The EMBO Journal vol.5 no.2 pp.223-229, 1986

The organisation and interviral homologies of genes at the 3′ end of tobacco rattle virus RNA1

Martine Boccara[1], William D.O.Hamilton and David C.Baulcombe

Molecular Genetics Department, Plant Breeding Institute, Maris Lane, Trumpington, Cambridge CB2 2LQ, UK

[1]Present address: Laboratoire de Pathologie Vegetale I.N.A., 16 rue Claude Bernard, 75005 Paris, France

Communicated by C.Leaver

The RNA1 of tobacco rattle virus (TRV) has been cloned as cDNA and the nucleotide sequence determined of 2 kb from the 3′-terminal region. The sequence contains three long open reading frames. One of these starts 5′ of the cDNA and probably corresponds to the carboxy-terminal sequence of a 170-K protein encoded on RNA1. The deduced protein sequence from this reading frame shows homology with the putative replicases of tobacco mosaic virus (TMV) and tricornaviruses. The location of the second open reading frame, which encodes a 29-K polypeptide, was shown by Northern blot analysis to coincide with a 1.6-kb subgenomic RNA. The validity of this reading frame was confirmed by showing that the cDNA extending over this region could be transcribed and translated *in vitro* to produce a polypeptide of the predicted size which co-migrates in electrophoresis with a translation product of authentic viral RNA. The sequence of this 29-K polypeptide showed homology with two regions in the 30-K protein of TMV. This homology includes positions in the TMV 30-K protein where mutations have been identified which affect the transport of virus between cells. The third open reading frame encodes a potential 16-K protein and was shown by Northern blot hybridisation to be contained within the region of a 0.7-kb subgenomic RNA which is found in cellular RNA of infected cells but not virus particles. The many similarities between TRV and TMV in viral morphology, gene organisation and sequence suggest that these two viral groups may share a common viral ancestor.

Key words: subgenomic RNA/tobacco mosaic virus/tobacco rattle virus/transport protein/viral evolution

Introduction

Tobacco rattle virus (TRV) is a soil-borne agent which infects a wide range of plants (Harrison and Robinson, 1978), including potato, in which it is an important pathogen. The genome consists of two RNA species (RNA1 and RNA2) which in strain SYM (Kurppa *et al.*, 1981), are 6.8 kb and 3.9 kb in length (Robinson *et al.*, 1983). RNA1 translates *in vitro* to produce two polypeptides of mol. wt 140 000 and 170 000 (Fritsch *et al.*, 1977). It is likely that the latter is produced by translation through the termination codon of the 140 000-dalton protein (Pelham, 1979). A third peptide of mol. wt 30 000 (approximately) is also encoded by RNA1, but is translated principally from a 1.5-kb subgenomic RNA, RNA4 (Robinson *et al.*, 1983).

RNA2 encodes the coat protein which is translated from a subgenomic RNA, RNA3 (Robinson *et al.*, 1983), and it now seems unlikely that RNA2 encodes other functions, as sequence analysis of TRV (strain CAM) shows only a single open reading frame (A.Siegel, personal communication). It is clear at least that viral RNA replication, symptom production and cross-protection between strains of TRV (Harrison and Robinson, 1978) are all RNA1 functions as these effects may all be produced in NM-type infections which occur naturally and which lack RNA2 (Harrison and Robinson, 1978).

As a step towards understanding the biology of TRV RNA1, a relatively simple pathogenic agent, we are characterising the gene products by sequencing. Here we present the entire coding sequence of the subgenomic RNA region of RNA1 and an additional 600 nucleotides immediately 5′ of that region. The deduced protein sequences identify three products from this region, two of which have significant homology with gene products in tobacco mosaic virus (TMV). The gene product from the 5′ region has homology also with products of alfalfa mosaic virus (AMV) and brome mosaic virus (BMV). The third gene product has no homology with gene products of these viruses and is synthesised on a subgenomic RNA which is not encapsidated in virus particles.

Results

Characterisation of the cDNA from the 3′ region of TRV-RNA1

From several cDNA clones of TRV-RNA1 a set of six were isolated which were determined, by restriction enzyme analysis and by cross-hybridisation, to form an overlapping set covering 6.8 kb, which is close to the length of RNA1 (Robinson *et al.*, 1983). One of these clones, 25B, included 2.0 kb of sequence at one end of this overlapping set of clones. This end was identified at the 3′ end of the RNA1 by showing hybridisation to fragments of RNA1 which had been labelled at the 3′ terminus. It was also shown (Boccara, unpublished) that this clone hybridised to RNA isolated from VS virion particles which are enriched for subgenomic RNAs (Robinson *et al.*, 1983). A second cDNA clone, 543B, was mapped by hybridisation and restriction enzyme sites to the 3′ region of 25B and extended the 3′ sequence in 25B by 200 bp.

Given the length of the overlapping set of cDNAs and the hybridisation results with 25B and 543B, it was concluded that these clones cover the sequence close to the 3′ terminus of RNA1 and include the sequence of the 3′-terminal subgenomic RNA.

