Tomato aspermy virus has an evolutionary relationship with other tripartite RNA plant viruses

David O'Reilly,²* Chris J. R. Thomas² and Robert H. A. Coutts¹

¹Department of Biology, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BB and ²Advanced Technologies (Cambridge) Ltd, Cambridge Science Park, Cambridge CB4 4WA, U.K.

The entire RNA 3 (2214 nucleotides) of a chrysanthemum isolate of tomato aspermy virus (C-TAV) has been cloned and its sequence determined. C-TAV possesses two open reading frames which encode a 3a protein (277 amino acids) and a coat protein (229 amino acids). Computer-assisted comparisons were made between C-TAV RNA 3 and its predicted protein sequences and those of two other tripartite RNA viruses, cucumber mosaic virus (CMV) and brome mosaic virus (BMV). Results from this study suggest that a close evolutionary relationship exists between C-TAV, Q-CMV and BMV. Divergence of nucleotide and amino acid sequences between these viruses is not reflected at the level of the predicted secondary structure of the encoded proteins, where conservation is strong.

Introduction

Tomato aspermy virus (TAV), like the type member of the cucumovirus group, cucumber mosaic virus (CMV), contains a tripartite, positive-sense ssRNA genome. The TAV genomic RNAs 1, 2 and 3 have $M_r$ of $1.26 \times 10^6$, $1.10 \times 10^6$ and $0.90 \times 10^6$ respectively (Habili & Francki, 1974a, b). Although CMV and TAV are distantly related antigenically (Devergne & Cardin, 1975; Rao et al., 1982) and there is very little sequence identity between their RNAs as determined by hybridization analysis (Gonda & Symons, 1978), they can form infectious pseudorecombinants (Habili & Francki, 1974a, b; Chen & Francki, 1990). CMV shares a number of basic features with other tripartite plant viruses which include (i) similarly sized RNAs encoding translation products of similar size; (ii) an untranslated 3'-terminal region conserved between the four RNAs of each virus; (iii) 5' m7G caps on all RNAs; (iv) a dicistronic RNA 3 encoding the 3a gene product at the 5' end and having a 3'-proximal cistron encoding the coat protein which is expressed via a subgenomic mRNA, RNA 4; and (v) only the largest three viral RNAs are required for infection. Computer-assisted comparisons (Davies & Symons, 1988) have suggested that an ancestral relationship exists between the 3a and coat proteins of the tripartite viruses, Q-CMV (Queensland strain) and brome mosaic virus (BMV), which has its genome organized in a similar fashion to CMV (Ahlquist et al., 1981). Little is known, however, of the genomic sequences of TAV. Only the RNA sequences of the 3' termini of two of the RNAs from two isolates of TAV, N- and V-TAV, have previously been determined by direct RNA sequencing (Wilson & Symons, 1981). In this paper the complete nucleotide sequence for RNA 3 of a chrysanthemum isolate of the virus, C-TAV, is reported. C-TAV RNA 3 is shown to be similar in genomic organization to Q-CMV and BMV and their evolutionary relationship is confirmed.

Methods

Virus isolate. The TAV strain used for these studies was originally isolated from naturally infected chrysanthemum and propagated in Nicotiana clevelandii. The virus was purified according to the method of Hollings & Stone (1971) and finally fractionated on 10 to 40% (w/w) sucrose gradients.

RNA extraction and fractionation. Purified virus was solubilized in the presence of 2% SDS for 10 min at 25 °C and electrophoresed on 1:2% agarose gels. Individual RNAs were eluted onto NA45 DEAE paper (Schleicher & Schuell) and further purified according to the method described by the manufacturers for RNA.

cDNA synthesis and cloning. Fractionated TAV RNA 3 was polyadenylated in vitro using poly(A) polymerase (BRL) according to the procedure of Sippel (1973). Polyadenylated TAV RNA was converted to ds cDNA, essentially as described by Haymerle et al. (1986). First-strand reactions were primed with oligo(dT) or random hexamers (Pharmacia). Double-stranded cDNAs were ligated into SalI-digested pBluescript M13 + KS (Stratagene) and the products were used for transformation of competent Escherichia coli XL1-Blue cells (Stratagene). Recombinants were screened by alkaline lysis minipreparations, restriction enzyme digestion, Southern blotting and
hybridization to a randomly primed $^{32}$P-labelled cDNA probe (produced from TAV RNA) as described by Maniatis et al. (1982).

