Nucleotide sequence and structural analysis of two satellite RNAs associated with chicory yellow mottle virus

L. Rubino,¹* M. E. Tousignant,² G. Steger³ and J. M. Kaper²

¹Dipartimento di Patologia Vegetale, Università degli Studi and Centro di Studio del CNR sui Virus e le Virosi delle Colture Mediterranee, Bari, Italy, ²Microbiology and Plant Pathology Laboratory, Plant Sciences Institute, ARS, U.S. Department of Agriculture, Beltsville, Maryland 20705, U.S.A. and ³Institut für Physikalische Biologie, Universität Düsseldorf, F.R.G.

The two satellite RNAs associated with CYMV infections were sequenced. The larger (sCYMV-L1) has only linear molecules 1145 nucleotides long, a poly(A) tail, a long open reading frame (ORF) coding for a protein of Mr 39636 resembling in composition those of other large nepovirus satellite RNAs, a 5' leader sequence of 16 nucleotides and a 3' non-coding region of 40 nucleotides. In vitro translation of sCYMV-L1 yielded a protein product with a size that corresponded to that predicted from the sequence. The smaller satellite (sCYMV-S1) is 457 nucleotides long, has no ORF of significant length and no in vitro messenger activity. Both linear and circular forms of this satellite RNA were detected in infected tissues. Comparison of the sCYMV-S1 primary structure with the sequences of other small nepoviral satellites reveals large regions of homology. Analysis of the secondary structures derived from the sequences of the plus and minus strands suggests possible consensus sequences for their self-cleavage.

Introduction

Three isolates of chicory yellow mottle nepovirus (CYMV) were obtained from chicory in different years from the same area (Vovlas et al., 1971) and named CYMV-RS, CYMV-C and CYMV-T. The symptomatology in chicory was the same each year and the isolates were not serologically distinguishable, but differed only in their nucleic acid composition. CYMV-RS particles contain two genomic RNAs, of approximately Mr 2·3 x 10⁶ and 2·0 x 10⁶. In addition to the genomic RNAs, CYMV-C particles also contain a third RNA with an estimated Mr of 0·5 x 10⁶ (sCYMV-L1), and those of CYMV-T contain a fourth RNA with an estimated Mr of 0·17 x 10⁶ (sCYMV-S1) (Piazzolla et al., 1986). The two low Mr species share no substantial sequence homology with each other or with the genomic RNAs and represent two different satellite RNAs supported by the same virus (Piazzolla & Rubino, 1984; Piazzolla et al., 1989). sCYMV-L1 has no significant influence on symptom expression (Piazzolla et al., 1986), but sCYMV-S1 is able to attenuate symptoms of CYMV infection in Nicotiana tabacum Xanthi nc and in N. glutinosa. sCYMV-L1 is polyadenylated with a free hydroxyl group at the 3' end and a 5' terminus which is blocked. The satellite is present in virions and infected tissue only as a monomeric linear molecule. sCYMV-S1 has a cyclic 2'-3' phosphate at the 3' end and a free hydroxyl group at the 5' end. It occurs in a linear form in virus particles and a circular form exists in infected tissue (Piazzolla et al., 1989).

The molecular characterization of the two satellites supported by CYMV is reported in this paper.

Methods

Virus purification and RNA extraction. Virus purification, RNA extraction and electroelution of satellite RNAs were performed as previously described (Piazzolla et al., 1989). Only purified satellite RNAs were used for cloning and sequencing.

Cloning and sequencing of sCYMV-L1. First and second strand cDNA synthesis was performed according to Gubler & Hoffman (1983) using the cDNA Cloning System Plus (Amersham); first strand synthesis was primed by an oligo(dT)₁₂₋₁₈. Double-stranded cDNA tailed with oligo(dC) and terminal transferase (IBI) according to the supplier was annealed with oligo(dG)-tailed pUC9 and cloned in Escherichia coli JM103 (Vieira & Messing, 1982). Plasmid DNA was extracted using the boiling method (Holmes & Quigley, 1981) and digested with PstI; the recombinants were identified by Southern blot analysis (Maniatis et al., 1982), using cDNA prepared with random hexanucleotide primers as a probe to gel-purified sCYMV-L1. Several overlapping deletion subclones were prepared from the clones containing the largest inserts by restriction enzyme digestion. Subcloned plasmid DNA was extracted and denatured according to Hattori & Sakaki (1980) and sequenced using the dideoxynucleotide chain termination method (Sanger et al., 1977) and modified T7 DNA polymerase (Sequenase, United States Biochemicals).
Cloning and sequencing of sCYMV-S1.