Sequencing of the cDNA clones

The nucleotide sequence of the cDNA clones 25B and 543B was determined using the dideoxy chain termination method on M13 templates generated by cloning either random fragments or specific restriction fragments. The DNA sequence equivalent of the viral strand and the deduced polypeptide sequences are shown on Figure 1. The sequence contains three long open reading frames. The first is a continuation of a reading frame initiated outside 25B and finishes at nucleotide 609. A second long open reading frame commences at ATG in position 846 and continues to position 1604, corresponding to a polypeptide of 252 amino

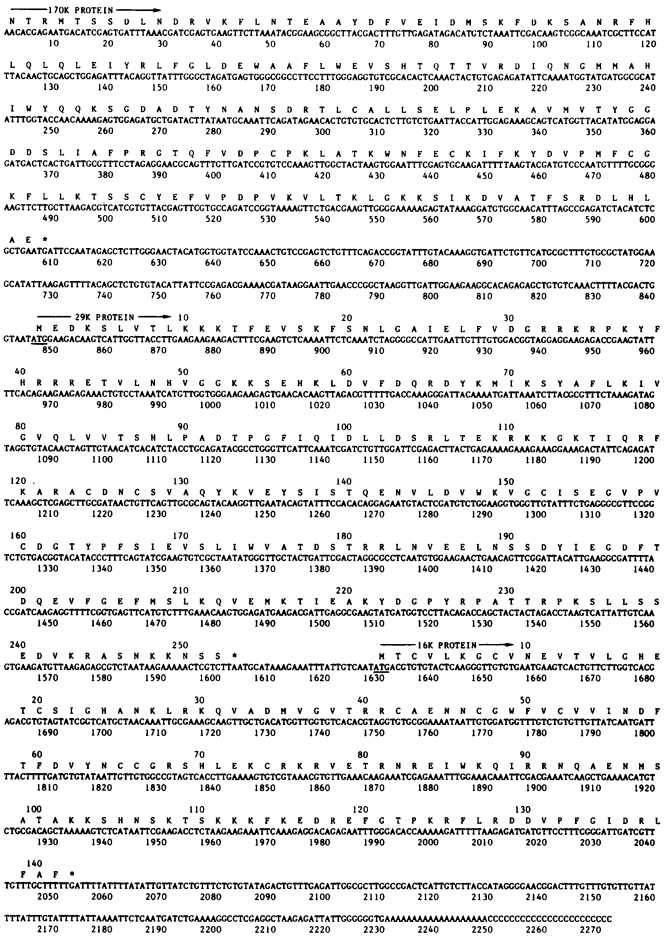

Fig. 1. The nucleotide sequence of the 3′ terminus region of TRV RNA1. The DNA sequence is shown as the equivalent of the viral RNA strand, numbered from the beginning of the cDNA. The three largest polypeptide sequences deduced from this sequence are shown using the single letter amino acid code and numbered from the initiating methionine.

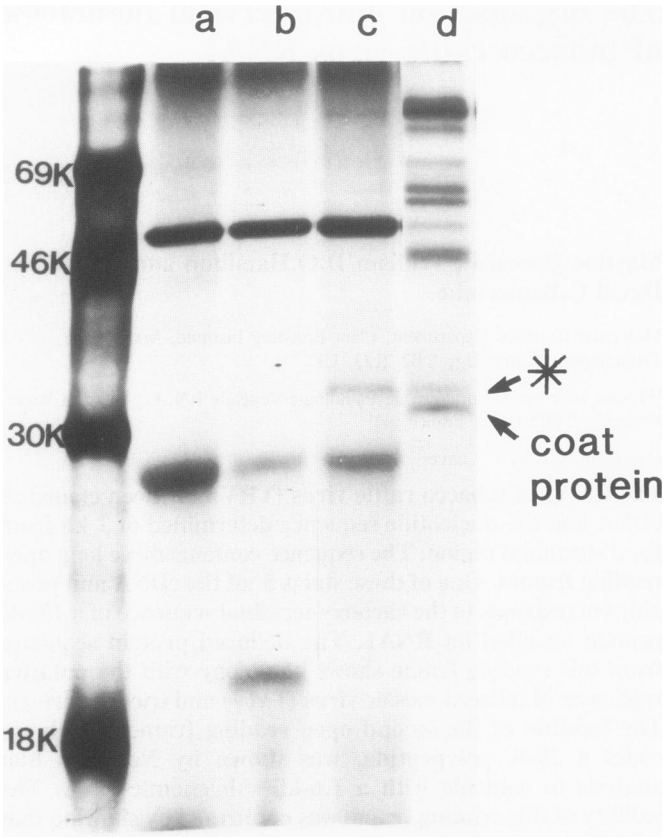

Fig. 2. *In vitro* translation of viral RNA and RNA synthesised *in vitro* from viral cDNA. The RNA samples were translated *in vitro* in the presence of [^{35}S]-methionine and fractionated by electrophoresis in SDS-polyacrylamide gels. The radioactive proteins were detected by fluorography. The RNA samples were (**a**) no RNA, (**b**) RNA transcribed *in vitro* from a cDNA clone of ribulose biphosphate carboxylase, (**c** RNA transcribed *in vitro* from a fragment of 25B extending over the 29-K protein reading frame, (**d** TRV RNA isolated from virus particles. The size markers were lactoglobulin (18 K), carbonic anhydrase (30 K), ovalbumin (46 K) and bovine serum albumin (69 K). The product of viral RNA which co-migrates with the product of RNA transcribed *in vitro* from 25B is indicated (*).

acids (mol. wt 28 826). A third long open reading frame starts from a methionine (ATG) codon at position 1629. This reading frame terminates at position 2054 and encodes a protein of 141 amino acids (mol. wt 16 337). In subsequent discussion the putative products of these reading frames are described as the 29-K and 16-K proteins.

The sequence to the 3′ of the gene for the 16-K protein contains no long open reading frames. At the 3′ terminus the oligo(dA) and oligo(dG) sequences are those added to the RNA 3′ end and plasmid vector during the cDNA cloning.

