Journal of General Virology (1990), **71**, 1259–1263. *Printed in Great Britain*

The nucleotide sequence of a satellite RNA associated with arabis mosaic nepovirus

Yuan Yi Liu,[1]* C. U. T. Hellen,[1]† J. I. Cooper,[2] D. J. Bertioli,[1] D. Coates[3] and G. Bauer[4]

[1] *Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB,* [2] *NERC Institute of Virology, Mansfield Road, Oxford OX1 3SR,* [3] *Genetics Laboratory, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU and* [4] *Max Planck Institut für Biophysikalische Chemie, Nikolaufberg, 3400 Göttingen, F.R.G.*

The nucleotide sequence of a satellite RNA (satRNA) associated with a lilac isolate of arabis mosaic virus (ArMV) was determined from cDNA copies. The sequence was 1104 nucleotides in length excluding the poly(A) tail, contained a long open reading frame which encodes a polypeptide of 360 amino acids, with an M_r of 39K. Nucleotide sequence comparisons revealed that the ArMV-associated satRNA shared 83% nucleotide identity with a satRNA from grapevine fanleaf nepovirus, but no extensive sequence homology was observed with other nepoviral satRNAs.

Introduction

Arabis mosaic virus (ArMV) is a nepovirus with many antigenic properties in common with grapevine fanleaf virus (GFLV; Hewitt *et al.*, 1971); however it differs in the breadth of its natural host range and in being carried usually by a distinct species of longidorid nematode. In common with other nepoviruses, ArMV has a bipartite RNA genome each molecule of which is 3′-polyadenylated and terminated at its 5′ end by a genome-linked protein, VPg (Mayo *et al.*, 1982; Hellen, 1987). In some instances, natural virus isolates contain a third encapsidated RNA molecule, a satellite RNA (satRNA) *sensu* Murant & Mayo (1982). Davies & Clark (1983) described an isolate of ArMV (from hops growing in Kent, U.K.) which contained a satRNA that closely resembles the satRNA of tobacco ringspot nepovirus (TobRV; Schneider, 1977; Buzayan *et al.*, 1986) in that it occurs as linear or circular molecules and has the ability to undergo self cleavage. The hop ArMV-associated satRNA has 50% sequence homology with the satRNA of TobRV (Kaper *et al.*, 1988).

When virion RNA from particles of a lilac (*Syringa vulgaris*) isolate of ArMV was analysed, a third RNA (RNA-3) was detected. This, like the satRNAs of tomato black ring virus (TBRV; Murant *et al.*, 1973 Mayo *et al.*, 1979; Koenig & Fritsch, 1982; Meyer *et al.*, 1984) and GFLV (Pinck *et al.*, 1987; Fuchs *et al.*, 1989), resembled the viral genomic RNAs structurally in being 3′-polyadenylated, and in having mRNA activity and a 5′-linked VPg.

Biological properties of the RNA-3 in lilac ArMV and evidence for its satellite nature will be published elsewhere but here we report its nucleotide sequence and make comparisons with sequences of other satRNAs.

Methods

Virus propagation and nucleic acid preparation. ArMV was propagated in *Chenopodium quinoa* and purified from systemically infected leaves as described by Harrison & Nixon (1960) except that particles were coacervated with 10% (w/v) polyethylene glycol in the presence of 1% (w/v) sodium chloride. RNA was extracted from virions as described by Massalski & Cooper (1986). Glyoxal-treated satRNA was separated from genomic species by electrophoresis in agarose gels (McMaster & Carmichael, 1977) and was recovered by electroelution using an ISCO model 1750 concentrator as described by Zassenhaus *et al.* (1982).

Synthesis and cloning of DNA. The RNA-3 separated from the genomic RNAs was used as a template for cDNA synthesis using oligo(dT)$_{12\text{-}18}$ primers and avian myeloblastosis virus (AMV) reverse transcriptase. The reaction conditions for synthesis of cDNA were based on those described by Cann *et al.* (1983) and Davies *et al.* (1978). In 50 µl of reaction mixture, 5 µl of satRNA was mixed with 5 µl of reverse transcriptase buffer (× 10) (containing 0·5 M-Tris–HCl pH 8·3, 1·0 M-NaCl, 100 mM-$MgCl_2$, 60 mM-DTT, 5 µl each of the four dNTPs (5 mM), 5 µCi[α-^{32}P]dATP (10 µCi/µl, 6000 Ci/mmol), 2 µl oligo(dT)$_{12\text{-}18}$ (1 mg/ml) and 40 units of AMV reverse transcriptase), and incubated at 37 °C for 30 min. After extraction with phenol–chloroform, cDNA–RNA hybrids were separated from unincorporated nucleotides

† Present address: Department of Microbiology, School of Medicine, Health Sciences Center, State University of New York at Stony Brook, New York 117794-8621, U.S.A.

by chromatography through a mini column of Sephadex G-50 (fine) and then precipitated by adding ethanol to 75%.

