Nucleotide sequence of tomato ringspot virus RNA1

Michael E. Rott, 1* Angus Gilchrist, 2 Lawrence Lee 2 and D’Ann Rochon 2*

1 Department of Plant Science, University of British Columbia, Vancouver, British Columbia V6T 1Z4 and
2 Agriculture and Agri-Food Canada, 6660 NW Marine Drive, Vancouver, British Columbia V6T 1X2, Canada

The nucleotide sequence of tomato ringspot nepovirus (TomRSV) RNA1 has been determined. TomRSV RNA1 is 8214 nucleotides in length, excluding the 3' poly(A) tail, and contains a single long open reading frame (ORF) of 6591 nucleotides beginning at the first AUG codon at nucleotide position 78. This ORF accounts for 80% of the RNA1 sequence and would give rise to a polyprotein with a predicted molecular mass of 244 kDa. Amino acid sequence comparisons between portions of the TomRSV RNA1-encoded polyprotein and proteins encoded by several members of the picornavirus superfamily have provided information concerning the genomic organization and putative functions of TomRSV-encoded proteins. The putative TomRSV protease retains a conserved histidine residue present in the proteases encoded by members of the como-, poty- and poliovirus groups which is thought to be involved in dipeptide cleavage site recognition. Interestingly, this histidine residue is replaced by a leucine in the proteases of other sequenced nepoviruses. This suggests that the TomRSV protease shares dipeptide cleavage site specificity with that of como-, poty- and picornaviruses rather than the other nepoviruses.

* Author for correspondence. Fax +1 604 666 4994. e-mail ROCHON@PARGVA.AGR.CA
† Present address: Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.

The nucleotide sequence data reported in this paper have been deposited in GenBank and assigned the accession number L19655.

Tomato ringspot virus (TomRSV), a member of the nepovirus group, is a 28 nm spherical virus. The bipartite genome is plus sense RNA which contains a VPg at the 5' termini and a poly(A) tail at the 3' termini. Expression of the TomRSV genome likely occurs through the production of a long polyprotein which is subsequently cleaved by a viral-encoded protease. Nepoviruses have been included in a larger picornavirus-like supergroup together with the plant como- and potyviruses and the animal picornaviruses (Goldbach, 1987).

The complete nucleotide sequence of TomRSV RNA2 has been determined and was shown to code for the coat protein (Rott et al., 1991a), a putative cell-to-cell movement protein (Wieczorek & Sanfacon, 1993) and an N-terminal polyprotein(s) of unknown function(s). The sequences of approximately 1·1 and 1·5 kb at the 5' and 3' termini, respectively, of TomRSV RNA1 have been previously described and show near perfect nucleotide sequence identity with corresponding sequences at the 5' and 3' termini of RNA2 (Rott et al., 1991b).

Preparation of TomRSV RNA1 cDNA clones has been described previously (Rott et al., 1988, 1991b). Cloning, sequencing and sequence analysis methods were as described previously (Rott et al., 1991a, b).

The complete nucleotide sequence of TomRSV RNA1 is shown in Fig. 1. RNA1 is 8214 nucleotides in length excluding the 3' poly(A) tail and contains one long open reading frame (ORF) initiating at the first AUG codon (nucleotide 78) and terminating at a UAA stop codon (nucleotide 6669). The polyprotein encoded by RNA1 has a predicted molecular mass of 244 kDa and accounts for approximately 80% of the RNA1 coding capacity. As previously noted, it is not known whether initiation of protein synthesis occurs at AUG 78 or at the next in-frame AUG codon at position 441 (see Rott et al., 1991a, b).

The predicted polyprotein sequence encoded by TomRSV RNA1 was examined for motifs characteristic of putative protease cofactors, NTP-binding proteins, viral cysteine proteases and RNA-dependent RNA polymerases. Fig. 2 aligns the motifs identified in the TomRSV polyprotein with those present in the polyproteins encoded by RNA1 of the nepoviruses tomato black ring virus (TBRV; Greif et al., 1988), grapevine chrome mosaic virus (GCMV; Le Gall et al., 1989), and grapevine fanleaf virus (GFLV; Ritzenthaler et al., 1991), as well as B RNA of the comovirus cowpea mosaic virus (CPMV; Lomonossoff & Shanks, 1983),
Fig. 1. For legend see page 468.
Short communication

Fig. 1. For legend see page 468.
and the genomic RNAs of the potyvirus tobacco etch virus (TEV; Allison et al., 1986) and poliovirus (Racaniello & Baltimore, 1981).

