Volume 10 Number 12 1982 Nucleic Acids Research

Comparative sequence and structure of viroid-like RNAs of two plant viruses

James Haseloff and Robert H.Symons

Department of Biochemistry, University of Adelaide, Adelaide, South Australia 5001

Received 18 May 1982; Accepted 1 June 1982

ABSTRACT

A newly discovered group of spherical plant viruses contains a bipartite genome consisting of a single-strand linear RNA molecule (RNA 1, M_r 1.5 x 10^6), and a single-strand, covalently closed circular viroid-like RNA molecule (RNA 2, $M_r \sim 125,000$). The nucleotide sequences of the RNA 2 of two of these, velvet tobacco mottle virus and solanum nodiflorum mottle virus, have been determined. RNA 2 of solanum nodiflorum mottle virus consists of 377 residues whereas that of velvet tobacco mottle virus consists of two approximately equimolar species, one of 366 residues and the other, with a single nucleotide deletion, of 365 residues. There is 92 - 95% sequence homology between the RNA 2 species of the two viruses. The predicted secondary structures possess extensive intramolecular base pairing to give rod-like structures similar to those of viroids. The structural similarities between the RNAs 2 of velvet tobacco mottle virus and solanum nodiflorum mottle virus and viroids may reflect functional similarities.

INTRODUCTION

The discovery of a new, unique group of plant viruses in Australasia was reported recently (1,2,3). The 30 nm diam. polyhedral virions contain two major single-strand RNA species; RNA 1 is a linear molecule of about 4,500 residues (M_r 1.5 x 10^6) whereas RNA 2 is a circular covalently closed molecule of 300 - 400 residues with a high degree of internal base pairing and physical properties similar to viroids, another group of plant pathogens.

So far, there are four members of this new group of plant viruses; velvet tobacco mottle virus (VTMoV), solanum nodiflorum mottle virus (SNMV), lucerne transient streak virus, and subterranean clover mottle virus (1,2,3). The best characterized are VTMoV and SNMV which possess a bipartite genome since both RNA 1 and the viroid-like RNA 2 are necessary for infection. The genetic functions provided by the two viral RNAs have not been determined except for that coding for the coat protein (4). VTMoV and SNMV are serologically related while hybridization analysis with complementary DNA gave estimates of sequence homology for the viral RNAs 1 of between 20% and 50%

0305-1048/82/1012-3681$2.00/0

depending on the stringency of the assay conditions (2). On the other hand, hybridization analysis indicated that the complete sequence of VMToV RNA 2 (M_r 1.2 x 10^5) is contained in SNMV RNA 2 (M_r 1.3 x 10^5) (2). Despite the close sequence similarities between the RNAs 2 of VTMoV and SNMV, neither RNA will support the replication of the heterologous RNA 1 (4) which indicates a highly specific relationship between the RNA 1 and RNA2 of each virus.

Although viroids are single-strand circular RNA molecules of similar size (250 - 400 residues) to these RNA 2 molecules and are internally base-paired to form rod-shaped molecules with helical segments separated by short single-strand regions, they are not encapsidated and replicate autonomously (5,6). In order to investigate the intriguing relationships between the RNAs of VTMoV and SNMV, we have sequenced the RNA 2 species of each virus and have compared their structures with those of viroids.

MATERIALS AND METHODS

Viruses and RNA

VTMoV and SNMV were kindly provided by Drs. R.I.B.Francki, J.W.Randles and A.R.Gould. Viruses were purified from infected *Nicotiana clevelandii* and viral RNAs isolated and purified essentially as described (1).

RNA Sequencing Procedures

(a) Partial enzymic digestion. Specific linear RNA fragments were obtained from circular RNA 2 molecules by partial RNase digestion under non-denaturing conditions as detailed previously (7) except that 150 units/ml of RNase T_1 and 0.25 units/ml of RNase U_2 were required for VMToV RNA 2 and 300 units/ml of RNase T_1 and 0.25 units/ml of RNase U_2 for SNMV RNA 2. The resultant RNA fragments were 5'-^{32}P-labelled *in vitro*, fractionated by polyacrylamide gel electrophoresis and sequenced by the partial enzymic digestion technique as in Haseloff and Symons (7).

(b) Dideoxynucleotide chain termination. RNA fragments produced by partial RNase T_1 digestion were also sequenced using the dideoxynucleotide chain termination technique (8). Specific purified 5'-^{32}P-fragments were dephosphorylated with calf intestinal phosphatase (9) and polyadenylated with *E.coli* poly(A) polymerase (10). Sequencing reactions were carried out using $d(T_8C)$ as the specific primer (11).