The cDNA–RNA hybrids were tailed with dC residues using terminal deoxynucleotidyl transferase and inserted into the *Pst*I site of dG-tailed pBR322 (Bethesda Research Laboratories) by the method described by Cann *et al.* (1983). The recombinant plasmid was used to transform *Escherichia coli* HB101 and transformants were selected by virtue of their resistance to tetracycline and their sensitivity to ampicillin. Colonies having plasmids containing RNA-3-specific inserts were selected by colony hybridization (Grunstein & Hogness, 1975) using cDNA prepared from the RNA-3 as the probe.

DNA sequencing. Clones (A3/3 which contains the longest satellite sequence and another of half this size, A3/2, representing the 3′ end) were both sequenced entirely by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). The inserts were cut out from the plasmids with the restriction enzyme *Pst*I, then digested with *Hin*dIII. The fragments were subcloned into the *Hin*dIII/*Pst*I site of M13 mp10 (Messing, 1983) and sequenced using the universal sequencing primer. Two primers at nucleotide positions 942 to 959 (5′ AACCGGCTATTATGGGAA 3′) or complementary to nucleotides 443 to 461 (3′ GAACGAAAGGTTAGGAAGA 5′) were used to facilitate extension of the sequence towards the 3′ and 5′ ends. These primers were synthesized by the phosphotriester method with an Applied Biosystems 380 B DNA Synthesizer. With M13 reverse sequencing primer (Pharmacia), double-stranded replicative forms of M13 mp10 containing the inserts were sequenced substantially as described by Chen & Seeburg (1985) from 5′ and 3′ ends. Finally, the cDNA inserts were cut with *Hpa*II and subcloned into *Pst*I/*Acc*I or *Acc*I sites of M13 mp10. All fragments were sequenced in both orientations.

RNA sequencing. A synthetic oligodeoxyribonucleotide complementary to the residues 130 to 149 (3′ AACATAACAGCACATTCGTG 5′) was used to determine the satRNA sequence at its 5′ end by primer extension with reverse transcriptase. Approximately 1 μg of RNA-3 and 2 ng of the primer in 10 μl reverse transcriptase buffer were incubated at 90 °C for 3 min. After gradual cooling to room temperature (22 °C), dideoxynucleotide chain termination sequencing was done using a Bio-Rad M13 sequencing RTase Kit and protocols supplied with it.

The terminal sequence was also determined using fluorescent dideoxynucleotide terminators and automated fluorescence detection with a DuPont Genesis 2000 DNA sequencer (Bauer, 1990). In this work a synthetic oligomer complementary to residues 25 to 41 (3′ TATAATTGCGCTTATGT 5′) was used with a preparation of virion RNA containing RNA-1, RNA-2 and RNA-3.

Sequencing analysis. This was done using a variety of programs (Staden 1982, 1984*a*, *b*, 1986; Devereux *et al.*, 1984), available on the Oxford University Computing Service VAX/VMS.

Results and Discussion

The complete nucleotide sequence of RNA-3 from ArMV (lilac) comprises 1104 nucleotides and is presented in Fig. 1. The clone A3/3 almost fully represented the satRNA and contained a run of 30 A residues (not shown in Fig. 1) indicating that the RNA sequence at its 3′ end had been copied. The sequence at the 5′ end, obtained when RNA-3 from virions was used as a template with the reverse transcription kit, terminated with two indecipherable nucleotides. These were identified as U and A using fluorescent dideoxynucleotide terminators; neither was present in clone A3/3.