**Blotting and nucleic acid hybridization.** DNA was electrophoresed through 1% agarose gels and capillary blotted onto Pall-Biodyne or GeneScreen Plus nylon membranes according to the manufacturer's instructions. TAV RNA was electrophoresed through denaturing formaldehyde gels as described by Miller (1987) and transferred onto Hybond-N membranes (Amer sham) according to the manufacturer's instructions. Nucleic acid probes were made either by reverse transcription of TAV RNA with avian myeloblastosis virus (AMV) reverse transcriptase (Amer sham) or by oligolabelling cDNA with the Klenow fragment of DNA polymerase I (Pharmacia) in the presence of [a-$^{32}$P]dATP. Oligonucleotide probes were labelled with [γ-$^{32}$P]ATP by T4 polynucleotide kinase (BRL), as described by Maniatis et al. (1982). Membranes were hybridized with cDNA or oligolabelled probes at 42 °C in 4 × SSC, 0.1% SDS, 0.05% Marvel (Cadborough), 50% formamide and then washed in 2 × SSC, 1% SDS at 70 °C. Membranes hybridized with oligonucleotide probes were incubated at 25 °C in 4 × SSC, 0.1% SDS, 0.05% Marvel and washed at 25 °C in 6 × SSC, 1% SDS.

**Nucleotide sequence determination.** Templates for sequencing of cloned cDNA were prepared either by single-stranded plasmid rescue using the f1 superinfecting phage K048 procedure (Stratagene) or by using Sephacryl 400 (Pharmacia) according to the manufacturer's instructions to purify double-stranded plasmid DNA. Dideoxynucleotide chain termination reactions were performed with T7 DNA polymerase (United States Biochemical) and [α-$^{32}$P]dATP as described by the enzyme manufacturer. TAV RNA templates for direct sequencing were purified as described above and sequenced using AMV reverse transcriptase according to the manufacturer's instructions (Promega). Sequencing products were electrophoresed on 6% acrylamide, 7 M-urea gels and read with an SAC Grafbar sonic digitizer.

**Analysis of sequencing data.** All sequence data was compiled and manipulated using the STADEN software package (Amer sham) and PROTYLZE (S. & E. Software).

**Oligonucleotides.** Nine oligonucleotides were used in this study; all were synthesized on a Biotech DNA synthesizer and read: NTAV1 (5'TTGGACATCCAGTGGGA3'), TAVCDNA1 (5'TGGGACATCCAGTGGGA3'), TAVPCR1 (5'TGGGACATCCAGTGGGA3'), TAVCDNA2 (5'ACGTCGATCAGTGGGA3'), TAVPCR2 (5'ACGTCGATCAGTGGGA3'), TAVCDNA3 (5'ACGTCGATCAGTGGGA3'), TAVPCR3 (5'ACGTCGATCAGTGGGA3'), TAVCDNA4 (5'ACGTCGATCAGTGGGA3'), TAVPCR4 (5'ACGTCGATCAGTGGGA3'). NTAV3' differed by two nucleotides from its targeted region in C-TAV RNA. Screening of the randomly primed library identified one clone, pTAV413, which, when used to probe a Northern blot of TAV RNA, allowed amplification of the rarer, longer molecules. This was a modification of a strategy described by Frohman et al. (1988) and requires sequence information at only one end of the template. First-strand cDNA was adenylated by terminal deoxynucleotidyl transferase (BRL) as described by Frohman et al. (1988). PCR was performed on the adenylated cDNA using oligonucleotide primers, TAVPCR1, -2 or -3, together with the oligo(dT) primer, LdT(15), using Taq polymerase (Perkin Elmer) and a Hybaid Thermal Reactor. Reaction conditions were essentially as described by Saiki et al. (1988). PCR reaction products were size-fractionated on 1% agarose gels using NA45 DEAE paper (Schleicher & Schuell) as described above for DNA, ligated into Smal-digested pHueter M13 + KS (Stratagene) and transformed into competent E. coli JM103 cells. Recombinants were screened by colony PCR (Gussow & Clackson, 1989) and by alkaline lysis (Maniatis et al., 1982).

