Journal of General Virology (1995), **76**, 465–473. *Printed in Great Britain*

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.

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.*, 1991*a*), 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.*, 1991*b*).

* 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.

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

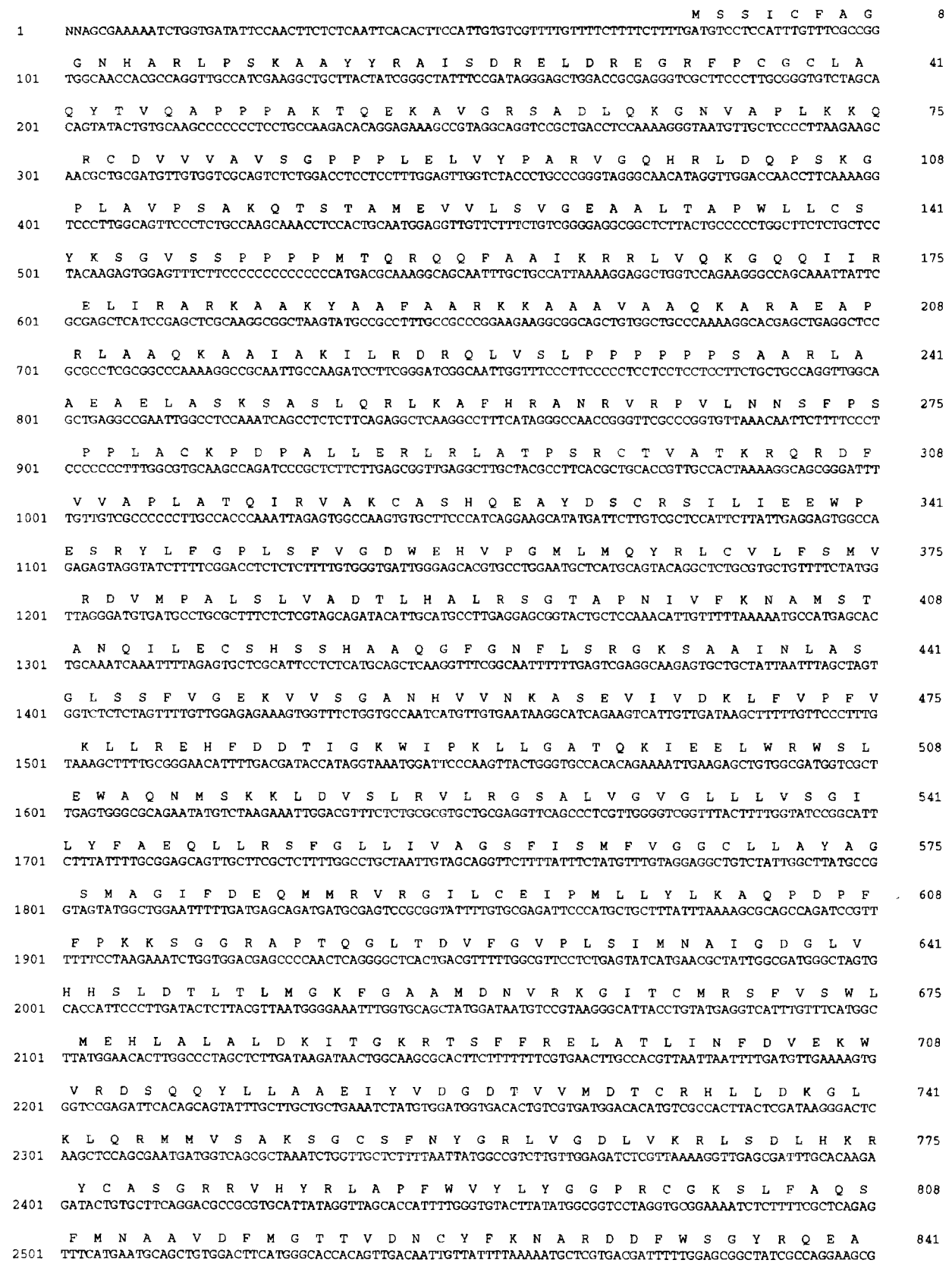

Fig. 1. For legend see page 468.

```
      I  C  C  V  D  D  L  S  S  C  E  T  Q  P  S  I  E  S  E  F  I  Q  L  I  T  T  M  R  Y  G  L  N  M  A   875
2601  ATATGCTGCGTTGATGATCTCTCCTCCTGCGAAACGCAACCCTCCATTGAGTCGGAATTCATTCAATTGATAACGACAATGAGATATGGATTAAATATGG

      G  V  E  E  K  G  A  S  F  D  S  K  M  V  I  T  T  S  N  F  F  T  A  P  T  T  A  K  I  A  S  K  A   908
2701  CAGGAGTTGAGGAAAAAGGAGCCTCATTTGATTCGAAGATGGTTATCACAACATCTAATTTTTTCACGGCTCCAACTACTGCTAAGATTGCTAGCAAAGC

      A  Y  N  D  R  R  H  A  C  I  L  V  Q  R  K  E  G  V  A  Y  N  P  S  D  P  A  A  A  A  E  A  M  F   941
2801  TGCCTACAACGATAGGCGTCACGCTTGCATTCTTGTTCAAAGAAAAGAAGGGGTTGCTTACAACCCAAGTGATCCTGCTGCTGCTGCGGAAGCGATGTTT

      V  D  S  T  T  Q  H  P  L  S  E  W  M  S  M  Q  E  L  S  A  E  L  L  L  R  Y  Q  Q  H  R  E  A  Q  H   975
2901  GTTGATAGTACTACTCAGCATCCGCTTTCCGAGTGGATGAGCATGCAGGAATTAAGTGCTGAGTTGTTGCTGCGTTACCAACAGCATAGGGAGGCTCAGC

      A  E  Y  S  Y  W  K  S  T  S  R  T  S  H  D  V  F  D  I  L  Q  K  C  V  N  G  D  T  Q  W  L  S  L  1008
3001  ATGCAGAATATAGCTATTGGAAATCCACTTCGCGCACTTCTCATGATGTTTTTGACATCTTGCAGAAGTGCGTGAATGGGGATACCCAGTGGCTATCACT

      P  V  D  V  I  P  P  S  I  R  Q  K  H  K  G  N  R  V  F  A  I  D  G  R  I  F  M  F  D  Y  M  T  L  1041
3101  TCCCGTTGACGTTATCCCTCCGTCTATTAGGCAGAAGCACAAGGGCAACCGAGTCTTCGCTATTGATGGAAGGATTTTTATGTTTGATTATATGACCCTA

      E  Y  D  E  I  K  E  K  E  N  L  D  A  R  H  L  E  A  R  I  L  E  K  Y  G  D  T  R  L  L  L  E  K  W  1075
3201  GAGTACGATGAAATCAAGGAAAAAGAGAATCTGGATGCTCGTCATCTGGAAGCTCGAATCCTTGAAAAGTACGGTGACACCCGCTTGCTTTTAGAAAAGT