The positive-sense RNA-3 sequence contains one long open reading frame whereas the negative-sense strand contains only short ones; the longest corresponds to 88 amino acids, a protein of approximately 9K. The first AUG, at positions 2 to 4, corresponds to an initiation codon for a putative peptide containing eight amino acids. In this, as in many other respects, the ArMV satRNA sequence closely resembles that described by Fuchs *et al.* (1989); in the GFLV RNA-3, the corresponding peptide has 22 amino acids. The longest ORF of ArMV RNA-3 has three in-frame methionine codons at positions 15 to 17, 126 to 128 and 216 to 218 in the 5′ end region. Proteins initiated at each of these methionines would have M_r values of 38849, 34673 and 31310 respectively. When ArMV RNA-3 was translated in rabbit reticulocyte lysates (Hellen, 1987), three products were obtained with M_r values of 41K, 37K and 36K. The 41K product was the most abundant, the 37K protein was present in lesser amounts and the 36K product was barely detectable. By contrast, Fuchs *et al.* (1989), using wheatgerm extracts and GFLV RNA-3, obtained only one translation product although the first three methionines fall in identical positions within the longest ORF of this sequence. We have not yet excluded the possibility that the three translation products obtained from ArMV RNA-3 result from premature termination after initiation at the first AUG or the possibility that another initiation codon is occasionally selected. Furthermore, although it is conceivable that the satRNA of ArMV is a mixture of molecules differing in the positions of their stop codons, we did not find sequence heterogeneity between the clones A3/3 and A3/2.

The 3′ end of the RNA-3 from ArMV has a non-coding region of seven nucleotides and is thereby shorter than the corresponding regions in the satRNAs of GFLV (74; Fuchs *et al.*, 1989) or TBRV (74; Meyer *et al.*, 1986). The region does not contain the polyadenylation signal AAUAAA implicated by Benoist *et al.* (1980) for animal mRNAs and viral RNAs or that proposed by Messing *et al.* (1983) for plant mRNAs i.e. G/AAUAA1-3. The 5′ end of the ArMV RNA-3 sequence has a non-coding leader of 14 nucleotides preceding the first AUG of the longest ORF. The 5′-terminal 11 nucleotides (UAUGAAAAAUU) fits the consensus sequence at the 5′ end of nepovirus RNAs noted by Fuchs *et al.* (1989); UG/UGAAAAAU/AU/AU/A).

The satRNA of lilac ArMV has a base composition of 22·7% A, 25·7% C, 28·0% G and 23·8% U. The distribution of bases in the third codon position of the longest ORF is 22·8% A, 22·5% C, 25·8% G and 28·5% U. These values indicate a lesser preference for U than