Synthesis and Cloning of Double-strand cDNA

SNMV RNA 2 (4 µg) in 10 µl distilled water was heated at 100°C for 30 min in a sealed capillary to generate randomly cleaved full-length linear molecules. Terminal 2'(3')-phosphates were removed with calf intestinal

phosphatase and the RNA molecules polyadenylated with poly(A) polymerase (9, 10). First strand ^{32}P-cDNA was synthesized using reverse transcriptase and oligo-$(dT)_{10}$ as primer (12). The reaction mixture was heated at 100°C for 1 min, snap cooled on ice and the RNA removed by digestion with 500 µg/ml RNase A for 20 min at 56°C. After phenol extraction, ether washing and ethanol precipitation, the ^{32}P-cDNA in 10 µl water was heated at 100°C for 1 min, snap cooled on ice and made to 20 µl containing 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 3 mM dithiothreitol, 1 mM of each of the four dNTPs, and 1 unit of Klenow fragment of *E.coli* DNA polymerase I. Incubation for synthesis of the second DNA strand using self-priming was at 37°C for 4.0 hr.

The double-strand cDNA was digested with *Sau*3A I restriction enzyme and the digest fractionated by electrophoresis in a 6% polyacrylamide gel in 90 mM Tris-borate, pH 8.3, 1 mM EDTA (13). A DNA fragment corresponding to residues 131 to 216 of SNMV RNA 2 was excised and eluted (14) and ligated into the *Bam*H I site of the replicative form of phage M13mp7 using phage T4 DNA ligase (15,16). Recombinant phage were screened by sequencing (17) using a specific M13 primer ($GTA_4CGACG_2C_2AGT$). Recombinant M13 replicative form was isolated (18), digested with *Sau*3A I and the cloned insert purified on a 6% polyacrylamide gel (13,14) and used as a primer for the sequencing of RNA 2 of VMToV and of SNMV by the dideoxynucleotide chain termination technique (19).

RESULTS

Approaches Used for Sequencing RNA 2 of VTMoV and SNMV

Since RNA 2 of VTMoV and SNMV exist as covalently closed single-strand circular RNAs (1,2), the approach used successfully for the sequencing of chrysanthemum stunt viroid (CSV) (7), avocado sunblotch viroid (ASBV) (19), and citrus exocortis viroid (CEV) (20) was also used here. Briefly the purified circular RNAs were partially digested with RNases T_1, U_2 and A and the exposed 5'-ends of the fragments labelled with γ-^{32}P-ATP and T4 polynucleotide kinase. The 5'-labelled fragments were fractionated by size on 80 cm long denaturing polyacrylamide slab gels, 20 - 30 labelled bands from each gel were eluted and sequenced using the partial enzymic sequencing procedure described in Haseloff and Symons (7). In addition, fragments produced by partial T_1 RNase digestion were also sequenced from the 3'-end by the dideoxynucleotide chain termination procedure (11) after removal of the 3'-terminal phospahte and polyadenylation. Sequencing of fragments in both the 5'- and 3'-directions allowed confirmation of sequences and the resolution of occasional band compressions (7,11) which were seen on sequencing in one direc-

tion but not in the other. The complete sequences of the two RNAs were eventually obtained from the sequences of numerous overlapping RNA fragments.

Primary Structures of RNA 2 of VTMoV and SNMV

The complete base sequences of these two RNAs are given in Fig. 1. Although the RNAs are covalently closed circular molecules, the sequences are presented in linear form for convenience and for ease of comparison. SNMV RNA 2 consists of 377 residues while VTMoV RNA 2 consists of two approximately equimolar species, one of 366 residues which, like SNMV RNA 2, has a U at residue 108, and another species of 365 residues where this residue is deleted.

This sequence heterogeneity within RNA 2 of VTMoV was determined by sequence analysis of individual purified fragments which differed in size by one residue and which were derived from either the 366 or 365 residue species. Further confirmation of this sequence heterogeneity and an estimate of the relative proportions of the two species were obtained using a cloned DNA fragment (derived from residues 131 to 216 of SNMV RNA 2) as a primer on the mixture of the intact VTMoV RNA 2 species. The 365 residue species is arbitrarily presented in Fig. 1 and numbering of the sequence will refer to this species.