Transcription in vitro of the 29-K sequence with SP6 RNA polymerase and in vitro translation

To confirm the validity of the open reading frames and compare the protein product with the product of TRV subgenomic RNAs, the cDNA sequence was transcribed into RNA and translated into protein *in vitro*. The transcription was carried out following insertion of the cDNA sequence between nucleotide 589 (*Bgl*II site) and the 3′ end of the cDNA clone (*Sal*I site in pUC9 polylinker) into plasmid pSP65 and using SP6 RNA polymerase (Melton *et al.*, 1984). The RNA product of this reaction was translated in a rabbit reticulocyte lysate containing ^{35}S-labelled methionine and the products, analysed by SDS electrophoresis, are shown in Figure 2. The products of the reaction containing the RNA product of the 25B (Figure 2c) include two components of mol. wt 50 000 and 28 000 which are also produced in the absence of exogenous RNA (Figure 2a) and a third peptide with an apparent mol. wt of 35 000, corresponding to the predicted 29-K polypeptide. This polypeptide co-migrated with a product of TRV RNA (Figure 2d) which was identified previously as the product of a subgenomic RNA encoded by RNA1. The translation products of TRV contained also a smaller peptide (apparent mol. wt 32 000) which is probably the coat protein product of RNA2 (Robinson *et al.*, 1983). The migration of these proteins in SDS-polyacrylamide gels is slower than predicted from the nucleotide sequence, but is similar to that shown by Robinson *et al.* (1983) for the products of the subgenomic RNAs in virus particles of TRV (strain SYM).

In these experiments the efficiency of transcription and translation of DNA into protein was low, as indicated by the prominence of the endogenous translation products in Figure 2c. This is likely to be a result of the techniques used rather than an intrinsic feature of the RNA because a polypeptide production from a construction in pSP65 containing the sequence for ribulose bisphosphate carboxylase small subunit from wheat was similarly inefficient (Figure 2b, Douglas Bradley, unpublished).

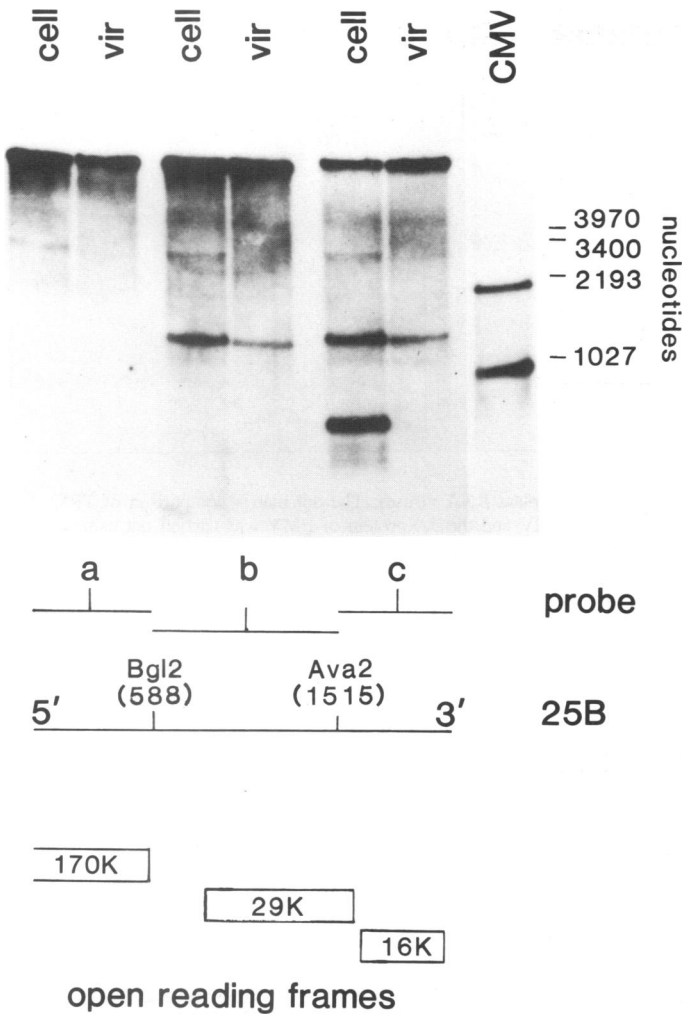

Fig. 3. Northern blot analysis of subgenomic RNAs. Total cellular RNA (5 μg) from TRV-infected tissue (cell) or virion RNA (vir) (50 ng) was fractionated by electrophoresis in a 1.2% agarose gel and hybridised with one of three different subcloned fragments (**a, b, c**) from 25B as shown in the diagram. 50 ng of cucumber mosaic virus RNA was fractionated on the same gel and detected after blotting by hybridisation with a cloned CMV probe. The size values of CMV RNAs are those given by Gould and Symons (1983).

Subgenomic RNA in the 3' region of RNA1

Translation *in vitro* of virion RNA showed no evidence for the production of the putative 16-K protein. In order to detect any subgenomic RNA which is not encapsidated and which might serve as translation template for the 16-K protein, Northern blot analysis was carried out using fragments of 25B as probe. With virion RNA this showed genome length RNA and a single abundant subgenomic species of length 1.6 kb (Figure 3a–c). However, in total cellular RNA from infected tissue, two subgenomic RNAs were observed; the 1.6-kb species and a 0.7-kb RNA (Figure 3c). The identification of these as subgenomic RNAs, rather than degradation products of genome RNA, is based on two criteria. First, the two RNAs showed specific hybridisation to regions of 25B, as would be predicted for the subgenomic mRNAs of the 29-K and 16-K proteins. Thus the 1.6-kb mRNA showed hybridisation only with probes extending 3' of the start of 29-K open reading frame (Figure 3b,c). The 0.7-kb mRNA hybridised only with the probes containing the sequence of the 16-K open reading frame (Figure 3c). Second, the failure to detect subgenome length RNAs with a probe extending to the 5' of the 29-K reading frame (Figure 3a) confirmed the lack of degradation in the RNA preparations.