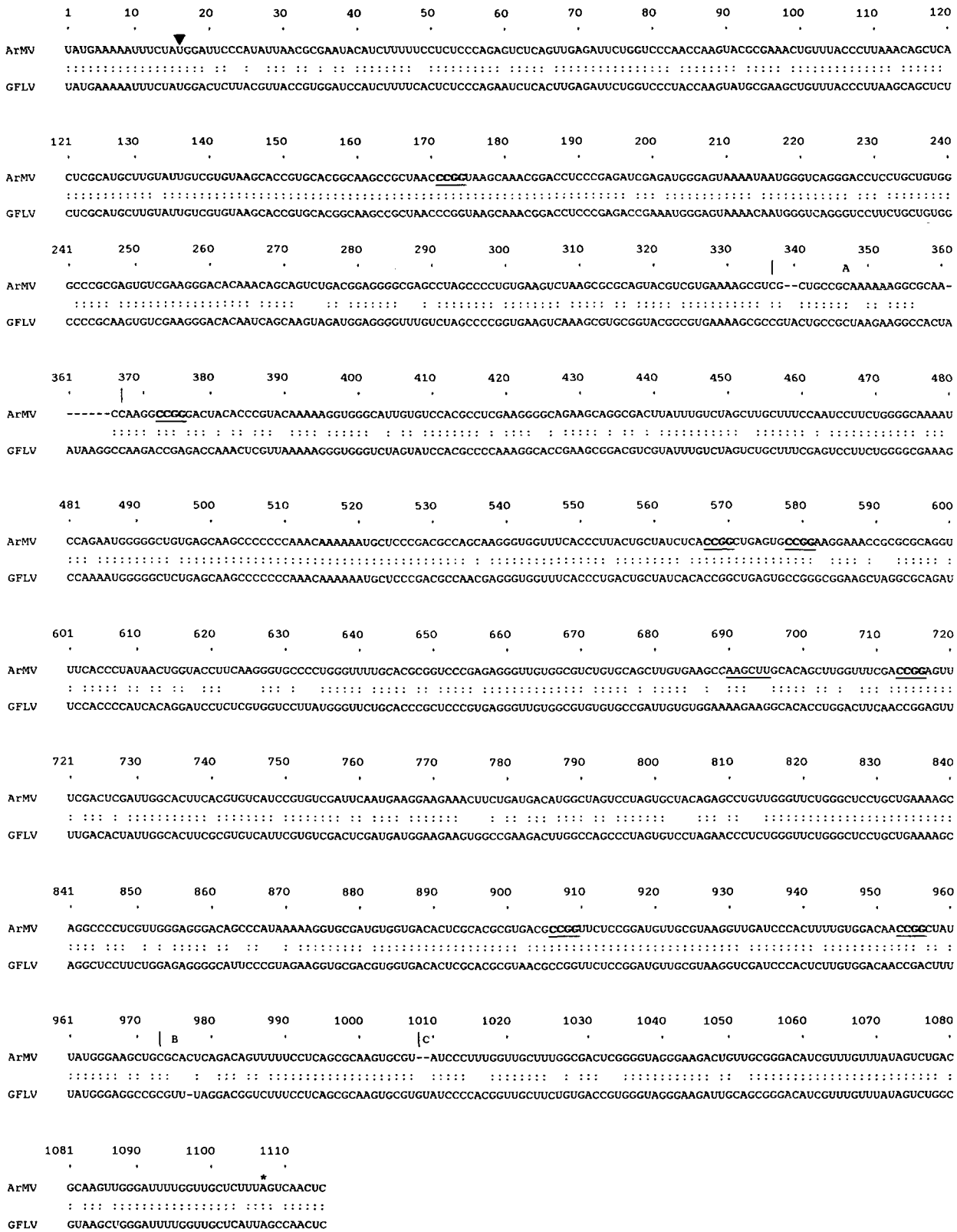

Fig. 1. Alignments of the nucleotide sequences of ArMV (lilac) RNA-3 (upper) and GFLV RNA-3 (lower). Bold and underlined nucleotides are sites recognized by *Hpa*II; the single *Hin*dIII cleavage site is underlined. Spaces introduced to optimize the alignments are marked with hyphens (-); ▼ and * indicate the start and stop codons, respectively, of the longest ORF of ArMV RNA-3.

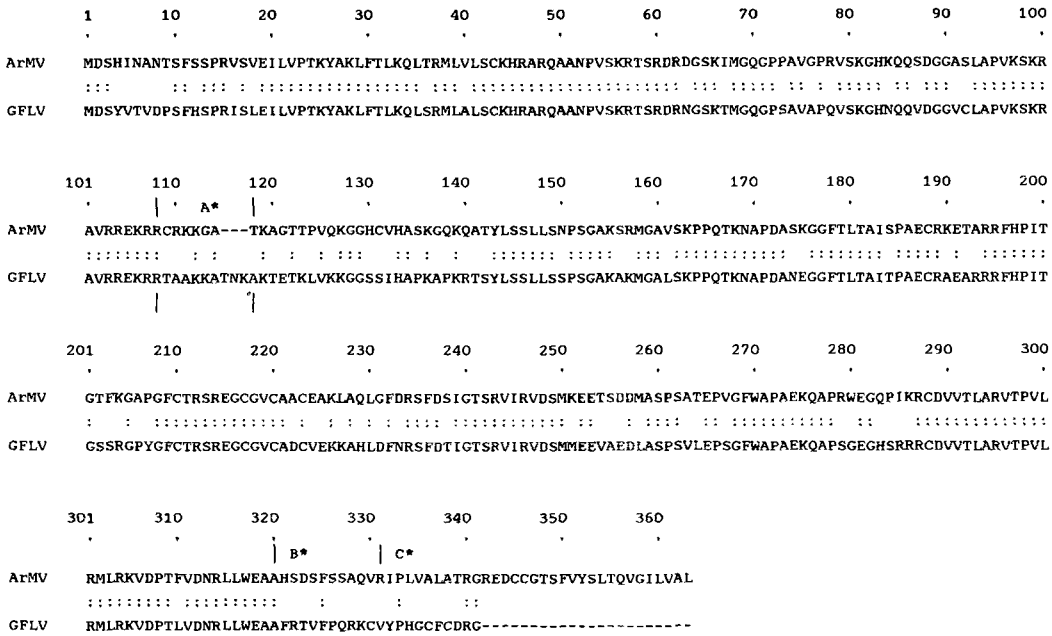

Fig. 2. Alignment between the deduced amino acid sequences from the ArMV (lilac) and GFLV RNA-3 sequences. Spaces introduced to optimize the alignments are marked with hyphens (-). A*, B* and C* correspond to regions A, B and C in Fig. 1.

many plant viral RNAs (Dasgupta & Kaesberg, 1982; van Wezenbeek *et al.*, 1983; Meyer *et al.*, 1986).