      G  A  N  G  V  V  A  Q  F  I  E  Q  L  L  E  G  P  S  N  V  A  S  L  E  V  L  S  K  D  S  L  E  S  1108
3301  GGGGTGCCAATGGAGTTGTTGCGCAATTTATTGAGCAACTTCTTGAGGGTCCTTCTAACGTTGCCTCCTTGGAGGTTTTATCTAAGGACTCCCTCGAGAG

      H  K  E  F  F  S  T  L  G  L  I  E  R  A  T  L  R  A  V  Q  K  K  I  D  A  A  R  E  D  L  M  H  L  1141
3401  TCACAAGGAATTTTTTTCTACCTTGGGACTTATCGAGAGAGCTACCTTGCGTGCTGTGCAGAAGAAAATTGATGCCGCGCGTGAGGATTTGATGCATTTG

      S  G  L  K  P  G  R  S  L  T  E  L  F  V  E  A  Y  D  W  V  Y  A  N  G  G  K  L  L  L  V  L  A  A  V  1175
3501  TCTGGTTTGAAACCAGGGCGCTCACTTACAGAATTGTTCGTTGAAGCGTATGACTGGGTTTACGCCAACGGTGGTAAGCTCCTTTTAGTGCTTGCTGCCG

      I  L  I  L  F  F  G  S  A  C  I  K  L  M  Q  A  I  F  C  G  A  A  G  G  T  V  S  M  A  A  V  G  K  1208
3601  TAATTTTGATTTTATTCTTTGGGTCTGCTTGTATAAAGTTGATGCAGGCCATTTTTTGTGGTGCCGCAGGTGGTACTGTCAGTATGGCTGCTGTCGGGAA

      M  T  V  Q  S  T  I  P  S  G  S  Y  A  D  V  Y  N  A  R  N  M  T  R  V  F  R  P  Q  S  V  Q  G  S  1241
3701  AATGACCGTTCAATCGACGATTCCCTCCGGTAGTTATGCAGACGTGTACAATGCGCGCAACATGACACGCGTTTTCCGCCCACAATCTGTACAGGGTTCT

      S  L  A  E  A  Q  F  N  E  S  H  A  V  N  M  L  V  R  I  D  L  P  D  G  N  I  I  S  A  C  R  F  R  G  1275
3801  TCTTTGGCGGAAGCGCAATTTAATGAATCGCACGCTGTGAATATGTTAGTGCGAATTGATTTACCTGATGGCAACATTATTTCTGCCTGCAGGTTTCGCG

      K  S  L  A  L  T  K  H  Q  A  L  T  I  P  P  G  A  K  I  H  I  V  Y  T  D  N  N  G  N  T  K  A  P  1308
3901  GAAAGTCTTTGGCTTTGACTAAACATCAGGCCTTAACCATACCGCCAGGTGCTAAAATCCATATTGTATATACTGACAACAATGGAAATACCAAAGCACC

      L  T  H  F  F  Q  P  T  G  P  N  G  E  H  F  L  R  F  F  N  G  T  E  V  C  I  Y  S  H  P  Q  L  S  1341
4001  GCTGACTCATTTTTTCCAACCTACTGGACCCAATGGAGAACATTTTTTGAGATTCTTCAACGGGACAGAGGTTTGTATTTATTCCCACCCTCAACTTTCA

      A  L  P  G  A  P  Q  N  Y  F  L  K  D  V  E  K  I  S  G  D  I  A  I  K  G  C  G  I  K  L  G  R  T  S  1375
4101  GCTTTGCCTGGCGCTCCACAAAATTATTTCTTGAAAGATGTGGAAAAAATATCTGGTGACATAGCCATTAAAGGTTGTGGCATCAAGCTAGGTAGAACCA

      V  G  E  C  V  G  V  K  D  N  E  P  V  L  N  H  W  R  A  V  A  K  V  R  T  T  K  I  T  I  D  N  Y  1408