Homology between the proteins of TRV and TMV

There are now several examples where similarity has been detected between proteins of viruses from different taxonomic groups (Murthy, 1983; Cornelissen *et al.*, 1984; Cornelissen and Bol, 1984; Haseloff *et al.*, 1984). The dot matrix computer program DIAGON has been used to detect any such similarities involving the TRV polypeptides deduced from the nucleotide sequence in the 3' region of RNA1. The program compares all possible combinations of 31 adjacent amino acids on two sequences and registers homology on a matrix diagram if the proportional match between the sequences is greater than a pre-set score. This score is influenced by the type of amino acids involved in the match and accounts for perfect matches and matches of amino acids with similar charge or shape.

The 29-K protein. DIAGON comparisons were made of the TRV 29-K protein sequence with proteins of similar mol. wt from TMV, AMV and BMV, but only with TMV were extensive matches recorded (Figure 4). At the score of 340 and considering two random sequences of the same length and the same amino acid composition, matches would be registered at a frequency of 10^{-4}. In fact, for the comparison between TMV 30-K and TRV 29-K sequences, matches were detected at six times this frequency and were grouped in four regions (Figure 4). Detailed inspection of the matches indicated that in only two of the four regions were there matches of identical amino acids in the TRV and TMV sequence. These two regions of similarity are included between amino acids 78 and 170 of the TRV protein and 69 and 160 of the TMV protein. Sequences from these regions of TRV and three different strains of TMV are aligned in Figure 5 and show that the homology in this region is almost as great between TRV and TMV as between different strains of TMV. Thus, there are 25 conserved residues between the vulgare strain of TMV (Goelt *et al.*, 1982) and the cowpea strain (Meshi *et al.*, 1982). This compared with respectively 24 and 25 conserved positions between TRV and these two strains. Sequence is available for part of this region in a third strain of TMV, cucumber green mottle mosaic virus (Meshi *et al.*, 1983), and shows 15 identical positions with TMV (vulgare strain) and 10 with TRV.

The 5' polypeptide of 25B. The translation products of RNA from TRV strain SYM include a polypeptide of mol. wt 170 000 (Robinson *et al.*, 1983) which is translated from the 5' of RNA1 (Pelham, 1979). It seemed likely therefore that the polypeptide product of the open reading frame at the 5' end of 25B included the C-terminal part of this peptide. The protein sequence of the 5' polypeptide was compared with the 183-K polypeptide product of TMV and the products of RNA2 in AMV and BMV. Extensive homology was observed in all three of these cases in a region which has been reported previously to be similar between AMV, BMV and TMV (Figure 6) (Cornelissen and Bol, 1984; Haseloff *et al.*, 1984). There were 71 amino acid matches with TMV, 52 with BMV and 58 with AMV. In the same regions the TMV protein showed 53 amino acid matches with AMV and 51 with BMV. The AMV and BMV proteins showed 79 matches.

The 16-K protein. The protein sequence of the 16-K protein was compared, using the DIAGON program, to the gene products of RNA3 from cucumber mosaic virus (Gould and Symons, 1982),

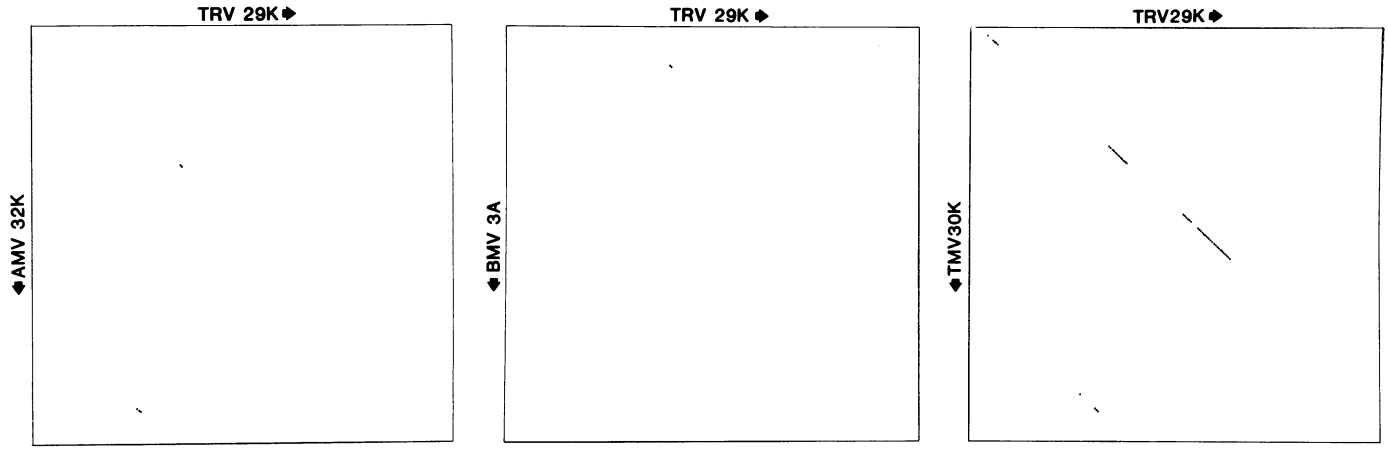

Fig. 4. Dot matrix comparison of TRV 29-K protein sequence with protein sequences from other plant RNA viruses. The dot matrix comparison of TRV 29-K protein sequence with the 30-K protein sequence of TMV, the 32-K protein sequence of AMV and the 3A protein of BMV was carried out using the DIAGON program of Staden (1982). The span length of each comparison was 31 amino acids and the score limit below which matches are not recorded, was 340.

