Nucleotide sequence of Hungarian Grapevine Chrome Mosaic Nepovirus RNA1

Olivier O. Le Gall, Thierry T. Candresse, Véronique Brault, Jean Dunez

Volume 17 Number 19 1989 Nucleic Acids Research

Nucleotide sequence of Hungarian grapevine chrome mosaic nepovirus RNA1

O.Le Gall, T.Candresse, V.Brault and J.Dunez

Station de Pathologie Végétale, INRA, BP 131, 33140 Pont de la Maye, France

Received May 24, 1989; Revised and Accepted August 25, 1989 EMBL accesssion no. X15346

ABSTRACT

The nucleotide sequence of the RNA1 of hungarian grapevine chrome mosaic virus, a nepovirus very closely related to tomato black ring virus, has been determined from cDNA clones. It is 7212 nucleotides in length excluding the 3′ terminal poly(A) tail and contains a large open reading frame extending from nucleotides 216 to 6971. The presumably encoded polyprotein is 2252 amino acids in length with a molecular weight of 250 kDa. The primary structure of the polyprotein was compared with that of other viral polyproteins, revealing the same general genetic organization as that of other picorna-like viruses (comoviruses, potyviruses and picornaviruses), except that an additional protein is suspected to occupy the N-terminus of the polyprotein.

INTRODUCTION

As in the other nepoviruses, the genome of hungarian grapevine chrome mosaic virus [GCMV, (1)] is constituted of two positive sense, single-stranded RNAs, separately encapsidated in icosahedral particles. The RNAs are polyadenylated at their 3′ end and probably covalently linked with a small polypeptide (VPg) at their 5′ end (2). The molecular weights of GCMV RNAs are estimated at about 2.8×10^6 Da (RNA1) and 1.6×10^6 Da (RNA2) upon gel electrophoresis under denaturing conditions. Studies on the distribution of genetic functions between genome segments of nepoviruses have shown that RNA1 is able to replicate independently in protoplasts (3) and thus codes for the information necessary for viral replication. RNA2 encodes a protein mediating viral cell-to-cell spread (3) as well as the coat protein (4).

GCMV is closely related to tomato black ring virus (TBRV) both serologically (5) and by nucleic acid hybridization (6). Pseudorecombinants have been obtained between GCMV and TBRV (7) and GCMV can cross protect *Chenopodium quinoa* Wild against the severe apical necrosis caused by the strain S of TBRV (8, 9). Those similarities made some authors regard GCMV as a strain of TBRV (8) although others considered them to be two distinct viruses (10). Sequencing of the TBRV genome has recently been achieved (11, 12, 13) and its genetic organization was found to be very similar to that of the comovirus cowpea mosaic virus (CPMV) (15, 16).

We have determined the nucleotide sequence of GCMV RNA1. Together with the sequence of GCMV RNA2 (Brault *et al.*, accompanying paper), this completes the sequencing of the genome of GCMV. Comparison of the deduced translation products of GCMV RNA1 with other viral proteins provides further information about the genetic organization of nepoviruses.

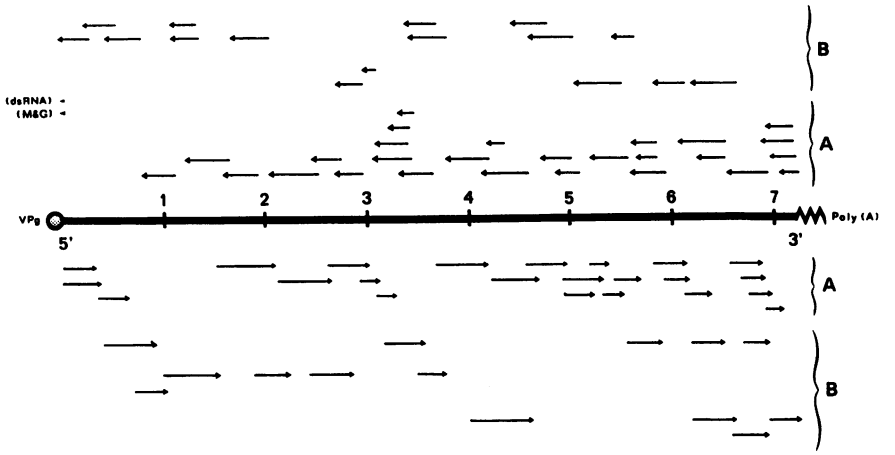


Figure 1 : Sequencing strategy of GCMV RNA1. The VPg at the 5′ end and the poly(A) tail at the 3′ end of the viral RNA (thick line) are represented. Arrows represent the extent of the data obtained from a single sequencing reaction. **A** : sequences determined from independent cDNA clones; **B** : sequences obtained from subclones; **(dsRNA)** : sequence obtained previously by direct RNA sequencing (2); **(M&G)** : data obtained by chemical sequencing of a 5′ terminal cDNA.

MATERIALS AND METHODS

Virus propagation and viral RNA preparation

GCMV was obtained from Dr. G.P. Martelli (Bari, Italy) and routinely propagated in the herbaceous host *Chenopodium quinoa* Wild. Virus and viral RNAs were purified as previously described (7).

Cloning of cDNA

Priming of the first strand of the cDNA was with oligo(dT) (pdT_{12-18}, Pharmacia) and the reverse transcriptase of avian myeloblastosis virus (AMV RTase, Genofit). Synthesis of the second strand was done either using RNase H, *Escherichia coli* DNA Polymerase I and DNA ligase (17) or using the Klenow fragment of DNA polymerase I after priming with a synthetic oligonucleotide homologous to the 5′ end of the viral RNA (18).