4201  GCGTTGGCGAGTGTGTTGGTGTTAAGGACAATGAACCCGTCTTAAATCACTGGCGCGCTGTCGCGAAGGTCCGCACCACCAAGATCACTATCGATAATTA

      S  E  G  G  D  Y  S  N  D  L  P  T  S  I  I  S  E  Y  V  N  S  P  E  D  C  G  A  L  L  V  A  H  L  1441
4301  TTCAGAGGGTGGTGATTATTCCAATGATCTTCCTACGTCCATCATCTCTGAGTACGTAAATTCACCAGAAGATTGTGGCGCGCTTTTAGTCGCCCATCTT

      E  G  G  Y  K  I  I  G  M  H  V  A  G  S  S  Y  P  V  E  V  D  G  V  Q  M  P  R  Y  I  S  H  A  S  F  1475
4401  GAAGGTGGTTACAAAATCATAGGGATGCACGTGGCGGGATCCTCTTATCCTGTCGAGGTTGATGGAGTGCAGATGCCAAGATACATATCTCATGCCTCCT

      F  P  D  Y  S  S  F  A  P  C  Q  S  S  V  I  K  S  L  I  Q  E  A  G  V  E  E  R  G  V  S  K  V  G  1508
4501  TCTTCCCCGATTATTCTTCTTTTGCTCCTTGCCAGTCTAGTGTTATCAAATCTCTAATTCAAGAGGCTGGCGTTGAGGAGCGTGGGGTTTCTAAAGTGGG

      H  I  K  D  P  A  E  T  P  H  V  G  G  K  T  K  L  E  L  V  D  E  A  F  L  V  P  S  P  V  E  V  K  1541
4601  ACATATTAAAGATCCTGCTGAGACGCCCCATGTTGGAGGGAAAACTAAGCTTGAATTGGTTGATGAAGCCTTCTTGGTGCCATCACCAGTTGAGGTAAAG

      I  P  S  I  L  S  K  D  D  P  R  I  P  E  A  Y  K  G  Y  D  P  L  G  D  A  M  E  K  F  Y  E  P  M  L  1575
4701  ATTCCCTCCATTCTGTCTAAAGATGACCCGCGCATTCCTGAAGCGTATAAGGGTTATGATCCATTGGGCGATGCCATGGAGAAGTTTTATGAGCCCATGT

      D  L  D  E  D  V  L  E  S  V  M  A  D  M  Y  D  E  F  Y  D  C  Q  T  T  L  R  I  M  S  D  D  E  V  1608
4801  TGGATCTGGACGAAGATGTCTTGGAGAGCGTTATGGCAGATATGTATGATGAATTCTATGATTGCCAAACGACTCTCCGTATTATGTCTGATGACGAAGT

      I  N  G  S  D  F  G  F  N  I  E  A  V  V  K  G  T  S  E  G  Y  P  F  V  L  S  R  R  P  G  E  K  G  1641
4901  TATCAATGGCAGCGATTTTGGTTTCAATATTGAAGCCGTTGTCAAAGGTACTTCTGAAGGCTACCCGTTCGTTTTGAGTCGGCGACCGGGCGAGAAGGGC

      K  A  R  F  L  E  E  L  E  P  Q  P  G  D  T  K  P  K  Y  K  L  V  V  G  T  E  V  H  S  A  M  V  A  M  1675
5001  AAAGCTCGCTTTTTAGAAGAGCTTGAACCCCAACCAGGTGACACTAAGCCTAAATATAAACTGGTTGTGGGCACTGAGGTGCATTCTGCTATGGTGGCGA

      E  Q  Q  A  R  T  E  V  P  L  L  I  G  M  D  V  P  K  D  E  R  L  K  P  S  K  V  L  E  K  P  K  T  1708
5101  TGGAACAACAGGCGCGTACTGAAGTTCCTTTGCTTATTGGTATGGATGTTCCGAAGGATGAGAGACTCAAACCGTCTAAGGTGTTGGAGAAGCCGAAGAC
```