The extensive sequence homology between these two RNAs, originally reported on the basis of hybridization analysis with cDNA (2), is confirmed by the sequence data. However, each RNA 2 species contains unique sequences; 95% of VTMoV RNA 2 is homologous with SNMV RNA 2 and 92% of SNMV RNA 2 is homologous with VTMoV RNA 2. The sequence differences are unevenly scattered throughout the two RNAs with a cluster of base differences around residues 339 - 359 of SNMV RNA 2 and residues 333 – 347 of VTMoV RNA 2 while there is almost complete sequence homology between residues 360 - 146 of SNMV RNA 2 and residues 348 - 145 of VTMoV RNA 2.

Secondary Structures of RNA 2 of VTMoV and SNMV

Secondary structure models for the two RNAs were constructed as described by Tinoco *et al.*(21) and are given in Fig. 2. Both RNAs form extensively base-paired rod-like structures which are similar to those described for viroids (7,19,20,22,23). The structures are consistent with the known sites of high sensitivity to ribonuclease under the conditions of high salt concentration used to generate specific RNA fragments for sequencing from the circular RNAs. Thus, the terminal single-strand hairpin loops and the central single-strand regions of both RNAs (residues 70 - 100 and 285 - 305) were especially susceptible.

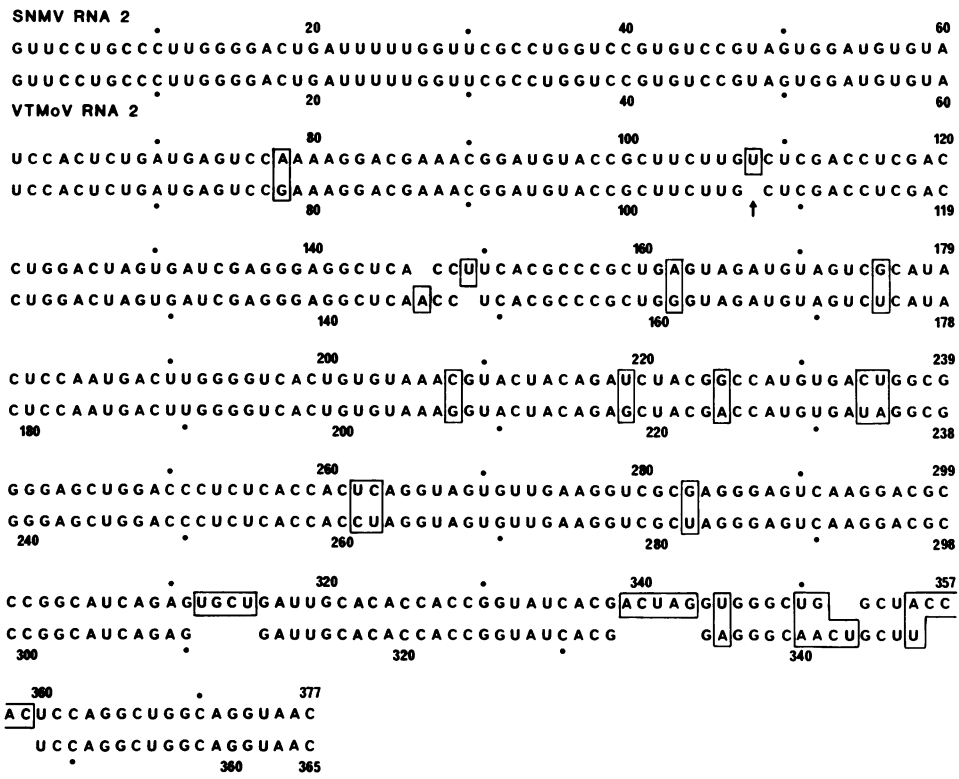

Figure 1. The primary sequences of SNMV RNA 2 and the 365 residue form of VTMoV RNA 2 are shown in linear form and aligned for maximum sequence homology. The 366 residue form of VTMoV RNA 2 has an extra UMP residue at position 108 (arrow). The sequence differences between VTMoV RNA 2 and SNMV RNA 2 are boxed. Residue 1 in each case corresponds to the left-hand end of the secondary structure model of Figure 2.

The properties of the proposed structures are summarized in Table 1 and are compared to those of the published structures of four viroids. The VTMoV and SNMV RNA 2 molecules possess proportions of G:C base pairing which are similar to those of the viroids potato spindle tuber viroid (PSTV), CSV and CEV, but higher than that of the smaller ASBV.