The double-stranded cDNA was inserted into the SmaI site of the plasmid Bluescribe (Stratagene) and the ligation mixture used to transform competent *E. coli* cells (strain DH5α, BRL). Colonies with the [$LacZ^-$, Amp^R] phenotype were selected and recombinant plasmids characterized by restriction analysis after extraction by the alkaline lysis method (19). Clones containing overlapping cDNA sequences were selected for the determination of the nucleotide sequence.

```
                    5'          3'
Virion RNA    VPg-UUGGAAAAUU---
                  ||||||||||---
(-) strand      OH-ACCUUUUAA---
                    3'          5'
```

Figure 2 : Evidence for the presence of an additional uridine at the 5′ end of the virion RNA, as compared with the 3′ end of the minus strand in a putative dsRNA molecule (2).

Subcloning of cDNA fragments
Some restriction fragments (EcoRI, HindIII, HpaII, PstI, PvuII, SstI, TaqI) obtained from two independent clones were subcloned (18, 19, 20) into pBluescribe. In addition, overlapping deletions in a full-length cDNA insert were produced using random deoxyribonuclease I digestion (21).
Nucleotide sequence determination and analysis
The nucleotide sequence was determined by the chain-termination method (22), using supercoiled plasmid DNA as the template. The DNA was extracted from a 5 ml overnight culture by a modified alkaline lysis method (23) and further treated with 4μg proteinase K in 60μl TE buffer (10mM Tris-Cl, 1mM EDTA, pH=8.0) containing 0.5% (w/v) SDS (37°C, 30 min), phenol extracted and precipitated in 0.5M NaCl, 6.5% PEG (0°C, 1h). The pellet was resuspended in 50μl TE and 8μl were used for sequencing. After a 5-min denaturation of the template in 0.4M NaOH, 15pmol of the sequencing primer were added and the DNA ethanol precipitated.

The sequencing reactions were made using a synthetic 17-mer oligodeoxynucleotide primer complementary to vector sequences of the T3 or of the T7 promoter, depending on the end of the insert to be sequenced. Sequenase™ (USB) was used following the supplier's indications and [^{35}S]dATPαS (Amersham) incorporated as the radiolabel. Alternatively, AMV RTase (Genofit) and [α-^{32}P] dATP were used according to Zagursky *et al.* (23).

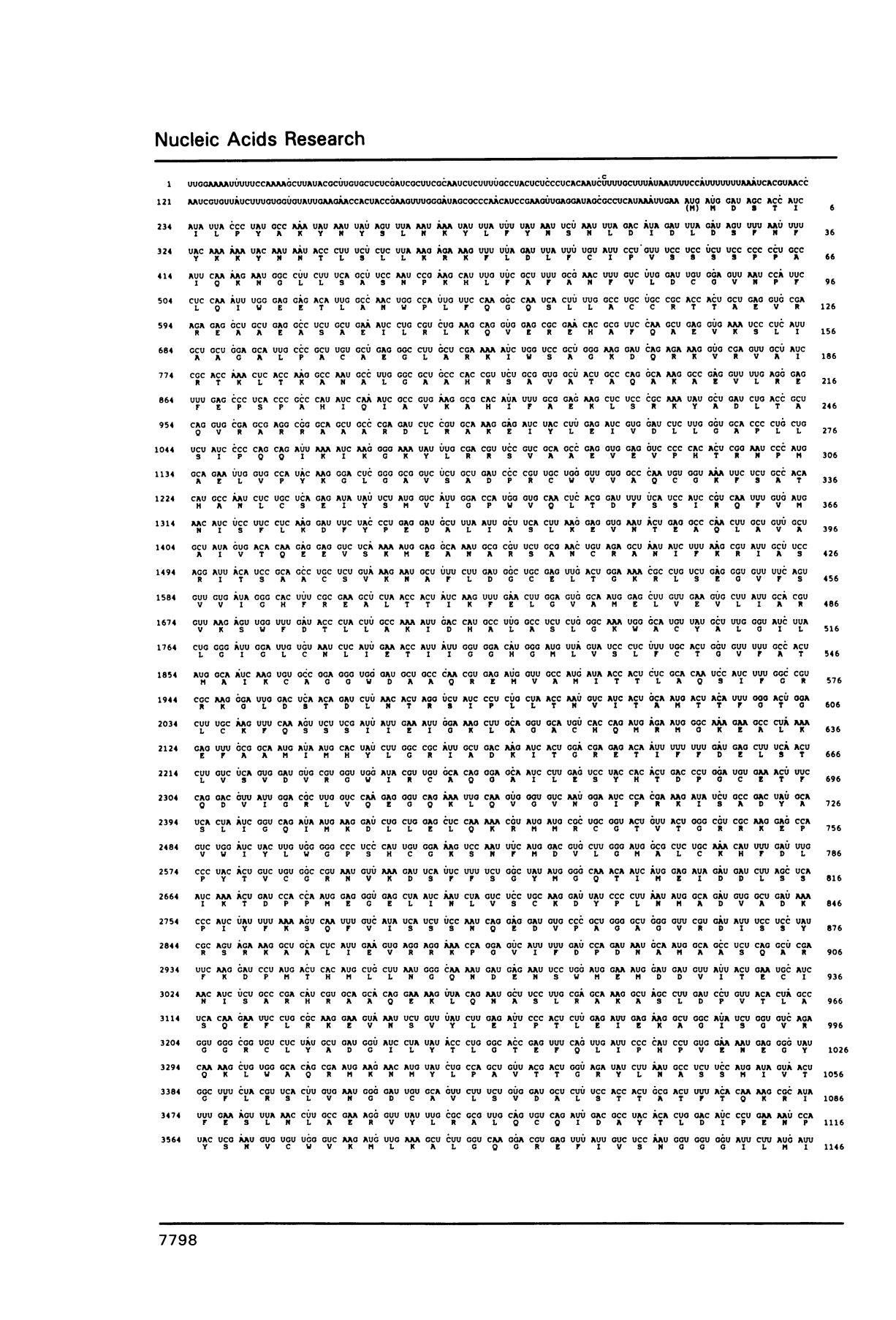
Analysis of the sequence was made using the Cornell DNA sequence analysis package (24), the NIH Molecular Biology PC-tools distribution package, the program CLUSTAL (25) and the PC-GENE package (Genofit), run on a Micral (Bull) microcomputer.

