig. 1, 2 and 4) show that the arrangement of the NTP-binding, protease and polymerase motifs in the PYFV polyprotein is similar to that in several viruses in the ‘picorna-like’ supergroup (Fig. 5), which includes animal picornaviruses and plant comoviruses, nepoviruses and potyviruses (Kamer & Argos, 1984; Argos et al., 1984; Goldbach, 1986, 1987; Candresse et al., 1990; Koonin, 1991). However, it was not possible to detect a close relationship between PYFV and other viruses in the picorna-like supergroup, and this justifies the placement of PYFV in a separate taxonomic group of plant viruses along with the anthriscus serotype of PYFV and, probably, dandelion yellow mosaic virus (Murant, 1988). The name ‘sequivirus’ has been proposed for this group (Murant, 1993).

We are grateful to D.G. Campbell of the MRC Protein Phosphorylation Unit, Department of Biochemistry, University of Dundee, U.K. for protein separation and amino-terminal sequencing work. We also acknowledge use of SEQNET facilities at the SERC Daresbury laboratory. This work was supported by a grant from the Scottish Office Agriculture and Fisheries Department under the Increased Flexibility Scheme.

References


(Received 5 October 1992; Accepted 24 November 1992)