The thermodynamic stabilities of the proposed models were calculated using values kindly provided by Dr. D.Riesner (Steger, Gross, Randles, Sänger and Riesner, in preparation). The values of -455 KJ/mol for SNMV RNA 2 and of -345 and -350 KJ/mol for the two forms of VTMoV RNA 2 (Table 1) are consistent with their thermal denaturation properties; thus VTMoV RNA 2 gave a T_m of 57°C in 0.15 M NaCl, 0.015 M sodium citrate, pH 7, while SNMV RNA 2 gave a higher T_m of 64°C under the same conditions (2,24). The predicted

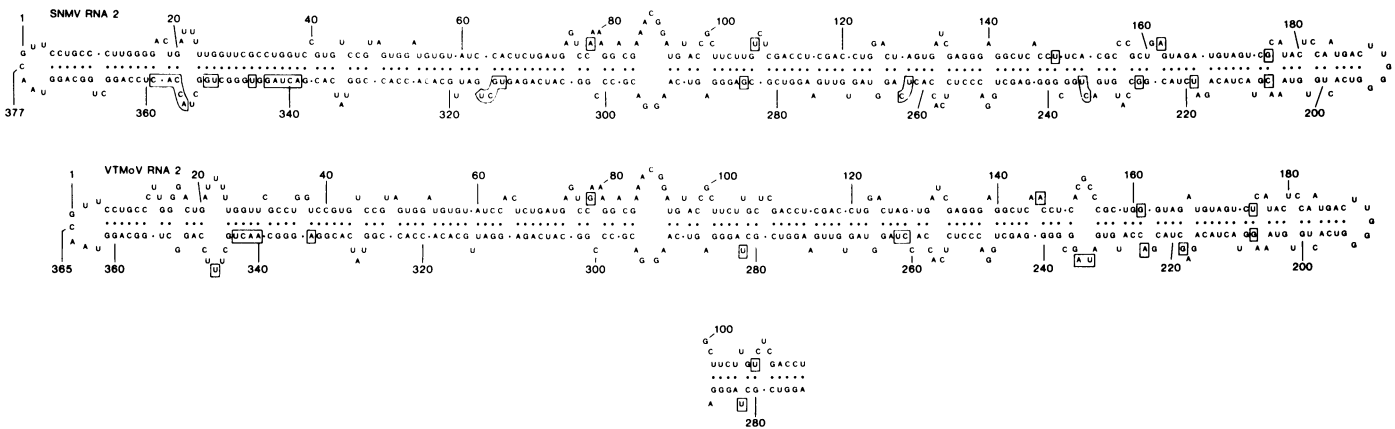

Figure 2. Predicted secondary structures of SNMV RNA 2 and the 365 residue form of VTMoV RNA 2 plus the segment of the 366 residue form of VTMoV RNA 2 containing the extra UMP residue (residue 108). The sequence differences between the RNAs are boxed.

stabilities of the RNA 2 of VTMoV and of SNMV are lower than those of the similarly sized viroids (PSTV, CSV and CEV) but higher than that of the smaller ASBV (Table 1).

Possible Polypeptide Translation Products from RNA 2 Species and their Complements

Since the RNA 2 species of VTMoV and SNMV are required with the homologous RNA 1 for viral infection (4), the RNA 2 molecules must either code for some protein product(s) and/or contain structural information essential for viral replication. Evidence suggests that, in almost all instances, a 5'-proximal AUG codon functions as the initiation signal for eukaryotic mRNA translation (25,26) and eukaryotic ribosomes do not interact with circular RNAs (27,28). Therefore, translation of the RNA 2 species would require the existence of specific linear RNA forms. Assuming that this condition is met, extensive homology allows SNMV RNA 2 and the 366 residue form of VTMoV RNA 2, and their complementary sequences, to code for several similar polypeptide products (Fig. 3). However, only one small polypeptide product is shared between SNMV RNA 2 and the 365 residue form of VTMoV RNA 2 and their complements. All possible translation products are less than 100 amino acids in length and therefore the genes coding for the viral coat proteins (approximately 300 amino acids; refs. 1,29) must reside in the RNA 1 species.

Table 1. Properties of proposed secondary structures for RNA 2 of VTMoV and SNMV compared with those of several viroids

RNA	No.of residues	No. of base pairs			G:C base pairs as % of total	Residues base paired %	ΔG* (KJ/mol at 25°C in 1M NaCl)	Reference
		A:U	G:C	G:U				
VTMoV RNA 2	365	38	72	13	59	67	-350	This work
VTMoV RNA 2	366	38	71	14	58	67	-345	This work
SNMV RNA 2	377	41	76	20	55	73	-455	This work
ASBV	247	43	28	12	34	67	-280	(19)
PSTV	359	37	73	16	58	70	-610	(23)
CSV	356	44	64	16	52	70	-540	(7)
CEV	371	34	72	18	58	67	-590	(20)

*Parameters for calculation provided by Dr. D. Riesner (Steger, Gross, Randles, Sänger and Riesner, personal communication).