RESULTS AND DISCUSSION

Nucleotide sequence and coding capacity of RNA1
Overlapping independent cDNA inserts (Fig. 1-A) were used to determine most of the nucleotide sequence of GCMV RNA1. In several regions where no independent clones were available, subclones were used (Fig. 1-B). The complete sequence was determined on both cDNA strands, and the major part of it was verified at least 3 times independently.

In a previous study (2), the 3′ terminal sequence of the minus strand in GCMV dsRNAs was unambiguously determined. In order to confirm this result on the virion strand, a 5′ terminal cDNA obtained by primer extension was chemically sequenced (26). As shown on figure 2, the comparison of these two sequences provide evidence for the presence of an extra non-templated uridine at the 5′ end of the virion strand. The GCMV RNA1 sequence is 7212 nucleotides in length, excluding the poly(A) tail and including the VPg-bound uridine.

Only one ORF, representing almost the full coding capacity of the RNA, has been identified in the positive (viral) sense. No other ORF, either in the positive or the negative strand, exceeds 396 nucleotides in length. The ORF begins at position 216 or 219, where the two first AUG codons in the sequence occur as a doublet. Both of the putative initiation codons are in a context known to enhance translation in eukaryotic cells (27, 28), with a G or an A at position −3, but the second AUG is in a better context than the first one for translation in plants (29), with A's at positions −3 and −4 and a G at position +4. Translation of RNA1 probably starts from one of these two AUG codons and ends at the opal codon at position 6972, resulting in a 2251 or 2252 amino acids protein, with a calculated molecular weight of about 250 kDa. A polypeptide of molecular weight 250