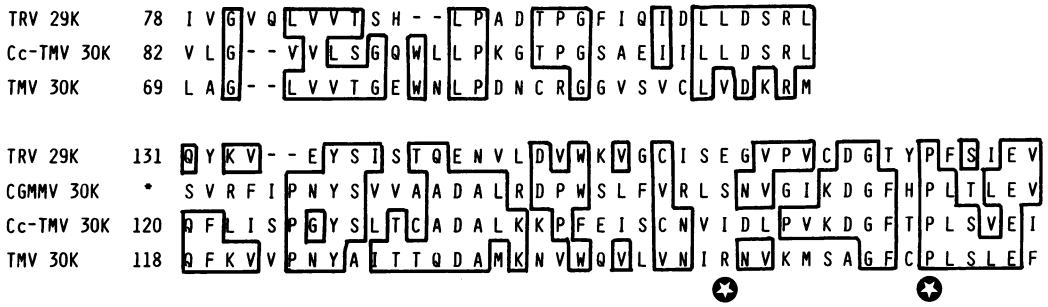

Fig. 5. The sequence of the 29-K protein of TRV aligned with sequences of the TMV 30-K protein. The sequences aligned are those from the two strongest homologies shown in Figure 4. The position of the N-terminal amino acid in each alignment is indicated. Positions where two or more sequences have the same amino acid are boxed. The sequences are TRV 29-K protein and the 30-K proteins of TMV strains vulgare (Goelet *et al.*, 1982) (TMV), cowpea (Zimmern, 1983) (Cc-TMV) and cucumber green mottle virus (Meshi *et al.*, 1983) (CGMMV). The sites of mutations in TMV strains LS1 and NI2519 (see Discussion) are indicated (*).

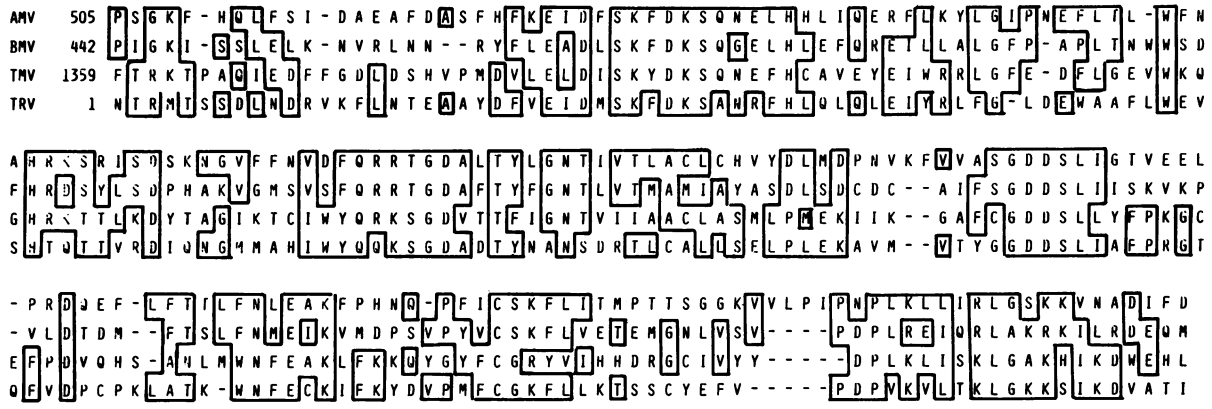

Fig. 6. The sequence of the protein in the 5′ reading frame of 25B aligned with protein sequences from other plant viruses. The protein sequences of the 183-K protein of TMV and the protein products of RNA2 of AMV and BMV are shown aligned to the sequence deduced from the 5′ reading frame of 25B. The position of the N-terminal amino acid in each sequence is indicated by the numbers at the beginning of each sequence. Positions where two or more sequences have the same amino acid are boxed.

the products of open reading frames in TMV (Goelet *et al.*, 1982), of BMV RNA3 (Ahlquist *et al.*, 1984), of AMV (Barker *et al.*, 1983) and of cowpea mosaic virus (Lomononosoff and Sharks, 1983; van Wezenbeck *et al.*, 1983). In no case was significant homology found.

Discussion

The nucleotide sequence of the 3′ region of TRV RNA1 contains three long open reading frames. These correspond to the coding sequence for two proteins identified previously by *in vitro*

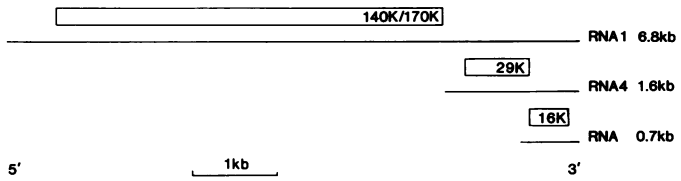

Fig. 7. A genetic map of TRV RNA1. The size and approximate position of the genes for the 140-K and 170-K readthrough proteins is as inferred from earlier translation analysis and from the termination codon in the sequence of 25B. The location of the other genes and subgenomic RNAs is described in this paper.

translation of viral RNA, and an additional 16-K protein which is translated from a non-encapsidated mRNA (Figure 3). A revised genetic map for TRV-RNA 1 (strain SYM) incorporating this new information is in Figure 7. The positions of the subgenomic RNAs in this map can be assigned only approximately from the data presented here. However, the position shown for the 1.6-kb mRNA is consistent with hybridisation data and fits well with the observed size of *in vitro* translation product from this RNA (Figure 2). Other locations for the 5′ end of this RNA would require that translational initiation occurs at an internal methionine, or would not allow production of a protein of the observed size. The location of the 0.7-kb mRNA as shown in Figure 7 is consistent with the hybridisation data, the position of the 16-K open reading frames and predicts that the 0.7-kb RNA is 3′ co-terminal with the 1.6-kb RNA.