A conserved amino acid sequence, $F-x_{27}W-x_{14}L-x_{21}L$, $L_xE$ ($x_e$ refers to the number of amino acids between conserved residues) is located near the N-terminal region of the TomRSV RNA1 polyprotein beginning at amino acid residue 482 where the second conserved L residue is replaced by an F in TomRSV (Fig. 2a). These conserved amino acids were previously identified in other nepo- and comovirus polyprotein sequences (Ritzenthaler et al., 1991) and a protease cofactor function was suggested (Ritzenthaler et al., 1991). The CPMV 24K protein contains this sequence and has been demonstrated to function as a cofactor for the CPMV 24K protease (Vos et al., 1986; Peters et al., 1986). The TomRSV RNA1 polyprotein sequence beginning at amino acid residue 791 includes the ‘A’ and ‘B’ motifs...
characteristic of NTP-binding proteins (Fig. 2b) (Gorbalenya & Koonin, 1989; Gorbalenya et al., 1989a). The highly conserved 'A' and 'B' site motifs are GxxG(x)GKS/T and D/ED, respectively, and can be widely separated (x refers to any amino acid; the residue in brackets is not necessarily conserved in all cases). Global similarity between amino acid sequences surrounding the NTP-binding domain of TomRSV and the other viruses is very low, making it difficult to align these regions except at the conserved motifs. The TomRSV sequence contains the hydrophobic residues LLVLAAVILILFT C-terminal to the NTP-binding domain (amino acid residues 1169–1182) which is predicted to be a transmembrane domain by the method of Argos (1984). The como-, poty- and picornavirus proteins which contain the NTP-binding domain also contain a very hydrophobic putative transmembrane spanning sequence at the C terminus. This sequence is thought to be important for anchoring the replication complex to the lipid membrane (Dorssers et al., 1984; Goldbach & van Kammen, 1985, and references therein; Takeda et al., 1986).

Viruses which use a polyprotein strategy encode the proteolytic enzymes required for polyprotein maturation. A region N-terminal to the putative TomRSV RDRP and beginning at amino acid residue 1283 contains a motif (H-x4-E-x10-CG-x8-GxxxGxxHxxG) characteristic of known viral cysteine proteases in which the residues H, E and C (italicized) form the putative catalytic triad of the enzyme (Bazan & Fletterick, 1989; Gorbalenya et al., 1989b; Hammerle et al., 1991; Dessens & Lomonossoff, 1992; Margis & Pinck, 1992) (Fig. 2c). Interestingly, the TomRSV protease domain contains an H residue (underlined) at amino acid position 1451 which is conserved among the como-, poty- and picornaviruses but not the other nepoviruses. Como-, poty- and picornavirus proteases preferentially cleave at only a few common dipeptide sites (Q/M, Q/S, Q/G, E/S and E/G; see Hellen et al., 1989) and it has been suggested that for the picornaviruses, the conserved H residue within the protease domain may be important for cleavage site recognition (Bazan & Fletterick, 1988). It has also been suggested that cleavage by proteases encoded by the nepoviruses TBRV, GCMV and GFLV, which have a different dipeptide specificity, may be due to replacement of the H residue with L in these proteins (Bazan & Fletterick, 1988; Demangeat et al., 1990; Ritzenthaler et al., 1991). The presence of this H residue in the TomRSV protease suggests that the cleavage sites for maturation of the TomRSV polyprotein may be similar to those of como-, poty- and picornaviruses (see below). Pairwise alignments (not shown) of the amino acid sequences of the putative proteases of TomRSV and the other viruses shown in Fig. 2(c) show that TomRSV shares a greater degree of amino acid sequence identity with GFLV (32%) than with the proteases of the other viruses compared (21–26%).

The C-terminal region of the TomRSV RNA1-encoded polyprotein contains sequences characteristic of known and putative RNA-dependent RNA polymerases (RDRP) (Argos, 1988) (Fig. 2d). The amino acid sequences surrounding the TomRSV RDRP motif could be aligned with varying degrees of success with the other RDRP-containing proteins. The best match occurred with GFLV which was 39% over 691 amino acids and was followed by 37%, 39% and 40% similarity over 476, 462 and 492 amino acids with TBRV, GCMV and CPMV, respectively. In the three latter alignments, similarity near the 3' terminus is not apparent. Only 25% and 23% similarity was detected with TEV and poliovirus over a reduced span of 191 and 260 amino acids, respectively. TEV and poliovirus encode smaller RDRPs with smaller C-terminal regions compared with the TomRSV putative RDRP (Fig. 3).