Fig. 1. For legend see page 468.

```
      R  T  F  V  V  L  P  M  H  Y  N  L  L  L  R  K  Y  V  G  I  L  C  S  S  M  Q  V  N  R  H  R  L  A    1741
5201  GCGTACATTCGTTGTTCTCCCAATGCACTATAACTTGCTGCTGCGTAAGTATGTGGGAATTTTGTGTTCTAGCATGCAAGTTAATAGGCATCGTCTAGCA

      C  A  V  G  T  N  P  Y  S  R  D  W  T  D  I  Y  Q  R  L  A  E  K  N  S  V  A  L  N  C  D  Y  S  R  F    1775
5301  TGTGCTGTGGGCACTAACCCATATTCGCGTGATTGGACGGACATTTATCAGCGCCTGGCTGAGAAAAATTCAGTGGCCTTGAATTGTGATTATAGCCGCT

      D  G  L  L  N  Y  Q  A  Y  V  H  I  V  N  F  I  N  K  L  Y  N  D  E  H  S  I  V  R  G  N  L  L  M    1808
5401  TTGATGGGCTCCTCAATTACCAGGCATATGTGCATATTGTTAATTTTATTAATAAATTGTACAACGATGAACATTCTATCGTGCGTGGCAATCTTTTGAT

      A  M  Y  G  R  W  S  V  C  G  Q  R  V  F  E  V  R  A  G  M  P  S  G  C  A  L  T  V  I  I  N  S  L    1841
5501  GGCTATGTATGGTAGGTGGAGTGTGTGTGGGCAGAGAGTTTTCGAAGTCCGCGCTGGCATGCCCTCTGGGTGTGCGCTCACCGTGATCATCAATTCACTT

      F  N  E  M  L  I  R  Y  V  Y  R  I  T  V  P  R  P  L  V  N  N  F  K  Q  E  V  C  L  I  V  Y  G  D  D    1875
5601  TTTAACGAAATGTTGATCAGGTATGTTTATCGCATCACCGTACCACGCCCCCTTGTAAATAATTTTAAACAGGAGGTGTGTTTGATTGTTTATGGTGATG

      N  L  I  S  I  K  P  D  T  M  K  Y  F  N  G  E  Q  I  K  T  I  L  A  K  Y  K  V  T  I  T  D  G  S    1908
5701  ATAATTTAATTTCTATTAAGCCGGACACCATGAAATATTTTAATGGTGAGCAAATCAAAACCATTCTGGCTAAATATAAAGTTACCATTACTGATGGCAG

      D  K  N  S  P  V  L  R  A  K  P  L  K  Q  L  D  F  L  K  R  G  F  R  V  E  S  D  G  R  V  L  A  P    1941
5801  TGATAAGAACTCACCTGTTCTTAGAGCCAAACCCTTGAAACAGCTCGATTTTTTGAAGAGAGGTTTCAGGGTTGAAAGTGATGGGAGGGTGCTTGCCCCT

      L  D  L  Q  A  I  Y  S  S  L  Y  Y  I  N  P  Q  G  N  I  L  K  S  L  F  L  N  A  Q  V  A  L  R  E  L    1975
5901  TTAGATTTGCAAGCTATCTATTCTTCCCTGTATTATATCAATCCGCAGGGAAATATATTAAAATCTTTGTTTTTGAATGCTCAAGTCGCTTTGAGAGAGT

      Y  L  H  G  D  V  E  Q  F  T  A  V  R  N  F  Y  V  N  Q  I  G  G  N  F  L  S  L  P  Q  W  R  H  C    2008
6001  TATATCTCCATGGCGATGTTGAGCAATTTACTGCTGTCAGGAATTTTTACGTCAATCAAATTGGCGGAAATTTCTTGAGTCTACCCCAGTGGAGGCACTG

      A  S  F  H  D  E  Q  Y  S  Q  W  K  P  W  S  P  V  K  F  L  E  V  D  V  P  D  A  K  F  L  Q  H  K    2041
6101  CGCTTCGTTCCATGATGAACAATATTCTCAGTGGAAGCCGTGGTCTCCCGTTAAATTCTTGGAGGTAGATGTGCCCGATGCTAAATTTTTACAGCATAAG

      A  P  A  T  A  L  S  I  V  A  D  R  L  A  V  A  G  P  G  W  R  N  K  D  P  D  R  Y  L  L  V  S  L  T    2075
6201  GCGCCAGCTACTGCCCTCTCGATTGTTGCTGATAGGCTTGCTGTTGCAGGACCTGGCTGGCGTAATAAAGATCCAGACAGGTATCTGTTAGTGAGCCTTA