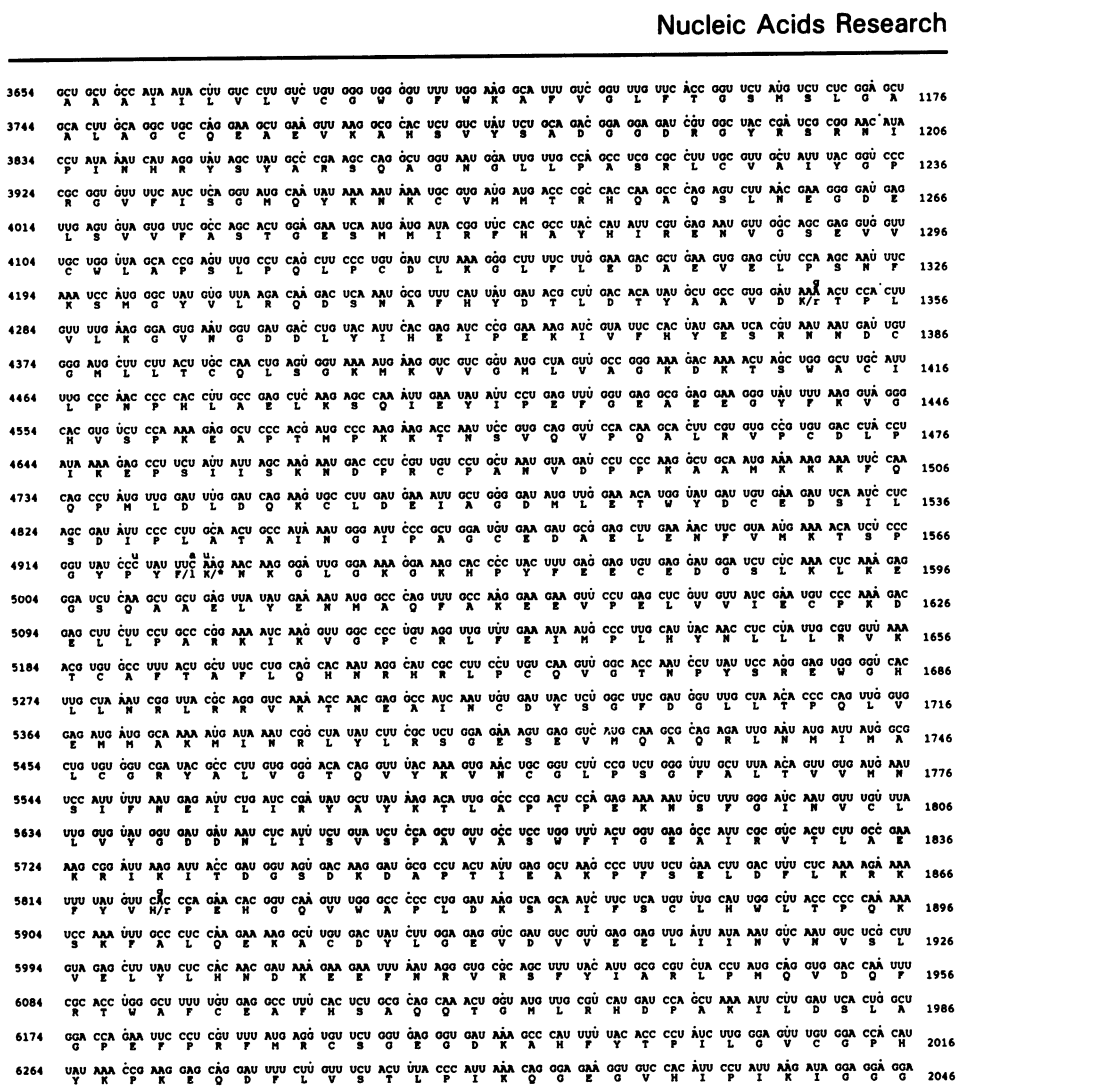


Figure 3 : Nucleotide sequence of GCMV RNA1 and deduced amino-acid sequence of the large open reading frame of GCMV RNA1. Sequence variations found in some cDNA clones are indicated in low letter case as well as the amino-acid exchange they might cause.

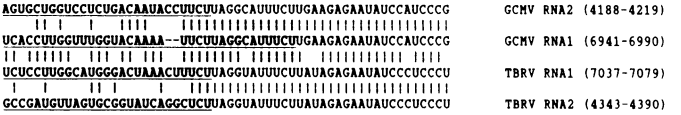


Figure 4 : Evidence for a frameshift in GCMV RNA1, based on the nucleotide sequence comparison between GCMV and TBRV RNAs in the region containing the stop codons, and implication for the termination of the coding region. The coding regions are underlined.

kDa is indeed observed upon *in vitro* translation of GCMV RNA1 (G. Demangeat, personal communication). The nucleotide sequence of RNA1 and the deduced amino acid sequence of the encoded protein are shown in figure 3.

Several nucleotide exchanges, shown in lower case in figure 3, have been found in some cDNA clones. They might represent mistakes made by the reverse transcriptase during the initial cloning step or alternatively correspond to sequence heterogeneities in the viral RNA population. One of them is found in the 5′ non-coding region, six in the ORF and ten in the 3′ non-coding region. From the six exchanges found in the coding region, two do not cause any amino-acid exchange, two do not radically change the structure of the mutated amino-acid, one introduces a completely different amino-acid and one introduces an amber codon, thus making the potential ORF end at position 4928. All of the exchanges found in the 3′ non-coding region, including the opal to ochre termination codon exchange,