Sequences beginning at amino acid residue 138 of the TomRSV RNA1 polyprotein could be aligned with the N-terminal region of the TBRV and GCMV RNA1 polyproteins but not with the N-terminal region of the GFLV polyprotein (Fig. 2e). The function of this region of the polyprotein is unknown.

In polio, CPMV and GFLV, the amino acid sequence for the small 5' genome linked protein (VPg) is located between the NTP-binding and protease proteins (Racaniello & Baltimore, 1988; Goldbach & Rezelman, 1983; Zabel et al., 1984; Pinck et al., 1991). Due to the small size of the VPg and its low degree of conservation the sequence of the VPg cannot be stated with certainty and is only tentatively located as described below and in Fig. 3.

As described above, the putative TomRSV protease may have a dipeptide cleavage site specificity which resembles that of como-, poty- and picornaviruses rather than other nepoviruses. Consequently, assignment of cleavage sites for delineating protein-coding regions was based on the known cleavage sites (Q/S, Q/M, Q/G, E/G and E/S) commonly used for maturation of como-, poty- and picornavirus polyproteins (Hellen et al., 1989; Palmenberg, 1990; Wellink et al., 1986). The locations of these sites were then compared with the locations of known cleavage sites for CPMV, as well as the putative sites for the RNA1-encoded polyproteins of the nepoviruses TBRV, GCMV and GFLV (Pinck et al., 1991; Ritzenthaler et al., 1991; Le Gall et al., 1989; Greif et al., 1988; Margis et al., 1994). Fig. 3 shows the proposed genomic organization and cleavage sites for the TomRSV RNA1 polyprotein compared with those previously described for CPMV B RNA (Wellink et al., 1986). A likely cleavage site between the TomRSV
Short communication

(a) 

TomRSV 482 FODITI--X13--IEELWRSLFLAFA--X11--L--X21--FAE
TBRV 463 FREDI--X13--IVMMIKKVVLH--X11--L--X21--LLE
GCMV 462 FREAL--X13--VEVLIAVVKSH--X11--L--X21--LLE
GFLV 471 FDUVTM--X13--LKKWEKLSLEF--X11--L--X21--LVE
CPMV 192 FEKVM--X13--LQLWLDIVGQ--X11--L--X21--LVE

Consensus F W L L E

(b) 

TomRSV 791 WYLVGPGCQLSLFASGFSN--X29--AICCVDDLSCE
TBRV 776 WYLFQGHCQSNFMATLND--X28--TFFHVDLSQV
GCMV 758 WYIAGPSHCQSFNMVDLGM--X28--TIMEIDDLSIK
GFLV 776 WYIFGASQXTTIANIII--X30--ACVKVDFAIE
CPMV 489 TIFPOQRTQKSLSMQYTK--X29--PFVLMDDFAAVV
TEV 1342 DLLVRQAVSKSTCGLPHLS--X68--DFVVIDCVDVND
Polio 1250 CLLVQHSGSPGTKSVLNLAR--X27--GVVIMDDLGNQP

Consensus G G GKS DD

"A" site "B" site

(c) 

TomRSV 1,283 X24--X46--E--X96--NSPEDCGALLVAHLEGQYKIIGMVAG
TBRV 1,270 X24--X38--E--X86--SRNEDCMIIIQIKSMRVVMVAG
GCMV 1,256 X24--X38--E--X86--SRNEDCGMLLLQIKSMRVVMVAG
GFLV 1,284 X24--X44--E--X91--AKYGEDALAVPIGSPKVIAMVC
CPMV 987 X24--X35--E--X86--TIPEPQGLVIANIQRKHKIVVAG
TEV 2,083 X24--X44--E--X64--TKVQGSPPLVSTRQ--FIVQSAS
Polio 1,607 X24--X30--E--X71--TRACGGQVITC--R--KVQIMVAG

Consensus H E CG G G H G

(d) 

TomRSV RNAI 138 LCLSYKSGVSSPPPMTQRQQFAIARKRLVLQKGOQIQIIREL--IRRARKAAKAYAAFAARKKA
TomRSV RNA2 138 LLRCCK--GEPPPQPPPQTPQQFAIARKRLVLQKGOQIQIIREH--IRRARKAYAAKIAAIKKAA
TBRV RNA 189 KLNNKARLGAHRSVARAVQAKARQVLEFEPSQQIPGQAELAQIFADRLRSKAYALTAVRAK
GCMV RNA 189 KLTKANAGLAHRSVATAQAKARVLEFEPSPAHIQIAVKAHIFARKLRSKYYALTAQVRAR