      S  L  K  A  N  E  G  G  L  Y  F  P  V  D  Y  G  E  G  T  G  Q  Q  A  T  E  A  S  I  R  A  Y  R  R    2108
6301  CCAGCTTGAAAGCAAATGAGGGTGGATTATACTTCCCCGTAGACTACGGGGAGGGTACAGGGCAGCAAGCCACGGAAGCTTCCATTAGAGCTTATCGTAG

      L  K  D  H  R  V  R  H  M  R  D  S  W  N  E  G  K  T  I  V  F  R  C  E  G  P  F  V  S  G  W  A  A    2141
6401  GCTAAAGGATCATCGCGTACGCCATATGCGCGATTCCTGGAATGAGGGAAAGACAATCGTGTTCCGATGCGAAGGTCCCTTTGTTTCAGGATGGGCAGCT

      A  I  S  F  G  T  S  V  G  M  N  A  Q  D  L  L  I  N  Y  G  I  Q  G  G  A  H  K  E  Y  L  G  R  Y  F    2175
6501  GCCATTTCCTTCGGTACTAGCGTTGGAATGAATGCCCAAGATCTCTTAATCAATTATGGTATACAAGGTGGCGCCCACAAGGAATATCTGGGACGCTACT

      V  G  A  R  F  K  E  L  E  R  Y  D  R  P  F  Q  S  R  I  I  A  S  *                                  2197
6601  TTGTTGGTGCTCGTTTTAAAGAGCTGGAGAGGTATGACCGACCTTTTCAGTCTCGTATAATTGCGAGCTAAATCCTCTTTGAGGCGAGTAGCTGCCGTTA
6701  GCAGCTTCCAAAAGGTGGCCTCTTAATTAGCTTTTAATAGGGGTTATCCAGCCTTAAGCAAGCTGGCACCGGTCCTGATGGACTACCAGGAAAGCACCTG
6801  GTTTGGAAGAATTCGAGTAAAATTCTTAAATCTTGTTTACTCGTGACTTATAGTACATTCAAGAGGAATGACTCATGTTTTGTTTATTTACATGATGGCA
6901  TAAAGAGTTAACGGCTCATATGGTGCTCATTACGTTCAAGTGTTGAAGGATCCAATAGCCTTGAACTGTGGTGCCATGTGAGGAAATCCACGTTATCTCT
7001  GATTGTCAAAATAGACTAGTCTAGGAGACGATAAATCCTATGTGGGTGAGTCCCACTCTGGCGAGACACGCAGTGCCTTTTATTTGTTTGAGGTTATCAA
7101  ACATCATATCTTGAGTCTGCATTTAAATTCGAATAATGTAGTTGTCATAGCCTACCGATGAGTCTGCGAGAAAGGTTCCATGAGGACTTGGGTTGGCTAA
7201  CCCCCACTTAATCTCTTCATAGATCATTCGACAGTGTGTCGAGAAACTATGGGTTTTGACACCTTAAGGGAAGCGAGAGTTCTCGTATGGATATCACTCT
7301  AATCATGGTACTTACCATGCCATGTTTAGAGTAAATCATCGCCTCGACGGTGTGATACTTTCCTTAAGTTCTAGTCAAACGATAGTTTCGTTGATCGTAT
7401  GTGAAGCGTGGAGCGAGTTCGAAACGAACTGATTACCCGAGGTAGGACGCTATTGTTCCAGGCGTTTCTTATGGGCATAAGCTGTAAACTTGGTTTCGCA
7501  AGCCATGCAGCACCTCCCGTTACTTGTGTACTTTCTAGGGGCTCCCGGCCTTCCTTCCGGTACAATACCTAGTGAAGCAAGTAATTGCGTTGAGGGATAA
7601  GAGTAGCATGTTCCTACTTAAGGAAGGAATATGTCGTGTTTTCCACACGTTAGTGTTGCAATGCTGTAATGGCACTGCAGTGCAGGAATGGTTCCCAGCC
7701  ACTTTTTCTGGGATTCTAATCGTACGTCACAATTGTGTGTGTATCGTTGACGGAGGAGTAGCGATCCTCTACCACGCGAGTCTGGAAGTGATTACCAGGG
7801  CCTAAGATGGCCAGCACACGGTACGATTAAATTTAGCTGTAATGTAGTGGTATGTTAAGTTGAGACTAACTTACCCGTACGAGTTAAACTCTAAGATGGA
7901  TGTGTGTTCTGCCATCTTAGAGGAAGTAGATGTGTTTTTACCAATCTGAGACGAGCCGTTAATTCGGTGCTTTAATACGTCAATGATAATACTCGTGCAG
8001  TTGCAGCTGCACGAGTATGTTGGTACACACAGTCTACTCGGATACGGTCGAGTTACCCTCACAATAGGGATTACTCTCTCAATCTTAACTACTGCAAGGA
8101  CGTTGTTTTCGCAGGGTTTTGTTGGTCCGTTTGTGTTTCAAAACGCTGCTTTGCAATTTTCTTTTTTGTTTTATTGCTTTCGTAGTGTCGAACTTTGTCC
8201  AAGTTCATAAAAGC [poly(A)]
```