The 5′ ends of the subgenomic RNA species also delimit the possible position of the origin of packaging which must be located within the 1.6-kb subgenomic RNA region but outside the 0.7-kb region. The internal location of this feature suggests a similar mode of viral RNA packaging to that of TMV.

The sequence located 3′ of the 16-K protein termination codon contains no long open reading frames. Based on comparison of the distance of the 3′ terminus from the start of the 16-K and 29-K open reading frames with the length of the subgenomic RNAs it is likely that the 3′ of the sequence presented here is close to the 3′ end of RNA1. The presence of the oligo(A) sequence is consistent with the cDNA sequence extending to the end of RNA1, since presumably the poly(A) polymerase would extend from a terminal 3′ OH residue. The products of RNA hydrolysis would produce a 3′ phosphate residue such that degraded RNA would not serve as a substrate for this enzyme. However, the terminal nucleotides differ from the $CCC_{(OH)}$ sequence determined by direct RNA sequencing on TRV strain CAM (Gugerli *et al.*, 1978) and further analysis will be necessary to show whether this sequence includes the true 3′ terminus of TRV RNA1.

This sequence analysis of RNA1 suggests that a maximum of four TRV-encoded proteins are necessary for pathogenesis. These are the 140-k and 170-K proteins (Pelham, 1979) and the 29-K and 16-K proteins described here. However, in TMV a larger number of gene products may be produced by translational initiation at sites within the three main open reading frames (Goelet and Karn, 1982; Hunter *et al.*, 1983; Sulzinski *et al.*, 1985) from specific subgenomic RNAs. The Northern blot analysis (Figure 3) suggests that this strategy is not used with TRV, as multiple discrete subgenomic length RNA species are not detected. If present they are certainly at a much lower abundance than in TMV (Goelet and Karn, 1982).

Sequence homologies between TRV and other viruses (Figures 5 and 6) imply particular functions for two open reading frames on TRV RNA1. As the gene products of AMV and BMV RNA2 are implicated in viral RNA synthesis (Kiberstis *et al.*,1981; Nassuth and Bol, 1983) it seems likely that the related 170-K protein has a similar role. The 29-K protein of TRV is related to the 30-K protein of TMV which, based on studies with two mutant strains of TMV, is thought to have a role in the cell-cell transport of virus (Ohno *et al.*, 1983; Zimmern and Hunter, 1983). In both of these mutants there are single base changes within the conserved TMV/TRV region. As a result of this, the mutant LS1 has a serine at position 153 which is proline in all other strains of TMV, and in TRV. The second mutant strain Ni 2519 is changed from arginine to glycine at amino acid 144.

The analysis of TMV does not show conclusively that the 30-K protein is a transport protein as other mutations responsible for the viral phenotype may be located in unsequenced regions of the viral genome. However, the clear homology between TMV and TRV in the mutated regions supports the idea of the postulated functional importance of those regions. By analogy with TMV this suggests a role for the 29-K protein of TRV in the cell to cell movement of the virus. This is the first report of homology between TMV and other groups of virus in genes not associated with the replicase function.

Precisely how the 29-K/30-K proteins might mediate viral transport is unknown. However, the occurrence of such a protein in TRV, where systemic spread of infection occurs in non-encapsidated RNA (NM) strains, suggests that a direct 29-K protein/viral RNA interaction is involved. The conclusion that the 30-K protein of TRV (CAM strain), which is equivalent to the 29-K protein and encoded by RNA2 (Bisaro and Siegel, 1980), would contradict this but is now known to be invalid as RNA2 encodes only coat protein (A.Siegel, personal communication). It will be possible now to prepare 29-K protein in large quantities using microbial expression systems and assay for specific protein/RNA binding *in vitro*.

It is interesting that the shared homology between TRV 29-K protein overlaps with the 'LAGLI' and other peptide regions shared between the TMV 30-K protein and the protein encoded by the 4th intron of the yeast mitochondrial apocytochrome b gene (Zimmern, 1983). However, no homology is detected between the yeast gene and the TRV 29-K protein and this must call into question the functional significance of the homology with the TMV 30-K protein.

Although sequence homologies do not suggest a function for the 16-K protein product of RNA1, other observations suggest a role in nematode transmissibility. In discussing the high degree of sequence and antigenic variation between coat proteins of different tobravirus strains, Harrison and Robinson (1986) suggest that the variation rules out a role of coat protein in determining vector transmissibility and specificity. This function is therefore encoded by default by RNA1 and consequently may be mediated by the 16-K protein. None of the other RNA1-encoded genes, on present knowledge, is likely to play a role in vector transmission.