Fig. 2. Comparison of conserved nucleotide sequences in large nepoviral satellites. Underlined nucleotides in the consensus are present in all sequences.

Fig. 3. Northern blot of total RNA preparations from infected tissues after denaturing electrophoresis and probing with a nick-translated clone of sCYMV-S1. Lane V contains RNA extracted from virions: only monomeric (L) and dimeric (D) linear forms are present. Other lanes contain different RNA preparations from infected tissues and show the circular (C) forms.

A synthetic 17-nucleotide oligomer (A) complementary to nucleotides (nt) 230 to 246 was used for primer extension to the 5' end of gel-purified sCYMV-L1. Primer and template were heated as above. This part of the sequence was confirmed by the same labelled oligonucleotide A.

Cloning and sequencing of sCYMV-S1. Gel-purified sCYMV-S1 was labelled at the 3' end with [5'-32P]pCp and T4 RNA ligase (BRL) (Piazolla et al., 1989) and at the 5' end with [7-32P]ATP and polynucleotide kinase (Haseloff & Symons, 1981), as a probe and identified by colony hybridization (Maniatis et al., 1982) using oligonucleotide A, which was 5' end-labelled with [7-32P]ATP. To determine the sequence proximal to the primer, sequencing was done with primer B 5' end-labelled with [7-32P]ATP.

The same oligonucleotide was used to prime first strand cDNA synthesis of sCYMV-S1. Second strand synthesis, tailing and cloning of the double-stranded cDNA was as described for sCYMV-L1. Recombinants were identified after Southern blotting and hybridization experiments (Maniatis et al., 1982). The DNA was extracted and sequenced as described above. A partial clone of about 300 bp was obtained and used as a probe in Northern blot hybridization experiments (Maniatis et al., 1982).

Sequence analysis. Sequences were compiled and analysed with the DNA Strider program (Marck, 1988) run on a Macintosh Plus computer. Nucleotide sequences were aligned according to Collins & Coulson (1987) and the most stable secondary structures were calculated as previously described (Steger et al., 1984).
Results

Cloning and sequence analysis of sCYMV-L1

About 400 transformants which contained sequences of sCYMV-L1 were selected. Eighteen clones with the largest inserts were further analyzed by restriction mapping and were found to be 3'-coterminal. However, sequence analysis showed that most did not contain the poly(A) tract. Four clones, containing inserts ranging between 500 and 1000 bp, were selected for sequencing on both strands after preparing deletion subclones of suitable size (approximately 200 bp).

Cloning of the 5' end region was completed using the oligonucleotide 5' TGATGGGAGTGGTGACA 3' complementary to nt 230 to 246 as a primer. About 40 clones were tested and several were sequenced on both strands. Two, pCY5-16 and pCY5-18, contained the 5' end of the RNA as confirmed by the sequence obtained by direct RNA sequencing using the same primer.

The sequence of sCYMV-L1 is 1145 nt long, not including the poly(A) tail (Fig. 1). The Mr determined from the sequence is approximately 0.37 x 10^6, which is lower than that previously estimated (Piazzolla et al., 1986).

Sequence analysis of the extreme 5' region of sCYMV-L1 suggests that it is probable that a genome-linked protein (VPg) is bound to the first nucleotide. Previous work has shown that the 5' terminus is not accessible to labeling (Piazzolla et al., 1989). Furthermore, the first nucleotide is U, which with many plant viruses and picornaviruses has been shown to be linked to a VPg. Finally, the 16 nt leader sequence contains the UGAAAAAU stretch, which has been proposed to be a consensus sequence at 5' termini linked to VPg in nepovirus satellite RNAs (Fuchs et al., 1989). The 3' end of sCYMV-L1 is characterized by a poly(A) tail about 15 nt long (Piazzolla et al., 1989). As in the satellite RNAs of grapevine fanleaf (sGFLV; Fuchs et al., 1989) and tomato black ring (sTBRV isolates E and C; Hemmer et al., 1987) nepoviruses and other plant viruses, there is no evidence for either a specific polyadenylation consensus sequence, or the eukaryotic messenger RNA polyadenylation signal (Fitzgerald & Shenker, 1981).