**Results**

**Determination of the 3' sequence of TAV RNA 3.** Confirmation that C-TAV was a genuine TAV isolate was achieved by Western blotting of the virus using antisera raised against the type isolate of TAV (strong reaction) and antisera raised against CMV (no reaction; results not shown). Gel electrophoresis of TAV in the presence of SDS facilitated the separation of the genomic RNAs of interest (Fig. 1a). As no definitive information was available on the 3'-terminal sequence of RNA 3 of this virus isolate, all the RNA species were polyadenylated in vitro. cDNA libraries were obtained from both oligo(dT) and randomly primed templates. An oligonucleotide, NTAV3', was synthesized to be complementary to the 3'-terminus of RNA 3 of N-TAV (Nicotiana glauca, South Australia) (Wilson & Symons, 1981) and used to screen, by probing the recombinant plasmids, for 3'-specific clones. An investigation of the cDNA library by dideoxynucleotide sequencing revealed that these clones averaged 200 to 300 bp in length and possessed poly(A) tails of 15 to 20 nucleotides. This analysis also revealed that the oligonucleotide probe NTAV3' differed by two nucleotides from its targeted region in C-TAV RNA. Screening of the randomly primed library identified one clone, pTAV413, which, when used to probe a Northern blot of TAV RNA, hybridized with RNAs 3 and 4 only (Fig. 1c) and was used later (see below) to assist cloning of the entire RNA 3 molecule. 3'-specific clones hybridized to all RNAs because of their known 3' sequence homology (results not shown). Sequencing of pTAV413 and selected 3'-specific clones generated a consensus of 650 nucleotides for the 3' region of RNAs 3 and 4 (Fig. 1d).

**PCR-directed cloning of RNA 3.** Owing to the difficulties experienced in obtaining long cDNAs from TAV RNAs by conventional cloning procedures, a PCR-based method was developed that allowed amplification of the rarer, longer molecules. This was a modification of a strategy described by Frohman et al. (1988) and requires sequence information at only one end of the template. First-strand cDNA
synthesis was performed on fractionated RNA 3 initially using the 3'-specific oligonucleotide primer TAVCDNA1. Adenylation of the 3' ends of these products, followed by size fractionation, allowed amplification of the longer cDNAs using the oligonucleotides LdT(15) and TAVPCR1. Since TAVPCR1 overlaps TAVCDNA1 by five nucleotides at its 3' end, this procedure prevents the efficient amplification of any contaminating adenylated TAVCDNA1. Electrophoresis of amplified cDNAs, followed by blotting and hybridization to α-32P-labelled pTAV413 probe, demonstrated that RNA 3- and 4-derived cDNAs had been amplified to generate molecules of up to 1 kb in length (Fig. 1b). One clone obtained by this method, pTAV347, and a series of smaller clones generated a consensus sequence of 1 kb in length of the 3' region of C-TAV RNAs 3 and 4. By using the overlapping primers TAVCDNA2 and TAVPCR2, which were made complementary to the 5' terminus of pTAV347, in similar PCR reactions as above, it was possible to generate clones that extended into the 5' region of TAV RNA 3. Products from this step generated the clone pTAV391 and increased the consensus sequence to 1700 bp from the 3' terminus. The origin of this clone was confirmed by its ability to hybridize specifically only to RNA 3 on a Northern blot of total genomic TAV RNA (data not shown). A final PCR reaction using the primers TAVCDNA3 and TAVPCR3, which were made complementary to the 5' terminus of pTAV391, amplified a major product of 500 bp in length. Cloning of this fragment produced pTAV302 and several other cDNAs possessing an identical 5' sequence, which was thought to represent the 5' terminus of RNA 3. The 5' sequence of RNA 3 was confirmed by direct dideoxynucleotide sequencing of the RNA with the oligonucleotide primer TAV3SEQ1 (data not shown). Fig. 1(d) shows the relationship of these clones with C-TAV RNA 3.