These analyses of the 3′ end of the TRV RNA1 suggest that tobraviruses represent an intermediate viral form between the tricornaviruses and tobamoviruses. The resemblance with tricornaviruses is based on the split genome organisation and the homologies between the sequence of the peptide encoded in the 5′ segment of the clone 25B with the protein product of RNA2. A stronger resemblance to the tobamovirus group is based upon the rod-like virion structure, from the similar behaviour of TMV mutant PM2, which produces no coat protein (Siegel *et al.*, 1962), and TRV NM strains, and from the polypeptide sequences and gene organisation in the region covered by clone 25B. Thus, the

5′ open reading frame of 25B, which presumably constitutes the 3′ end of the gene for the TRV 170-K protein, is in the same relative position (Figure 7) on RNA1 as its TMV homologue, the gene for the 183-K protein. The 29-K gene is immediately 3′ of this and also in the same relative position on the genome as the 30-K TMV homologue. The resemblance with TMV also extends to the strategy of translation of the 29-K/30-K proteins which are both derived from subgenomic RNAs and of the 170-K/183-K protein which is likely to result from translation through the 140-K/120-K protein termination signal of TRV and TMV (Pelham, 1979). However, the reading frames of the 29-K and 170-K proteins do not overlap in TRV as is the case in TMV, and this would seem to be another example, besides the gross genome organisation where genes are more widely separated in TRV than in TMV.

The extension in this report of observed homologies between genes of different viral group (Cornelissen and Bol, 1984; Cornelissen *et al.*, 1984; Haseloff *et al.*, 1984) to include an additional viral group (tobravirus), and another gene (29-K/30-K) gives additional support to the view that these genes had a common evolutionary origin. This does not necessarily distinguish whether the common ancestral forms were in an archaic virus or were cellular genes (Haseloff *et al.*, 1984). However, given the additional similarities in viral structure and gene organisation shared by tobamoviruses and tobraviruses it is likely that these two classes of virus have diverged from a common viral ancestor. In the course of this evolution the gene located 3′ to the 29-K/30-K gene has changed in one or both of these viral classes, perhaps by recombination of RNA molecules (King *et al.*, 1982).

Materials and methods

Preparation of RNA

TRV strain SYM (Kurppa *et al.*, 1981) was propagated in *Nicotiana clevelandii*. RNA was isolated from purified and fractionated virion particles and was provided by D.Robinson (Scottish Crop Research Institute). Total cellular RNA was isolated from *N. tabacum* var Samsun NN at 10 days post-inoculation by the method of Apel and Kloppstech (1978).

Construction of cDNA clones

TRV RNA1 was polyadenylated *in vitro* using poly(A) polymerase (Bethesda Research Laboratories) and cDNA synthesis primed with oligo(dT) was according to Maniatis *et al.* (1982), but in the presence of 4 mM sodium pyrophosphate. The DNA/RNA hybrid product of this reaction (1 μg) was fractionated on a 5 ml sucrose gradient in 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM EDTA by centrifugation at 50 000 r.p.m. for 2 h. The largest molecules of cDNA (~30% of the preparation) were recovered by ethanol precipitation and tails of oligo(dC) were added with terminal transferase (14 U, BRL). The reaction was carried out according to Maniatis *et al.* (1982).

The vector plasmid (20 μg) was cut with *Pst*I and tailed with oligo(dG) according to Maniatis *et al.* (1982). The tailed vector (50 ng) and cDNA (50 ng) were annealed in 100 mM NaCl, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA from 65°C to 4°C overnight and transformed into calcium-treated (Maniatis, 1982) *Escherichia coli* strain MC1022 (Casadaban and Cohen, 1980). Clones containing the largest inserts (1.3–2.3 kb) were identified and selected for further investigation. Plasmid DNA was prepared from these clones by the method of Hansen and Olsen (1978).

DNA sequencing

DNA fragments from TRV RNA1 cDNA clones were transferred to the mp derivatives of M13 (Messing, 1983) and the nucleotide sequence determined by the dideoxy termination method of Sanger *et al.* (1977). Most of the sequence was generated by shotgun sequencing of random fragments from the cDNA inserts (Anderson, 1981). Specific restriction fragments were cloned in M13 vectors and sequenced to ensure that the sequence was determined on both strands. The sequence data was compiled using the computer programs of Staden (1982).

Transcription in vitro with SP6 RNA polymerase and in vitro translation

A *Bgl*II/*Sal*I restriction fragment from cDNA clone 25B was transfefred into the *Bam*HI/*Sal*I sites of pS65 (Melton *et al.*, 1984). The derivative plasmid (20 μg) was linearised by digestion with *Hind*III which cuts 3′ of the insert relative to the SP6 promoter, extracted with phenol, precipitated with ethanol and transcribed with SP6 RNA polymerase. The reaction (50 μl) contained 40 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$, 10 mM dithiothreitol, 4 mM spermidine, 0.5 mM of each of the four ribonucleoside triphosphates, 5 μg template DNA, BSA 1 mg/ml, 20 units RNasin (BRL), 7.5U SP6 polymerase (New England Nuclear) and 0.5 mM m7GpppGOH (Ahlquist and Janda, 1984). The RNA was processed as described by Krieg and Melton (1984) and translated *in vitro* in a rabbit reticulocyte lysate containing [^{35}S]methionine (Jackson and Hunt, 1983). The radiolabelled proteins were fractionated by polyacrylamide gel electrophoresis in the presence of SDS (Laemmli, 1971) and detected by fluorography (Bonner and Laskey, 1974).

Gel electrophoresis of RNA hybridisation analysis

RNA was fractionated by electrophoresis in agarose gels containing formaldehyde as described by Rave *et al.* (1979). The gels were blotted onto nitrocellulose and hybridised with nick-translated probes, as described by Baulcombe and Buffard (1983).

Protein sequence comparisons

The DNA sequences of the viral strand were translated and the protein sequences compared using the dot matrix DIAGON program (Staden, 1982) run on a VAX computer.

Acknowledgements

We thank Bryan Harrison and Mike Mayo for discussions, David Robinson for discussion and RNA, and David Zimmern for assistance and advice on the analysis of protein homology. This work was supported by the Rockefeller Foundation. M.B. held an award from the Royal Society while on sabbatical leave at the Plant Breeding Institute.