Fig. 1. Nucleotide sequence and deduced amino acid sequence of TomRSV RNA1 cDNA. The first two nucleotides are undetermined and are each represented by an N. Numbers to the left of the sequence refer to nucleotide sequence position and numbers to the right refer to amino acid sequence position.

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_{11}-L-x_{21}-Lx_1E (x_n 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. 2*a*). 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 32K protein contains this sequence and has been demonstrated to function as a cofactor for the CPMV 24K protease (Vos *et al.*, 1988; Peters *et al.*, 1992).

The TomRSV RNA1 polyprotein sequence beginning at amino acid residue 791 includes the 'A' and 'B' motifs

characteristic of NTP-binding proteins (Fig. 2*b*) (Gorbalenya & Koonin, 1989; Gorbalenya *et al.*, 1989*a*). The highly conserved 'A' and 'B' site motifs are Gxx(G)xGKS/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 *et al.* (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*-x_{46}-*E*-x_{101}-*C*G--x_8-GxxxxxGxHxxG) 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.*, 1989*b*; Hammerle *et al.*, 1991; Dessens & Lomonossoff, 1992; Margis & Pinck, 1992) (Fig. 2*c*). 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. 2*d*). 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. 2*e*). 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

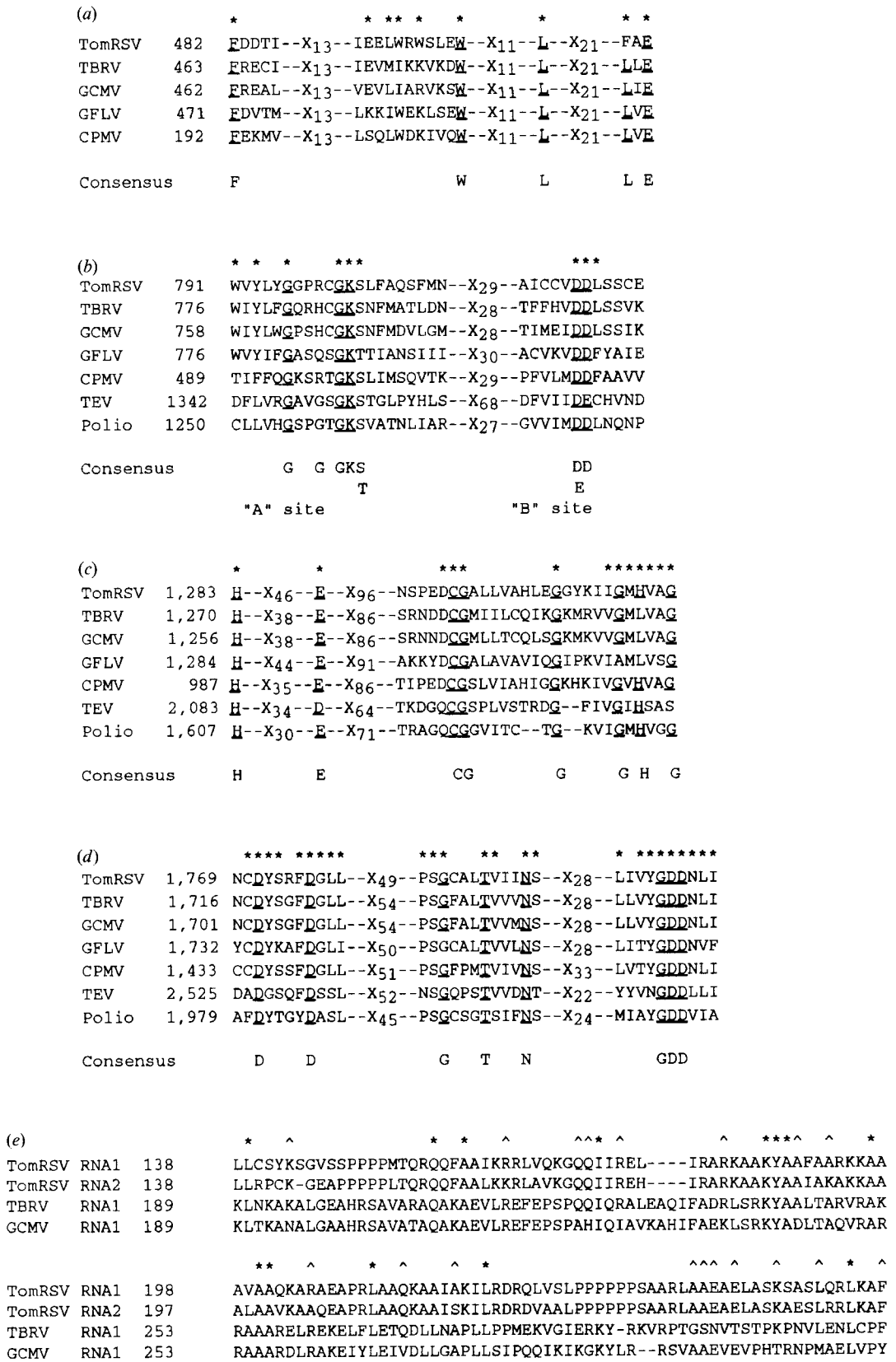

Fig. 2. For legend see opposite.