---

Fig. 4. Alignment of nucleotide sequences of sCYMV-S1, sTobRV and sArMV in plus strand orientation. Boxed regions encompass the sequences that can form the ‘hammerhead’ structure (---) and the minus strand self-cleavage consensus sequence (:: :) respectively.

Positions agreeing with the consensus are given in upper case. A letter in the consensus sequence is given if at least two of the satellite sequences contain the same nucleotide at this position.

---

Corrections of the nucleotide sequences of sCYMV-L1, sTobRV and sArMV in plus strand orientation. Boxed regions encompass the sequences that can form the ‘hammerhead’ structure (---) and the minus strand self-cleavage consensus sequence (:: :) respectively.

Positions agreeing with the consensus are given in upper case. A letter in the consensus sequence is given if at least two of the satellite sequences contain the same nucleotide at this position.

---
Following the non-coding leader sequence of 16 nt, the sCYMV-L1 has a large open reading frame (ORF), with an AUG at positions 17 to 19. The sequence context of this AUG as a start codon conforms in part to the consensus proposed by Kozak (1984) and Lütke et al. (1987), as it has a G in position +4 and a C in position −2; the next available AUG (nt 221) would give a very much smaller product and is therefore unlikely to be used for initiation. The termination codon is at positions 1103 to 1105, giving a putative protein of 363 amino acids (M, 39636, 40K). The sCYMV-L1 nucleotide sequence is terminated by a non-coding region of 40 nt and the poly(A) tail.

Analysis of protein coding region and in vitro translation product

Analysis of the in vitro 35S-labelled translation products of sCYMV-L1 showed one major polypeptide of estimated M, 41000 (not shown). This value is close to the size expected from sequence analysis. The sCYMV-L1 40K protein resembles those encoded by sTBRV (Hemmer et al., 1987) and sGFLV (Fuchs et al., 1989) in two ways. It contains more basic residues (17%) than acidic residues (7.4%); the first 100 residues are particularly basic (24 basic and four acidic). Also the hydrophilicity profile (Hopp & Woods, 1981) of the sCYMV-L1 protein (data not shown) shows a hydrophobic domain at the N terminus, followed by a highly hydrophilic stretch of amino acids.

Comparison of the nucleotide sequence of sCYMV-L1 with the sequences of other large nepoviral satellites

We have constructed dot plots with different stringencies comparing the sequences of sCYMV-L1 and sGFLV (Fuchs et al., 1989), sTBRV (Hemmer et al., 1987) or the large satellite RNA of arabis mosaic virus (J. I. Cooper, quoted in Fuchs, 1989) and have found a region of homology between nt 94 to 133 of sCYMV-L1 and regions of other nepoviral satellites as indicated in Fig. 2. In addition, as previously mentioned, sCYMV-L1 shares with other nepovirus satellites the 5' sequence UGAAAATAU.

Linear and circular forms of sCYMV-S1

Our previous work showed that a free hydroxyl group is present at the 5' end of the sCYMV-S1 molecule, but that 3' end labelling could only be accomplished after hydrolysis with maleic acid at pH 2 and dephosphorylation with calf intestinal phosphatase, suggesting the presence of a cyclic 2'-3' phosphate at the 3' end (Piazzolla et al., 1989). This indicated that sCYMV-S1 could also be present in a circular form. Northern blot analysis of total nucleic acid from virions after electrophoresis on denaturing gels showed that no circular molecules are encapsidated (Piazzolla et al., 1989). However, denaturing gel electrophoresis and Northern blot hybridization of RNA extracted from infected tissues gives clear evidence that there are two forms of sCYMV-S1, which differ in electrophoretic mobility, present in infected leaves (Fig. 3). To confirm that the slower migrating band C represented a circular form of sCYMV-S1, the two bands, C and L, were eluted separately from denaturing polyacrylamide gels and submitted to the procedure for 5' end labelling. Only the material in the L band was readily labelled, whereas that in the C band, representing the putative circular form, was not (results not shown).

In some experiments, nucleic acid preparations were also analysed by bidimensional gel electrophoresis (Schumacher et al., 1983), blotted and probed with
cloned probes to either sCYMV-S1 or L1. They confirmed the presence of circular forms for sCYMV-S1 only (results not shown).