Nucleotide sequence comparisons showed no significant homology between the ArMV satRNA and those of TBRV (Meyer *et al.*, 1984), TobRV (Buzayan *et al.*, 1986), ArMV hop isolate (Kaper *et al.*, 1988), or cucumber mosaic virus (CARNA5; Collmer *et al.*, 1983). It was surprising to discover that the ArMV satRNA shares 83% of its nucleotide sequence with that of the satRNA of GFLV (Fuchs *et al.*, 1989); to make the best possible fit in the alignments between the nucleotide sequences of these two satRNAs (Fig. 1) only four breaks are required. With regard to the fact that these two satRNA sequences are so similar, two of these breaks (in region A, at positions 338 and 339 and positions 360 to 366) could be treated as self-correcting 'frameshift' mutations resulting in a loss of three amino acids to the coding potential of ArMV satRNA. The possible 'frameshift' area A* (at positions 108 to 118) in the alignments between the deduced amino acid sequences of these two satRNAs (Fig. 2) shows this effect, with only two amino acid identities maintained, and a three amino acid break. Following this 'frameshift' area, there is a region of low amino acid sequence homology corresponding to amino acid positions 120 to 143. Although this region can be considered in-frame (nucleotide sequence homology of 73%), the amino acid sequence differs greatly (with 43% homology) and most of the differences are at the first and second base positions. The other two breaks (Fig. 1, in B and C) may be compound 'frameshift' alterations which completely change the amino acid sequence at the C terminus of the predicted proteins (Fig. 2, B* and C*).

This work was done while D. J. Bertioli was in receipt of a NERC Studentship and in part under a contract (PECD/7/8/106) with the Department of the Environment which has authorized publication. We are also grateful to Susan Jackson for text capture.

References

BAUER, G. J. (1990). RNA sequencing using fluorescent-labelled dideoxynucleotides and automated fluorescence detection. *Nucleic Acids Research* **18**, 879–884.

BENOIST, D., O'KARE, K., BREATHNACH, R. & CHAMBON, P. (1980). The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Research* **8**, 127–142.

BUZAYAN, J. M., GERLACH, W. L., BRUENING, G., KEESE, P. & GOULD, A. R. (1986). Nucleotide sequence of satellite RNA and its relationship to multimeric forms. *Virology* **151**, 186–199.

CANN, A. J., STANWAY, G., HAUPTMANN, R., MINOR, P. D., SCHILD, G. C., CLARKE, L. D., MOUNTFORD, R. C. & ALMOND, J. W. (1983). Poliovirus type 3: molecular cloning of the genome and nucleotide sequence of the region encoding the protease and polymerase proteins. *Nucleic Acids Research* **11**, 1267–1281.

CHEN, F. Y. & SEEBURG, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4**, 165–170.

COLLMER, C. W., TOUSIGNANT, M. E. & KAPER, J. M. (1983). Cucumber mosaic virus-associated RNA 5: X. The complete nucleotide sequence of a CARNA 5 incapable of inducing tomato necrosis. *Virology* **127**, 230–234.

DASGUPTA, R. & KAESBERG, P. (1982). Complete nucleotide sequences of the coat protein messenger RNAs of brome mosaic virus and cowpea chlorotic mottle virus. *Nucleic Acids Research* **10**, 703–713.

DAVIES, D. L. & CLARK, M. F. (1983). A satellite-like nucleic acid of arabis mosaic virus associated with hop nettlehead disease. *Annals of Applied Biology* **103**, 439–448.

DAVIES, J. W., VERVER, J. W. G., GOLDBACH, R. W. & VAN KAMMEN, A. (1978). Efficient reverse transcription of cowpea mosaic virus RNAs. *Nucleic Acids Research* **5**, 4643–4660.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

FUCHS, M., PINCK, M., SERGHINI, M. A., RAVELONANDRO, M., WALTER, B. & PINCK, L. (1989). The nucleotide sequence of satellite RNA in grapevine fanleaf virus, strain F13. *Journal of General Virology* **70**, 955–962.

GRUNSTEIN, M. & HOGNESS, D. S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences, U.S.A.* **72**, 3961–3965.