**Organization of the RNA 3 sequence of C-TAV**

A consensus sequence of 2214 nucleotides for RNA 3 of C-TAV (Fig. 2) was determined by dideoxynucleotide chain termination sequencing of the three sets of cDNAs derived by PCR and of those derived by direct cDNA cloning. We are confident that the sequence presented is that of C-TAV RNA 3 and there are no unresolved discrepancies in the consensus. Two open reading frames (ORFs) were predicted from this sequence. The first ORF is preceded by a 5' non-coding region of 89 nucleotides and encodes a protein of 277 amino acids. This ORF is thought to encode the 3a protein (Schwinghammer & Symons, 1977). An intercistronic region of 306 nucleotides separates the 3a ORF from the putative coat protein ORF (229 amino acids). The size predicted for this protein is similar to that of the capsid protein determined by Western blotting (data not shown). This ORF is followed by a 3' non-coding 306 nucleotide region. The transcription start site for the generation of the mRNA of the coat protein (RNA 4) has not been determined.

**Sequence comparisons with other tripartite viruses**

The RNA and predicted protein sequences of TAV were compared with their equivalent regions from two other tripartite viruses, Q-CMV (Gould & Symons, 1982 and corrected, Davies & Symons, 1988) and BMV (Ahlquist et al., 1981). Initial analysis was performed using a dot matrix comparison program (DIAGON; Staden, 1982) and applying an 'identities algorithm' score of 6/11. From this analysis, TAV showed no sequence identity with BMV (results not shown) but had some limited identity with Q-CMV at its 5' and 3' termini within the 3a ORF and the intercistronic regions (Fig. 3a). Using similar parameters as above, for the predicted amino acid sequences of the 3a and coat proteins of C-TAV and Q-CMV only the 3a proteins showed any direct identity.
Fig. 2. The entire nucleotide sequence and predicted ORFs of C-TAV RNA 3. Protein 3a and the coat protein are initiated at nucleotide positions 90 and 1219 from the 5' end. Termination codons are represented by an asterisk.
Fig. 3. Dot matrix comparisons for (a) Q-CMV v. C-TAV RNA 3; (b, d) Q-CMV v. C-TAV 3a protein; (c, e) Q-CMV v. C-TAV coat protein; (f) BMV v. C-TAV 3a protein; (g) BMV v. C-TAV coat protein. Plots (a) to (c) are scored by identity (DIAGON) and (d) to (g) by double matching probability (McLachlan, 1977).

(Fig. 3b). No identity was detectable between the coat proteins of C-TAV and Q-CMV (Fig. 3e) or the 3a and coat proteins of C-TAV and BMV (not shown). Similarly, Q-CMV and BMV did not show any direct sequence identity in these protein sequences. However, when the 'double matching probability' algorithm of McLachlan (1971) was used to analyse the data, different results were obtained. This procedure uses the odds score matrix MDM78 (Dayhoff, 1969) which was derived by looking at accepted point mutations in families of closely related proteins and allows a method for identifying distant relationships between amino acid sequences. Using a percentage score of 132 the dot matrix plots indicated near relatedness of both the C-TAV and Q-CMV proteins (Fig. 3d, e) and a more distant relationship between C-TAV and BMV proteins (Fig. 3f, g). For the same predicted proteins, Q-CMV, when compared with BMV, offered a plot similar to that found when C-TAV was compared with BMV (results not shown; see Davies & Symons, 1988). This suggests that C-TAV and Q-CMV are equally related to BMV. These findings also demonstrate that although C-TAV, Q-CMV and BMV show limited, or no direct, sequence identity in either their RNA 3s or encoded proteins when examined by dot matrix analysis, they are related to each other by a series of point mutations.