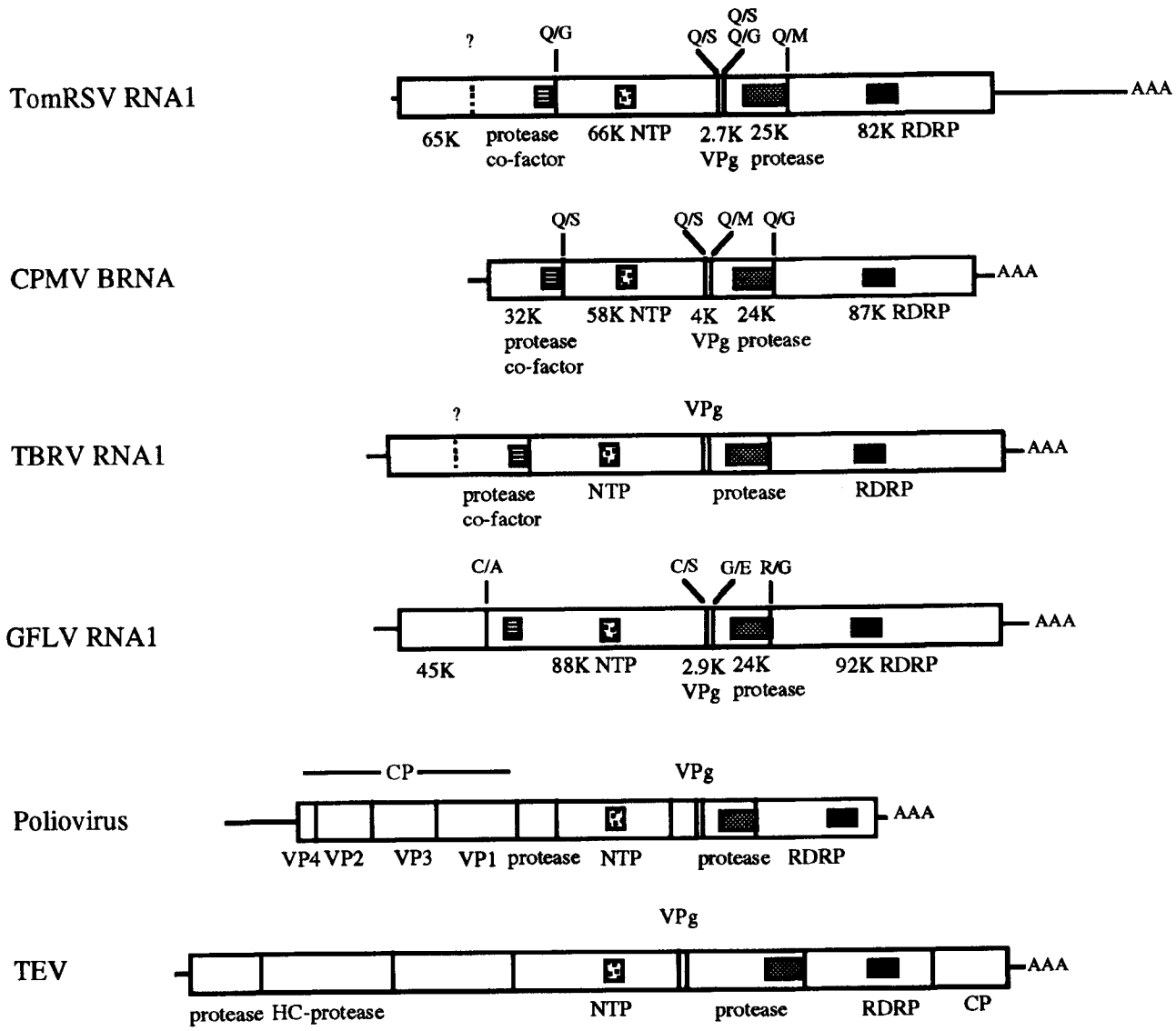

Fig. 3. Proposed genomic organization of TomRSV RNA1 and comparison with other members of the picornavirus superfamily. Lines indicate noncoding sequence and the bars represent the polyprotein sequence encoded by the long ORFs. Vertical lines through the bars indicate known and putative protease cleavage 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_n 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 side 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.*, 1991*b*). 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

ALLISON, R. JOHNSTON, D. E. & DOUGHERTY, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* **154**, 9–20.

ARGOS, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Research* **16**, 9909–9919.

ARGOS, P., KAMER, G., NIKLIN, M. J. H. & WIMMER, E. (1984). Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. *Nucleic Acids Research* **12**, 7251–7267.

BAZAN, J. F. & FLETTERICK, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proceedings of the National Academy of Sciences, USA* **85**, 7872–7876.

DEMANGEAT, G., GRIEFF, C., HEMMER, O. & FRITSCH, C. (1990). Analysis of the *in vitro* cleavage products of the tomato black ring virus RNA-1-encoded 250K polyprotein. *Journal of General Virology* **71**, 1649–1654.

DESSENS, J. T. & LOMONOSSOFF, G. P. (1992). Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology* **184**, 738–746.

DORSSERS, L., VAN DER KROL, S., VAN DER MEER, J., VAN KAMMEN, A. & ZABEL, P. (1984). Purification of cowpea mosaic virus RNA replication complex: identification of a virus-encoded 110,000 dalton polypeptide responsible for RNA chain elongation. *Proceedings of the National Academy of Sciences, USA* **81**, 1951–1955.

GOLDBACH, R. (1987). Genomic similarities between plant and animal RNA viruses. *Microbiological Sciences* **4**, 197–202.

GOLDBACH, R. & REZELMAN, G. (1983). Orientation of the cleavage maps of the 200-kilodalton polypeptide encoded by the bottom component RNA of cowpea mosaic virus. *Journal of Virology* **46**, 614–619.

GOLDBACH, R. & VAN KAMMEN, A. (1985). Structure, replication and expression of the bipartite genome of cowpea mosaic virus. In *Molecular Plant Virology*, vol. 2, pp. 83–120. Edited by J. W. Davies. Boca Raton, FL: CRC Press.

GORBALENYA, A. E. & KOONIN, E. V. (1989). Viral protein containing the purine NTP-binding sequence pattern. *Nucleic Acids Research* **17**, 8413–8440.

GORBALENYA, A. E., BLINOV, V. M., DONCHENKO, A. P. & KOONIN, E. V. (1989*a*). An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. *Journal of Molecular Evolution* **24**, 256–268.