**Nucleotide sequence of sCYMV-S1**

The complete nucleotide sequence of the sCYMV-S1 was determined directly on the RNA with approximately two-thirds of the sequence corroborated by DNA sequencing of the available clones. The sequence is 457 nt long; the longest ORF, initiated by AUG in position 224, would give a translation product of only 51 amino acids, which was not detected in *in vitro* translation experiments (data not shown). Dot plot analysis of the sequence of sCYMV-S1 compared to the sequence of sCYMV-L1 shows no significant homology (data not shown).
Structural features of sCYMV-S1 compared with other small nepoviral satellites

Fig. 4 aligns sCYMV-S1 with the published sequences of the satellite of tobacco ringspot virus (sTobRV; Buzayan et al., 1986) and the small satellite of arabis mosaic virus (sArMV; Kaper et al., 1988). A high degree of homology (about 75%) between the three satellite RNAs was found in the first 60 nt of the 5' end. This sequence and the terminal nucleotides (ACTGTC) at the 3' end can be folded into a 'hammerhead' structure (Fig. 5), which has been suggested to determine the self-cleavage site for a number of viral satellites (such as sTobRV RNA; Fig. 5) and avocado sunblotch viroid (Foster & Symons, 1987).

A third region of homology is located in the middle of the three sequences: alignments between nt 255 to 305 of sCYMV-S1 and nt 106 to 156 of sArMV show 80% homology; the same percentage of homology is also found between nt 255 to 305 of sCYMV-S1 and nt 175 to 224 of sTobRV. This conserved sequence gives rise to a stable 'hairpin' below the central stem in the secondary structure of sCYMV-S1 (Fig. 6); this is also present in structures proposed for sArMV and sTobRV (Kaper et al., 1988). It has been demonstrated that the minus strand self-cleavage site of sTobRV is located in a 'catalytic domain' determined by this conserved region which is different from the 'hammerhead' consensus sequence of the (+) strand (Hampel & Tritz, 1989); a similar structure can be obtained by folding the (−)(sCYMV-S1 and (−)sArMV sequences (Fig. 7).

Discussion

CYMV supports the replication of two satellite RNAs (L1 and S1) that have no structural features in common, i.e. size, nucleotide sequence and primary structure of the termini or protein coding ability.

sCYMV-L1 resembles the satellites reported for GFLV, TBRV and ArMV (large) in size, messenger activity and the size of in vitro protein products. In addition, the 5' terminus of sCYMV-L1 is likely to have a VPg, as there is a short nucleotide stretch at the 5' terminus that is identical to the consensus sequence linked to a VPg in other nepoviral satellites. The blocked 5' end and the presence of a free hydroxyl group at the 3' end excludes the possibility that sCYMV-L1 is replicated by a rolling circle mechanism and makes greater the likelihood of its replication by a linear mechanism similar to that operating for the helper viral RNA.

sCYMV-S1 resembles the small nepoviral satellites sArMV and sTobRV in having no messenger activity and in the nature of the RNA termini which are compatible with a circular configuration (Konarska et al., 1982). Such circular forms have been reported for sArMV (Kaper et al., 1988) and sTobRV (Linthorst & Kaper, 1984) and in this report are shown to occur in CYMV-infected tissues (Fig. 3). The three small nepoviral satellite RNAs are also characterized by similar secondary structures and by their consensus sequence domain for a 'hammerhead' structure involved in self-cleavage of sTobRV. The presence of a 'hammerhead' structure and the detection in vivo of a circular form strongly supports the notion that sCYMV-S1, like sTobRV, is replicated by a rolling circle mechanism. This is further reinforced by our finding of dimeric forms of sCYMV-S1 in virions (Fig. 3).

Satellites or satellite-like RNAs have been grouped according to (i) helper virus species or group, (ii) whether or not the satellite encodes its own coat protein, (iii) molecular size and (iv) ability to modulate plant disease expression (for reviews see Murant & Mayo, 1982; Francki, 1985; Kaper & Collmer, 1988). Satellites can also be divided into two groups according to whether or
not they possess at their termini structural features similar to those at the termini of their helper virus RNAs. Such a grouping might imply that the different types of satellite depend on their helper viruses for different functions.

The two satellite RNAs supported by CYMV belong to the two types and therefore the CYMV system could be a good model for the understanding of mechanisms and consensus sequences involved in viral parasitism.

L. R. wishes to thank Dr Marcello Russo for helpful discussions and advice given during the course of this work.

**References**


(Received 8 February 1990; Accepted 24 April 1990)