To investigate further the nature of this relatedness, the predicted amino acid sequences for the 3a and coat proteins were subjected to the Chou–Fasman (Chou & Fasman, 1978a, b) method for the prediction of secondary structure (PROTYLZE; S. & E. Software). Fig. 4 shows the results of this analysis, in graphic form,

Fig. 4. Chou–Fasman graphic plots comparing the predicted α-helix and β-sheet formation for the 3a and coat proteins of C-TAV (——), Q-CMV (-----) and BMV (——). Values on the vertical axes are average propensities for structure formation.
of α-helix and β-sheet formation for the 3a and coat proteins. A striking similarity between the structures of the corresponding proteins of each virus is clearly evident. As before, C-TAV appears to be more closely related to Q-CMV than to BMV, although less obviously so than from the dot matrix comparisons. Thus, despite the fact that the proteins of these viruses do not share the same primary structures, the mutations that have given rise to this feature have not significantly altered their secondary structures and possibly their tertiary conformations.

**Discussion**

Characterization of the 3'-terminal sequence of RNA 3 of C-TAV and the application of a PCR technique (Frohman et al., 1988) has facilitated the determination of the entire RNA 3 nucleotide sequence and that of its encoded proteins. This has revealed that C-TAV RNA 3 possesses a genetic organization similar to that of other tripartite RNA viruses. Identity at the RNA level, as determined by dot matrix analysis, was seen only between C-TAV and Q-CMV. The fact that this identity is restricted to the non-coding regions and the 3a ORFs might suggest that they are under mutational constraint. However, BMV shows no such identity to either C-TAV or Q-CMV in these areas and yet all three RNAs demonstrate similar properties, such as the ability to form tRNA-like structures at their 3' termini (Wilson & Symons, 1981; Gould & Symons, 1982; Ahlquist et al., 1981) which may be due to differing mutation rates between these viruses. It is not unusual for non-structural proteins to be tolerant to a high degree of amino acid substitution. Miller et al. (1979) found that of approximately 1500 single amino acid substitutions at 142 positions in a lac repressor protein, about 50% were phenotypically silent. The discrepancy between the conservation of primary and secondary structures of these proteins can be explained by the observation that mutation frequencies in viral RNA genomes are high compared to DNA genomes (Holland et al., 1982) and that functional constraints on these proteins have only allowed mutations that substitute chemically similar amino acids (McLachlan, 1971). It is still not clear, however, whether the structural similarities of these proteins reflect divergent or convergent evolution. The genomic organization of CMV, BMV and AMV (Haseloff et al., 1984), and now TAV, implies a common ancestry (Goldbach, 1986). Indeed, even RNA viruses with such dissimilar genomic organization as AMV, tobacco mosaic virus and the animal alphavirus, Sindbis virus (Haseloff et al., 1984), have been shown to possess similar amino acid sequences in proteins believed to perform common functions. Further investigations into the roles of these viral proteins and RNA sequences using recombinant DNA techniques will not only facilitate an understanding of the molecular mechanisms involved in viral infection but may also give a more accurate indication of which features should be considered when attempting to establish their evolutionary relationships.

David O'Reilly is supported by a Science and Engineering Research Council CASE award with Advanced Technologies (Cambridge) Ltd.

**References**


(Received 6 June 1990; Accepted 12 October 1990)