Consensus D D G T N GDD

Fig. 2. For legend see opposite.
Short communication

Fig. 3. Proposed genomic organization of TomRSV RNA1 and comparison with other members of the picornavirus superfamily. Lines indicate noncoding sequence and the bars represent the polyprotein sequence encoded by the long ORFs. Vertical lines through the bars indicate known and putative protease cleavage sites (see below). Dashed lines indicate additional cleavage sites possible in this region. The conserved amino acid sequences are indicated by the similarly shaded boxes. Abbreviations are: coat protein (CP), NTP-binding domain (NTP), RNA-dependent RNA polymerase (RDRP), helper component (HC) and poly(A) tail (AAA) and designate putative functions for the predicted proteins. The TomRSV polyprotein sequence was scanned for the dipeptides E/S, E/G, Q/G, Q/M and Q/S, which are common cleavage recognition sites within como-, poty- and picornavirus polyproteins. Potential protease cleavage sites in the TomRSV RNA1 polyprotein are shown. Cleavage sites of the CPMV B RNA polyprotein are shown as a comparison and aided in the identification of the TomRSV cleavage sites. The sizes of the cleaved products of TomRSV, GFLV and CPMV are indicated along with their known and putative functions.

Fig. 2. Alignment of portions of the TomRSV-encoded polyprotein with motifs identifying the putative protease cofactor, NTP-binding protein, cysteine protease, RDRP and N-terminal coding region of several members of the picornavirus superfamily. (a) Protease cofactor. The region containing the protease cofactor motif of TomRSV is shown aligned with that of nepoviruses and CPMV. (b) NTP-binding domain. Asterisks indicate identical amino acid residues in at least four of seven sequences. (c) Cysteine protease. Asterisks indicate identical amino acid residues in at least four of seven sequences. (d) RDRP. Asterisks indicate identical amino acid residues in at least four of seven sequences. (e) N-terminal region of TomRSV polyprotein. Asterisks indicate identical residues in all four sequences and a caret indicates identical amino acid residues in three of four sequences. Dashes are spaces inserted into the sequence to maximize alignment. In all figures, X refers to the number of amino acids separating the conserved residues or sites. Underlined residues are the most highly conserved as determined from these alignments, and numbers to the left of the sequence refer to amino acid residue position in each viral polyprotein.
putative protease and the RDRP domain could occur at the Q/M dipeptide located at amino acid position 1465–1466. This site aligns with the Q/G cleavage site of CPMV. Three potential cleavage sites were found between the NTP-binding and protease domains. It is possible that cleavage at the protease side could occur at either the Q/S site located at position 1236–1237 or at the Q/G site at position 1239–1240. Another potential cleavage site at the NTP-binding domain site occurs at the Q/S site at position 1212–1213. The region between the two Q/S sites is 24 amino acids in length. This is identical in size to the GFLV VPg protein (Pinck et al., 1991) and only four amino acid residues shorter than the CPMV VPg protein (Goldbach & Rezelman, 1983). However, proper identification of this region as encoding the TomRSV VPg cannot be confirmed since the proteins are too short to be aligned with certainty. A likely cleavage site between the putative protease cofactor and the NTP-binding domains is the Q/G site at position 620–621. This site corresponds closely with the Q/S site of CPMV. Additional N-terminal cleavage sites are possible but cannot be predicted due to limited amino acid sequence similarities with the proteins of the other viruses compared.

A comparison of the genomic organization of TomRSV RNA1 with those of TBRV, GFLV, CPMV, TEV and poliovirus is shown in Fig. 3. The comparison demonstrates the conservation of sequence and relative order of proteins encoded by TomRSV and the other members of the picornavirus-like supergroup. The TomRSV genome contains a number of unusual features when compared to that of other nepoviruses. TomRSV has a relatively large genome of over 15000 bases (RNA1 and RNA2 combined), due in part to the unusually large size of the 3′ noncoding regions (Rott et al., 1991b). In addition, the 5′ noncoding regions are identical and consequently there is extensive amino acid sequence identity in the N-terminal regions of the polyproteins encoded by RNA1 and RNA2. The many unique features suggest that TomRSV be considered a member of a distinct subgroup of nepoviruses as previously suggested by Martelli (1975). Partial sequence data from the cherry leafroll virus genome (Scott et al., 1992) indicate that it may also belong to this subgroup. Sequence analysis of peach rosette mosaic and myrobalan latent ringspot virus may indicate that all of these viruses belong to a distinct nepovirus subgroup.

This work was partially supported by NSERC operating grant number OGP0043840. We thank Fabienne Hans for her excellent critical review of this manuscript.

References


(Received 18 April 1994; Accepted 19 October 1994)