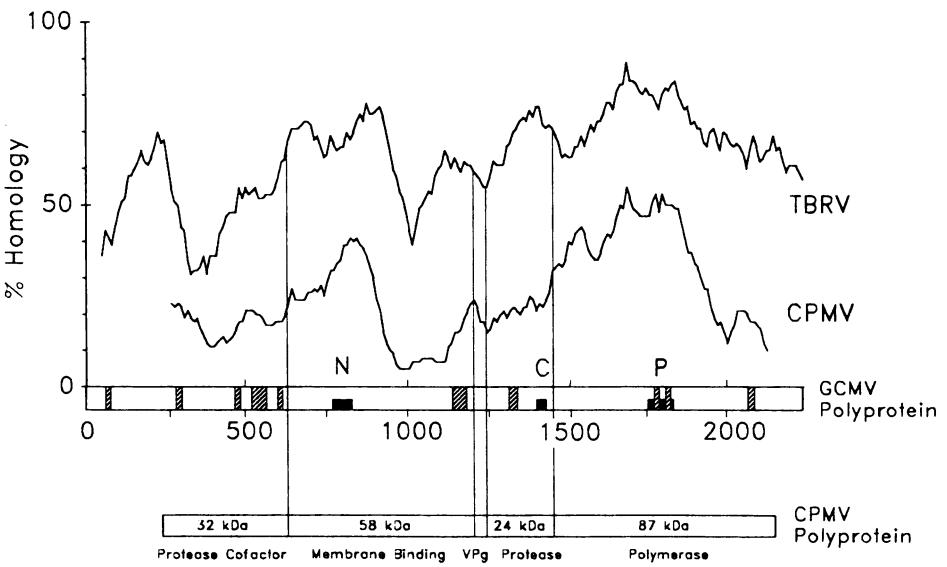


Figure 5 : Amino-acid homologies between the polyproteins encoded by GCMV and, respectively, TBRV and CPMV. Percent homology are calculated in a window of 100 residues around the plotted position. The CPMV polyprotein is symbolized at the bottom of the figure and its cleavage sites indicated by thin lines. The GCMV polyprotein is also presented. Shaded areas are the membrane associated domains predicted by the programs RAOARGOS and HELIXMEM and black areas are the regions presented in figure 5 (**N** : Nucleotide binding site, **C** : Protease active site and **P** : Polymerase conserved domain).

```
Nucleotide binding site (N)

                        .  *    ***   . .   . ..
GCMV            EPVWIYL-WGPSHCGKSNFMDVLGMALCKHFDLPYTV
TBRV            EPVWIYL-FGQRHCGKSNFMATLDNALAKHFGLPNTT
CPMV            MPFTIFF-QGKSRTGKSLLMSQVTKDFQDHYGLGGET
TEV             SPARDFLVRGAVGSGKS---TGLPYHLSKR-GRVLML
Polio           IEPVCLLVHGSPGTGKSVATNLIARAIAER---ENTS

Consensus                  GxxGxGKs
                             t


Cysteine protease active site (C)

                      .  ..**  .      *   ..*.  ..   .. .
GCMV            HYESRNNDCGMLLTCQLSGKMKVVGMLVAG--KDKTS
TBRV            HYESRNDDCGMIILCQIKGKMRVVGMLVAG--KDKTS
CPMV            EAPTIPEDCGSLVIAHIGGKHKIVGVHVAG-IQGKIG
TEV             MFPTKDGQCGSPLVSTRDGF--IVGIHSASNFTNTNN
Polio           ---TRAGQCGGVITC--TG--KVIGMHVGG--NGSHG

Consensus         TxxxxCG         G      GxHvaG
                                           sg


RNA Polymerase conserved domain (P)

                    . *  ** . * . *...  . .          .   . *** . .     .
GCMV            KVNCGLPSGFALTVVMNSIFNEILIRY--X17--VCLLVYGDDNLISVSPAVA
TBRV            KVNCGLPSGFALTVVVNSVFNEILIRY--X17--VCLLVYGDDNLISVSPSIA
CPMV            RVECGIPSGFFMTVIVNSIFNEILIRY--X22--IGLVTYGDDNLISVNAVVT
TEV             KKHKGNNSGQPSTVVDNTLMVIIAMLY--X11--IVYYVNGDDLLIAIHPDKA
Polio           CVKGGMPSGCSGTSIFNSMINNLIIRT--X13--LKMIAYGDDVIASYPHEVD

Consensus         GxxsGxxxTxxxNs                        GDD
                     t         t
```

Figure 6 : Alignment of the three regions indicated in figure 5 (N, C, P : see figure 5) between GCMV, TBRV, CPMV, TEV and Poliovirus. Positions 755–790 (N), 1378–1412 (C) and 1760–1822 (P) of the polyprotein of GCMV are shown. The multiple alignment was obtained using the program CLUSTAL (25). The stars indicate the amino acids conserved in the five polypeptides, the points the conservative substitutions and the consensual amino acids in each type of site are underlined.

were found in a single cDNA clone, possibly derived from an RNA molecule having accumulated several mutations.

Terminal non-coding regions

The 5′ and 3′ terminal non-coding regions are 215 (from positions 1 to 215) and 241 (from positions 6972 to 7212) nucleotides in length, respectively.

As in TBRV, the 3′ non-coding regions of GCMV RNAs 1 and 2 are 100% homologous (accompanying paper) but it should be noticed that a frameshift caused by an apparent deletion at position 6956 in RNA1 leads to termination of the coding sequence at position 6971 instead of 6960, the position homologous to the termination position of RNA2 and both TBRV RNAs (Fig. 4).

The difference in size between GCMV RNA1 and TBRV RNA1 (7212 versus 7356 nucleotides in length respectively) is mainly located in the non-coding sequences, 260 (5′ non-coding) and 304 (3′ non-coding) nucleotides in length respectively in TBRV RNA1, the size differences being due to long deletions/insertions (see accompanying paper).

Comparisons of the protein with other viral proteins and search for significant motifs in the amino-acid sequence

Nepoviruses are similar to comoviruses, potyviruses and picornaviruses in many ways, including their genome composed of single-stranded, positive sense RNA ending with a VPg and a poly(A) tail, and translation of the genomic RNA(s) into a polyprotein further cleaved by (a) virus-encoded protease(s), to yield the mature viral proteins [for reviews see (15, 16)]. Thus the 250 kDa protein potentially encoded by GCMV RNA1 is a polyprotein. This is confirmed by *in vitro* translation results showing that the primary translation product of GCMV RNA1 is processed (G. Demangeat, personal communication), as also described for other nepoviruses (4, 14).

In order to locate the viral proteins along the precusor, we aligned colinearly the RNA1-encoded polyprotein with that encoded by another nepovirus [TBRV RNA1, (13)], a comovirus [CPMV RNA-B, (30)], a potyvirus [Tobacco etch virus, TEV, (31)] and a picornavirus [Poliovirus strain Sabin 1, (32)], using the program FASTP (33).

Figure 5 presents the percentage of identical amino-acids between GCMV and TBRV or GCMV and CPMV respectively. This procedure did not allow the detection of sufficient homology between GCMV and TEV or Poliovirus to be drawn. The average homology with TBRV is 62.6% and that with CPMV is 21.1% but several regions show a much higher homology level, suggesting that these conserved domains support some important biological function, probably similar to that of their CPMV counterpart. Such conserved blocks are found at several locations along the sequence.

At the N-terminus of the polyprotein of GCMV, a block of high homology with TBRV was found between residues 90 and 275. The NH_2-terminal protein of CPMV (32 kDa protein), a cofactor for the viral protease (34), does not overlap with this region. Either this cofactor is bigger in nepoviruses than it is in comoviruses, or an additional protein is encoded by the 5′ region of the ORF in nepoviruses, cleaved from the polyprotein at a position around residues 250−350. A data bank search, using the program FASTP, between the part of the polyprotein comprised between amino-acids 90−275 with the peptide sequences stored in two data banks (SWISSPROT and NBRF) did not reveal significant homologies between the potential NH_2-terminal protein and any peptide with a known function.

The PC/GENE programs RAOARGOS and HELIXMEM predict several membrane-associated blocks in the terminal region (Fig. 5), mainly in the part homologous to the CPMV 32 kDa protein. The protease helper could thus be associated with membranes *in vivo*.