References

Ahlquist,P. and Janda,M. (1984) *Mol. Cell. Biol.*, **4**, 2876–2882.
Ahlquist,P., Dasgupta,R. and Kaesberg,P. (1984) *J. Mol. Biol.*, **172**, 369–383.
Anderson,S. (1981) *Nucleic Acids Res.*, **9**, 3015–3027.
Apel,K. and Kloppstech,K. (1978) *Eur. J. Biochem.*, **85**, 581–588.
Barker,R.F., Jarvis,N.P., Thompson,D.V., Loesch-Fries,L.S. and Hall,T.C. (1983) *Nucleic Acids Res.*, **11**, 2881–2891.
Baulcombe,D.C. and Buffard,D. (1983) *Planta (Berl.)*, **157**, 493–501.
Bisaro,D. and Siegel,A. (1980) *Virology*, **107**, 194–201.
Bonner,W.M. and Laskey,R.A. (1974) *Eur. J. Biochem.*, **46**, 83–88.
Casadaban,M.J. and Cohen,S.N. (1980) *J. Mol. Biol.*, **138**, 179–207.
Cornelissen,B.J.C. and Bol,J.F. (1984) *Plant Mol. Biol.*, **3**, 379–384.
Cornelissen,B.J.C., Janssen,H., Zuidema,D. and Bol,J.F. (1984) *Nucleic Acids Res.*, **12**, 2427–2437.
Fritsch,C., Mayo,M.A. and Hirth,L. (1977) *Virology*, **77**, 722–732.
Goelet,P. and Karn,J. (1982) *J. Mol. Biol.*, **154**, 541–550.
Goelet,P., Lomononosoff,G.P., Butler,P.J.G., Akam,M.E., Gait,M.J. and Karn,J. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5818–5822.
Gould,A.R. and Symons,R.H. (1982) *Eur. J. Biochem.*, **126**, 217–226.
Gugerli,P., Darby,G. and Minson,A.C. (1978) *J. Gen. Virol.*, **38**, 273–280.
Hansen,J.B. and Olsen,R.H. (1978) *J. Bacteriol.*, **135**, 227–238.
Harrison,B.D. and Robinson,D.J. (1978) *Adv. Virus Res.*, **23**, 25–77.
Harrison,B.D. and Robinson,D.J. (1986) In van Regenmortel,M.H.V. (ed.), *Rod-shaped Plant Viruses*. Plenum Press, New York, in press.
Haseloff,J., Goelet,P., Zimmern,D., Ahlquist,P., Dasgupta,R. and Kaesberg,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4358–4362.
Hunter,T., Jackson,R. and Zimmern,D. (1983) *Nucleic Acids Res.*, **11**, 801–821.
Jackson,R.J. and Hunt,T. (1983) *Methods Enzymol.*, **96**, 50–74.
Kiberstis,P.A., Loesch-Fries,L.S. and Hall,T.C. (1981) *Virology*, **112**, 804–808.
Kurppa,A., Jones,A.T., Harrison,B.D. and Bailiss,K.W. (1981) *Ann. Appl. Biol.*, **98**, 243–254.
King,A.M.Q., McCahon,D., Slade,W.R. and Newman,J.W.I. (1982) *Cell*, **29**, 921–928.
Krieg,P.A. amd Melton,D.A. (1984) *Nucleic Acids Res.*, **12**, 7057–7070.
Laemli,U.K. (1970) *Nature*, **227**, 681–685.
Lomononosoff,G.P. and Shanks,M. (1983) *EMBO J.*, **2**, 2253–2258.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Meshi,T., Ohno,T. and Okada,Y. (1982) *Nucleic Acids Res.*, **10**, 6111–6117.
Meshi,T., Kiyama,R., Ohno,T. and Okada,Y. (1983) *Virology*, **127**, 54–64.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Murthy,M.R.N. (1983) *J. Mol. Biol.*, **168**, 469–475.
Nassuth,A. and Bol,J.F. (1983) *Virology*, **124**, 75–85.

Ohno,T., Takamatsu,N., Meshi,T., Okada,Y., Nishiguchi,M. and Kiho,Y. (1983) *Virology*, **131**, 255–258.
Pelham,H.R.B. (1979) *Virology*, **97**, 256–265.
Rave,N., Crkvenjakov,R. and Boedtker,H. (1979) *Nucleic Acids Res.*, **6**, 3559–3567.
Robinson,D.J., Mayo,M.A., Fritsch,C., Jones,A.T. and Raschke,J.H. (1983) *J. Gen. Virol.*, **64**, 1591–1599.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Siegel,A., Zaitlin,M. and Sehgal,O.P. (1962) *Proc. Natl. Acad. Sci. USA*, **48**, 1845–1851.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951–2961.
Sulzinski,M., Gabard,K., Palukaitis,P. and Zaitlin,M. (1985) *Virology*, **145**, 132–140.
van Wezenbeek,P., Verver,J., Harmsen,J., Vos,P. and van Kammen,A. (1983) *EMBO J.*, **2**, 941–946.
Zimmern,D. (1983) *J. Mol. Biol.*, **171**, 345–352.
Zimmern,D. and Hunter,T. (1983) *EMBO J.*, **2**, 1893–1900.

Received on 22 October 1985

Note added in proof

The nucleotide sequence of TRV (strain CAM) RNA2, referred to in this paper, is now published [Bergh,S.T., Koziel,M.G., Huang,S.-C., Thomas,R.A., Gilley,D.P. and Siegel,A. (1985) *Nucleic Acids Res.*, **13**, 8507–8518].