GORBALENYA, A. E., DONCHENKO, A. P., BLINOV, V. M. & KOONIN, E. V. (1989*b*). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. *FEBS Letters* **243**, 103–114.

GREIF, C., HEMMER, O. & FRITSCH, C. (1988). Nucleotide sequence of tomato black ring virus RNA-1. *Journal of General Virology* **69**, 1517–1529.

HAMMERLE, T., HELLEN, C. U. T. & WIMMER, E. (1991). Site-directed mutagenesis of the putative catalytic triad of poliovirus 3C proteinase. *Journal of Biological Chemistry* **266**, 5412–5416.

HELLEN, C. U. T., KRÄUSSLICH, H. & WIMMER, E. (1989). Proteolytic processing of polyproteins in the replication of RNA viruses. *Biochemistry* **28**, 9881–9890.

LE GALL, O., CANDRESSE, T., BRAULT, V. & DUNEZ, J. (1989). Nucleotide sequence of Hungarian grapevine chrome mosaic nepovirus RNA1. *Nucleic Acids Research* **17**, 7795–7807.

LOMONOSSOFF, G. P. & SHANKS, M. (1983). The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO Journal* **2**, 2253–2258.

MARGIS, R. & PINCK, L. (1992). Effects of site-directed mutagenesis on the presumed catalytic triad and substrate-binding pocket of grapevine fanleaf nepovirus 24-kDa proteinase. *Virology* **190**, 884–888.

MARGIS, R., VIRY, M., PINCK, M., BARDONNET, N. & PINCK, L. (1994). Differential proteolytic activities of precursor and mature forms of the 24K proteinase of grapevine fanleaf nepovirus. *Virology* **200**, 79–86.

MARTELLI, G. P. (1975). Some features of nematode-borne viruses and their relationship with the host plants. In *Nematode-borne Vectors of Plant Viruses*, pp. 223–252. Edited by F. Lamberti, C. E. Taylor & J. W. Seinhort. London & New York: Plenum Press.

PALMEMBERG, A. C. (1990). Proteolytic processing of picornaviral polyprotein. *Annual Review of Microbiology* **44**, 603–623.

PETERS, S. A., VOORHORST, W. G. B., WERY, J., WELLINK, J. & VAN KAMMEN, A. (1992). A regulatory role for the 32K protein in proteolytic processing of cowpea mosaic virus polyproteins. *Virology* **191**, 81–89.

PINCK, M., REINBOLT, J. A. M., LOUDES, A. M., LE RET, M. & PINCK, L. (1991). Primary structure and location of the genome-linked protein (VPg) of grapevine fanleaf nepovirus. *FEBS Letters* **284**, 117–119.

RACANIELLO, V. R. & BALTIMORE, D. (1981). Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proceedings of the National Academy of Sciences, USA* **78**, 4887–4891.

RITZENTHALER, C., VIRY, M., PINCK, M., MARGIS, R., FUCHS, M. & PINCK, L. (1991). Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. *Journal of General Virology* **72**, 2357–2365.

ROTT, M. E., TREMAINE, J. H. & ROCHON, D. M. (1991*a*). Nucleotide sequence of tomato ringspot virus RNA-2. *Journal of General Virology* **72**, 1505–1514.

ROTT, M. E., TREMAINE, J. H. & ROCHON, D. M. (1991*b*). Comparison of the 5′ and 3′ termini of tomato ringspot virus RNA1 and RNA2: evidence for RNA recombination. *Virology* **185**, 468–472.

SCOTT, N. W., COOPER, J. I., LIU, Y. Y. & HELLEN, C. U. T. (1992). A 1·5 kb sequence homology in the 3′-terminal regions of RNA-1 and RNA-2 of a birch isolate of cherry leaf roll nepovirus is also present, in part, in a rhubarb isolate. *Journal of General Virology* **73**, 481–485.

TAKEDA, N., KUHN, R. J., YANG, C. F., TAKEGAMI, T. & WIMMER, E. (1986). Initiation of poliovirus plus-strand RNA synthesis in a membrane complex of infected HeLa cells. *Journal of Virology* **60**, 43–53.

VOS, P., VERVER, J., JAEGLE, M., WELLINK, P. & GOLDBACH, R. (1988). Two viral proteins involved in the proteolytic processing of the cowpea mosaic virus polyprotein. *Nucleic Acids Research* **16**, 1967–1985.

WELLINK, J., REZELMAN, G., GOLDBACH, R. & BEYREUTHER, K. (1986). Determination of the proteolytic processing sites in the polyprotein encoded by the bottom-component of cowpea mosaic virus. *Journal of Virology* **59**, 50–58.

WIECZOREK, A. & SANFACON, H. (1993). Characterization and subcellular localization of tomato ringspot nepovirus putative movement protein. *Virology* **194**, 734–742.

ZABEL, P., MOERMAN, M., LOMONOSSOFF, G., SHANKS, M. & BEYREUTHER, K. (1984). Cowpea mosaic virus VPg: sequencing of radiochemically modified protein allows mapping of the gene on B RNA. *EMBO Journal* **3**, 1629–1634.

(*Received 18 April 1994; Accepted 19 October 1994*)