The large domain of homology with both TBRV and CPMV, located between GCMV residues 600 and 950 (Fig. 5), contains the motif GxxGxGK-S/T (where x may be any amino acid), recently described as the nucleotide binding site of nucleic-acids processing proteins (35, 36, 37, 38), at position 763. Homologies between GCMV polyprotein, TBRV polyprotein, CPMV 58 kDa protein, TEV cytoplasmic inclusion protein and Poliovirus protein 2C in this region (indicated by N in fig. 5) are shown in figure 6-N. The corresponding domain in CPMV is a membrane-associated protein (39) thought to play a role in RNA replication (40), as does a precursor of the poliovirus protein 2C (41). The possible membrane-location of the protein is confirmed by the presence, at the carboxy end of the domain homologous to the CPMV 58 kDa protein, of a highly hydrophobic domain (located between residues 1138 and 1184, see fig. 5), predicted by the program HELIXMEM to be a transmembrane helix.

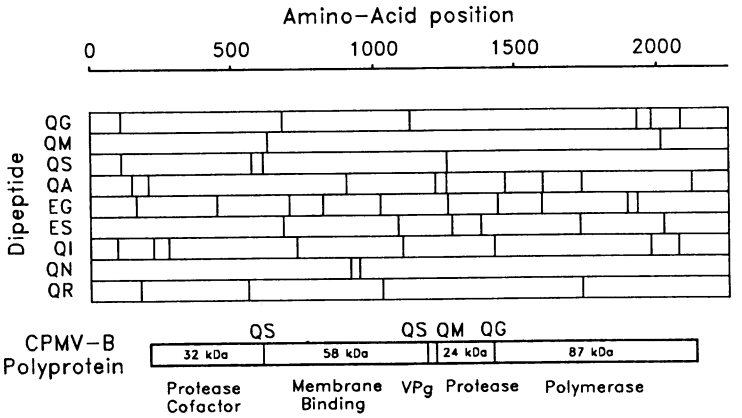


Figure 7 : Location of potential cleavage sites along the polyprotein encoded by GCMV RNA1. Selected dipeptides are those indicated (44, 46). The polyprotein of CPMV, its cleavage sites and the CPMV proteins are indicated at the bottom of the figure.

The active site of cysteine proteases classically contains the motif TxxGxCG-x_{8-9}-G-x_{1-4}-ppGxH-S/V-A/G-G where x may be any amino acid residue and p is a polar amino acid (42, 43). This motif is found between residues 1381 and 1407 of the GCMV polyprotein (indicated C in fig. 5) with some modifications (Fig. 6-C) : the threonine is changed into a serine, also a polar residue, and, more strikingly, the histidine supposed to be part of the substrate binding pocket in that type of proteases [for a review see (44)] is replaced by a leucine in GCMV. The histidine/leucine substitution in the active site of a viral cysteine protease has also been reported in the case of the only other nepovirus so far sequenced, TBRV (13).

The last block of homology concerns the polymerase region where a large domain, extending from residues 1450 to 1850, is highly homologous with both TBRV and CPMV (Fig. 5). Extensive homologies in a more restricted region containing the consensus sequence of active sites of RNA-dependent polymerases (45), indicated P in figure 5, is shown in figure 6-P. Both the GxxS/T-GxxxTxxxN-S/T and the GDD flanked by hydrophobic regions are found in GCMV, in positions 1764 and 1810 respectively.

Thus the gene cluster characteristic of the picorna-like viruses, Nucleotide binding protein/VPg/Protease/Polymerase is found in GCMV except for the VPg, which is too small and contains insufficient homology to be located by the comparative method used here.

Search for potential cleavage sites of the polyprotein

The homology study, as well as the search for consensus sequences, reveals the probable functions of each domain in the polyprotein. However, since none of the GCMV proteins, except the coat protein, has been isolated or characterized, the exact size of the viral proteins remains unknown. For picorna-like viruses, a few dipeptides are known to be potential cleavage sites of polyproteins by the viral protease(s). These are those listed in figure 7 (44). None of these dipeptide is found in positions homologous to the cleavage sites of CPMV (Fig. 7), except a Q/A at position 1218, corresponding exactly in position to the VPg/Protease junction of CPMV and another Q/A at position 208, possibly allowing cleavage of the putative N-terminal protein of nepoviruses. Thus the cleavage of the GCMV RNA1-encoded polyprotein occurs either at positions not exactly colinear with the cleavage

sites of CPMV or at dipeptides different from those usually found in picorna-like viruses.

Mutational analysis (46) showed that the cleavage site can be modified in its second position without extensive loss of efficiency and that pairs such as Q/I, Q/R or Q/N are recognized about as well as the wild type Q/S site in TEV. It is thus possible that the Q/I at position 1428 is the cleavage site between the protease and the polymerase, since its position corresponds almost exactly to that observed in CPMV.

In addition, the coat protein of GCMV seems to be matured by cleavage of a R/A dipeptide (see accompanying paper). Such a cleavage is thus likely to occur also in the translation product of RNA1. Actually, cleavage of one of the R/A dipeptides at positions 249, 252 and 259 could yield the putative N-terminal protein of nepoviruses.