HARRISON, B. D. & NIXON, H. L. (1960). Purification and electron microscopy of three soil-borne plant viruses. *Virology* **12**, 104–117.

HELLEN, C. U. T. (1987). *Studies on the structure, replication and expression of the cherry leaf roll genome.* D.Phil. thesis, University of Oxford.

HEWITT, W. B., MARTELLI, G., DIAS, H. F. & TAYLOR, R. H. (1971). Grapevine fan leaf virus. *CMI/AAB Descriptions of Plant Viruses*, no. 28.

KAPER, J. M., TOUSIGNANT, M. E. & STEGER, G. (1988). The nucleotide sequence predicts circularity and self-cleavage of 300-ribonucleotide satellite of arabis mosaic virus. *Biochemical and Biophysical Research Communications* **154**, 318–325.

KOENIG, I. & FRITSCH, C. (1982). A protein linked at the 5′ end of satellite and genomic tomato black ring virus RNAs: study of *in vitro* translation after protease treatment. *Journal of General Virology* **60**, 343–353.

MCMASTER, G. K. & CARMICHAEL, G. G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels using glyoxal and acridine orange. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 4835–4838.

MASSALSKI, P. R. & COOPER, J. I. (1986). Comparison of the genome RNA sequence homologies between isolates of cherry leaf roll virus by complementary DNA hybridization analysis. *Journal of General Virology* **67**, 1169–1172.

MAYO, M. A., BARKER, H. & HARRISON, B. D. (1979). Polyadenylate in the RNA of five nepoviruses. *Journal of General Virology* **43**, 603–610.

MAYO, M. A., BARKER, H. & HARRISON, B. D. (1982). Specificity and properties of the genome-linked proteins of nepoviruses. *Journal of General Virology* **59**, 149–162.

MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101**, 20–77.

MESSING, J., GERAGHTY, D., HEIDECKER, G., HU, N.-T., KRIDL, J. & RUBENSTEIN, J. (1983). Plant gene structure. In *Genetic Engineering of Plants: An Agricultural Perspective*, pp. 211–227. Edited by T. Kosuge, C. P. Meredith & A. Hollaender. New York & London: Plenum Press.

MEYER, M., HEMMER, O. & FRITSCH, C. (1984). Complete nucleotide sequence of a satellite RNA of tomato black ring virus. *Journal of General Virology* **65**, 1575–1583.

MEYER, M., HEMMER, O., MAYO, M. A. & FRITSCH, C. (1986). The nucleotide sequence of tomato black ring virus RNA-2. *Journal of General Virology* **67**, 1257–1271.

MURANT, A. F., MAYO, M. A., HARRISON, B. D. & GOOLD, R. A. (1973). Evidence for two functional RNA species and a 'satellite' RNA in tomato black ring virus. *Journal of General Virology* **19**, 275–278.

MURANT, A. F. & MAYO, M. A. (1982). Satellites of plant viruses. *Annual Review of Phytopathology* **20**, 49–70.

PINCK, L., FUCHS, M., PINCK, M., RAVELONANDRO, M. & WALTER, B. (1987). A satellite RNA in grapevine fanleaf virus strain F13. *Journal of General Virology* **69**, 233–239.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SCHNEIDER, I. R. (1977). Defective plant viruses. In *Beltsville Symposia in Agricultural Research. I. Virology in Agriculture*, pp. 201–219. Edited by J. A. Romberger, J. D. Anderson & R. L. Powell. Montclair, New Jersey: Allanheld, Osmun & Co.

STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731–4751.

STADEN, R. (1984*a*). Computer methods to locate signals in nucleic acid sequences. ANALYSEQ: promoters, ribosome binding sites, intron/exon junctions. *Nucleic Acids Research* **12**, 505–519.

STADEN, R. (1984*b*). Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Research* **12**, 521–538.

STADEN, R. (1986). The current status and portability of our sequence handling software. *Nucleic Acids Research* **14**, 217–231.

VAN WEZENBEEK, P., VERVER, J., HARMSEN, J., VOS, P. & VAN KAMMEN, A. (1983). Primary structure and gene organisation of the middle component RNA of cowpea mosaic virus. *EMBO Journal* **2**, 941–946.

ZASSENHAUS, P. H., BUTOW, R. A. & HANNON, Y. P. (1982). Rapid electroelution of nucleic acids from agarose and acrylamide gels. *Analytical Biochemistry* **125**, 125–130.

(*Received 14 November 1989; Accepted 6 February 1990*)