CONCLUSION

A striking feature found in this work is the presence of a non-templated uridine at the 5′ end of the virion strand of both GCMV RNAs (see accompanying paper). For several viruses, the unpaired 5′ end of the virion strand or 3′ end of the minus strand have been described (47, 48, 49, 50). In most cases, the minus strand contains an additional nucleotide at its 3′ end, but in poliovirus, several variations are found : the minus strand may either be exactly complementary to the virion strand (this is actually the case in most dsRNA molecules), or alternatively it may have an additional adenosine or lacks one or two of the terminal adenosines (50). Thus, the case of GCMV seems related to that found in some of the molecules of poliovirus dsRNA. The biological signification of this feature is obscure because the poliovirus dsRNA molecules found to have an unpaired end are thought to be accumulated non-functional species (50). We cannot, however, rule out the possibility that the dsRNA underwent some form of trimming and lost the 3′-terminal base.

Homologies between GCMV and TBRV are, on average, about 60%, both at the nucleotide and at the amino acid levels. GCMV and TBRV should thus be regarded as closely related but distinct viruses rather than as strains of the same virus. Strong homologies between the 3′ non-coding regions are probably related to the viability of pseudorecombinants between these viruses, because this region of the RNA is thought to contain signal sequences necessary for replication.

Amino-acid sequence homologies confirm that the general organization of the polyprotein encoded by GCMV RNA1 is similar to that of the other picorna-like viruses, i.e. picornaviruses, comoviruses and potyviruses since homologous genes are found in the same order in these viruses (16). An additional protein might be encoded at the N-terminus of the polyprotein in nepoviruses because this region is conserved between TBRV and GCMV but has no CPMV counterpart. The existence of this hypothetical protein has already been suspected in TBRV (13) and the homology profile presented in figure 5, as well as the conservation of its presumed cleavage site, provide additional evidence for its existence.

The steps of the processing of the polyprotein are far from clear since the cleavage sites usually found in picorna-like polyproteins are absent from the expected regions in the two nepoviruses sequenced so far. This could be explained in three differnet ways. Firstly, we have to consider the possibility that other dipeptides could be cleaved by the viral protease. This could perhaps be related to the fact that, both in TBRV and GCMV, the histidine usually found in cellular or viral cysteine proteases (44) is replaced by a leucine, possibly leading to a modified specificity of the enzyme. Indeed, the cleavage sites could

be quite original since at least the coat protein of GCMV seems to be released by cleavage of an R/A dipeptide. Secondly, contrary to the assumption we made during our search, the cleavage sites of GCMV might not be colinear with those of other picorna-like viruses. Thirdly, another viral protein could act as a protease with a different specificity, as reported in picornaviruses and recently also in potyviruses (51).

ACKNOWLEDGEMENTS

We would like to thank the members of the group of Dr C. Fritsch for the communication of unpublished results. We also thank Kathryn Mayo-Candresse for help with the English of the manuscript.

REFERENCES

1. Martelli, G.P. and Quacquarelli, A. (1972) Grapevine chrome mosaic virus. In Murant, A.F. and Harrison, B.D. (Eds), C.M.I./A.A.B. Descriptions of plant viruses, Old Woking, England, # 103.
2. Le Gall, O., Candresse, T. and Dunez, J. (1988a) Nucleotide sequence of the 3' ends of the double-stranded RNAs of grapevine chrome mosaic nepovirus. *J. Gen. Virol.* **69**, 423–428.
3. Robinson, D.J., Barker, H., Harrison, B.D. and Mayo, M.A. (1980) Replication of RNA1 of tomato black ring virus independently of RNA-2. *J. Gen. Virol.* **51**, 317–326.
4. Fritsch, C., Mayo, M.A. and Murant, A.F. (1980) *J. Gen. Virol.* **46**, 381–389.
5. Kerlan, C., and Dunez, J. (1983) Applications de l'immunoélectro-microscopie à la détection de deux souches d'un même virus. *Ann. Virol. (Institut Pasteur)* **134E**, 417–428.
6. Dodd, S.M., and Robinson, D.J. (1984) Nucleotide sequence homologies among RNA species of tomato black ring virus and other nepoviruses. *J. Gen. Virol.* **65**, 1731–1740.
7. Doz, B., Macquaire, G., Delbos, R. and Dunez, J. (1980) Caractéristiques et rôle du RNA3, RNA satellite du virus des anneaux noirs de la tomate. *Ann. Virol. (Inst. Pasteur)* **131E**, 489–499.
8. Doz, B., Delbos, R., and Dunez, J. (1982) Prémunition: une compétition entre souches faibles et souches sévères pour une voie commune d'expression des symptômes. *Les colloques de l'INRA* **11**, 29–44.
9. Bretout, C. (1987) Contribution à l'étude du phénomène de la prémunition entre deux souches du virus des anneaux noirs de la tomate (TBRV) sur *Chenopodium quinoa* Wild. PhD thesis, Université de Bordeaux II.
10. Martelli, G.P. and Quacquarelli, A. (1970) Hungarian chrome mosaic of grapevine and tomato black ring : two similar but unrelated plant viruses. In the proceedings of the 4th international conferences on viruses and virus diseases of grapevine, Colmar, France.
11. Meyer, M., Hemmer, O., and Fritsch, C. (1984) Complete nucleotide sequence of a satellite RNA of tomato black ring virus. *J. Gen. Virol.* 65, 1575–1583.
12. Meyer, M., Hemmer, O., Mayo, M.A., and Fritsch, C. (1986) The nucleotide sequence of tomato black ring virus RNA-2. *J. Gen. Virol.* **67**, 1257–1271.
13. Greif, C., Hemmer, O., and Fritsch, C. (1988) Nucleotide sequence of tomato black ring virus RNA-1. *J. Gen. Virol.* **69**, 1517–1529.
14. Morris-Krsinich, B.A.M., Forster, R.L.S., and Mossop, D.W. (1983) *Virology* 130, 523–526.
15. Goldbach, R. (1987) Genome similarities between plant and animal RNA viruses. *Microbiol. Sci.* **4**, 197–202.
16. Goldbach, R., and Wellink, J. (1988) Evolution of plus-strand RNA viruses. *Intervirology* **29**, 260–267.
17. Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.
18. Le Gall, O., Candresse, T., Brault, V., Bretout, C., Hibrand, L., and Dunez, J. (1988b) Cloning of full-length cDNA of grapevine chrome mosaic nepovirus genomic RNAs. *Gene*, **73** (1), 67–75.
19. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning : A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
20. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1987) Current protocols in molecular biology, Greene publishing associates and Wiley-interscience, New-York.
21. Lin, H.C., Lei, S.P., and Wilcox, G. (1985) An improved DNA sequencing strategy. *Anal. Biochem.* 147, 114–119.

22. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors.*Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
23. Zagursky, R.J., Baumeister, K., Lomax, N., and Berman, M.L. (1985) Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. *Gene Anal. Techn.* **2**, 89–94.
24. Fristensky, B., Lis, J., and Wu, R. (1982) Portable microcomputer software for nucleotide sequence analysis. *Nucl. Acids Res.* 10, 6451–6463.
25. Higgins, D.G., and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. 2Gene **73**, 237–244.
26. Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific cleavages. *Meth. Enzymol.* 65, 499–559.
27. Kozak, M. (1987a) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* 15, 8125–8148.
28. Kozak, M. (1987b) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* 196, 947–950.
29. Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Sceele, G.A. (1987) Selection of AUG initiation codons differs in plants and animals. *EMBO J.* 6, 43–48.
30. Lomonossof, G.P., and Shanks, M. (1983) The nucleotide sequence of cowpea mosaic virus B RNA. *EMBO J.* **2**, 2253–2258.
31. Allison, R., Johnston, R.E., and 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.
32. Nomoto, A., Omata, T., Toyoda, H., Kuge, S., Horie, H., Kataoka, Y., Genba, Y., Nakano, Y., and Imura, N. (1982) Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome. *Proc. Nat. Acad. Sci. U.S.A.* **79**, 5793–5797.
33. Lipman, D.J. and Pearson, W.R. (1985) Rapid and sensitive protein similarity searches. *Science* **227**, 1435–1441.
34. Vos, P., Verver, J., Jaegle, M., Wellink, J., Van Kammen, A, and Goldbach, R. (1988) Two viral proteins involved in the proteolytic processing of the cowpea mosaic virus polyproteins. *Nucl. Acids Res.* **16**, 1967–1985.
35. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1988a) A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.* 235, 16–24.
36. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1988b) A conserved NTP-motif in putative helicases. *Nature* **333**, 22.
37. Gorbalenya, A.E., and Koonin, E.V. (1988) One more conserved sequence motif in helicases. *Nucl. Acids Res.* **16**, 7734.
38. Hodgman, T.C. (1988) A new superfamily of replicative proteins. *Nature* **333**, 22–23.
39. Goldbach, R., Rezelman, G., Zabel, P., and Van Kammen, A. (1982) Expression of the bottom-component RNA of cowpea mosaic virus : evidence that the 60-kilodalton VPg precursor is cleaved into single VPg and a 58-kilodalton polypeptide. *J. Virol.* 42, 630–635.
40. Eggen, R., Kaan, A., Goldbach, R. and Van Kammen, A. (1988) Cowpea mosaic virus RNA replication in crude membrane fractions from infected cowpea and *Chenopodium amaranticolor*. *J. Gen. Virol.* **69**, 2711–2720.
41. Takegami, T., Semler, B.L., Anderson, C.W., and Wimmer, E. (1983) Membrane fractions active in poliovirus RNA replication contain VPg precursor polypeptides. *Virology* 128, 33–47.
42. Argos, P., Kamer, G., Nicklin, M.J.H., and 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. *Nucl. Acids Res.* **12**, 7251–7267.
43. Bazan, F. and Fletterick, R.J. (1988) Viral cysteine proteases are homologous to the trypsin-like family of serine proteases : structural and functional implications. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7872–7876.
44. Wellink, J., and Van Kammen, A., (1988) Proteases involved in the processing of viral polyproteins. *Arch. Virol.* 98, 1–26.
45. Kamer, G., and Argos, P. (1984) Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucl. Acids Res.* 12, 7269–7282.
46. Dougherty, W.G., Carrington, J.C., Cary, S.M., and Parks, T.D. (1988) Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* 7, 1281–1287.
47. Weber, H. and Weissman, C. (1970) The 3′-termini of bacteriophage Qβ plus and minus strands. *J. Mol. Biol.*, **51**, 215–224.

48. Wengler, G.,Wengler, G., and Gross, H.J. (1979) Replicative form of Semliki forest virus RNA contains an unpaired guanosine. *Nature*, **282**, 754–756.
49. Collmer, C.W. and Kaper, J.M., (1985) Double-stranded RNAs of cucumber mosaic virus and its satellite contain an unpaired guanosine : implications for replication. *Virology* **145**, 249–259.
50. Richards, O.C. and Ehrenfeld, E. (1980) Heterogeneity of the 3′ end of minus-strand in the poliovirus replicative form. *J. Gen. Virol.* **36**, 387–394.
51. Carrington, J.C., Cary, S.M., Parks, T.D. and Dougherty, W.G. (1989) A second proteinase encoded by a plant potyvirus genome. *EMBO J.* **8**, 365–370.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.