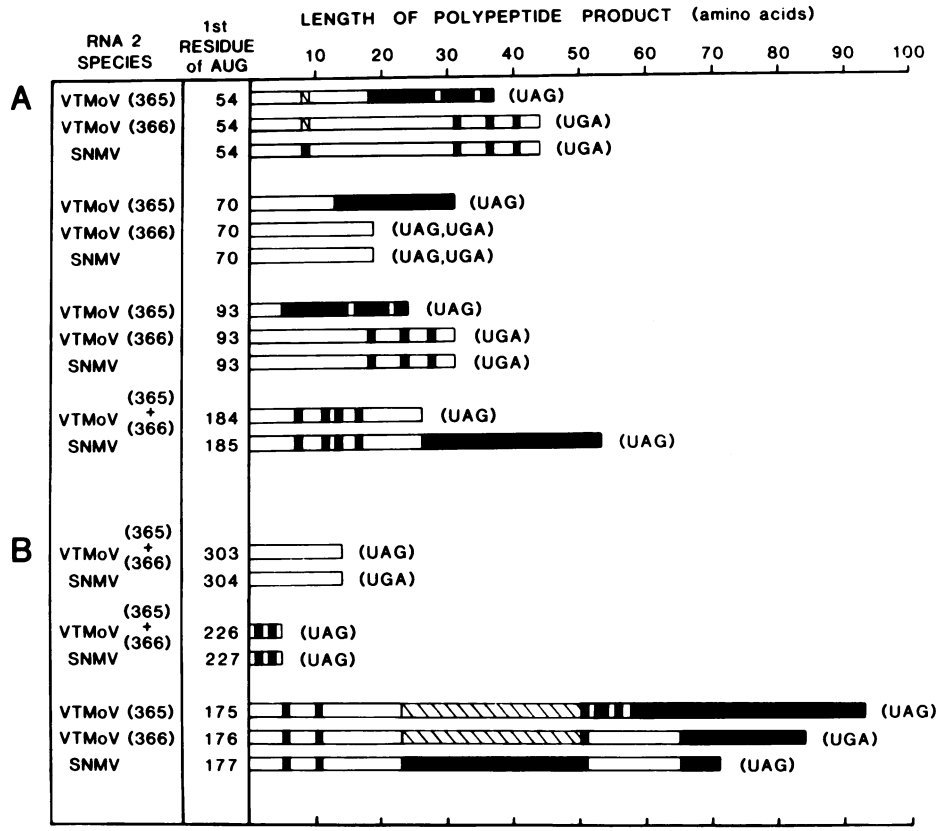

Figure 3. Possible polypeptide products of SNMV RNA 2 and the 365 and 366 residue forms of VTMoV RNA 2 (A) and their putative complementary RNAs (B) are shown in schematic form. Each possible translation product is given with the residue number of the first base of the initiation codon plus the termination codon(s) in parentheses. For the complementary sequences, the same residue numbers are retained and therefore run in the 3' to 5' direction. The clear areas represent regions of amino acid sequence homology and the black areas regions of non-homology in the RNAs shown. The cross-hatched areas are regions of sequence homology between different products of the two VTMoV RNAs which correspond to regions of non-homology in SNMV RNA 2. Where an identical product is obtained from the two forms of VTMoV RNA 2, only one is presented.

DISCUSSION

In overall structure, VTMoV and SNMV RNA 2 resemble viroids in being small single-strand covalently closed circular RNA molecules with extensive base-paired regions interspersed with single-strand regions. However, in contrast with viroids which replicate autonomously and are not encapsidated (5,6) the RNA 2 species are essential components of a bipartite genome and

are encapsidated (4). While viroids do not appear to code for functional protein products (7,30,31,32), there is no information available for the RNA 2 species.

The lack of conservation of possible translation products between SNMV RNA 2 and the 365 and 366 residue forms of VTMoV RNA 2, despite greater than 90% sequence homology, suggests either that the 365 residue form of VTMoV RNA 2 may be non-functional or that RNA 2 coded translation products may have no function in viral replication. Although the involvement of RNA 2 coded translation products in viral replication cannot be excluded, it seems likely that the unique viroid-like structures of the RNA 2 molecules encode some function besides that of a template, and that this function is required for the replication of both RNA 1 and RNA 2 species.

Since neither VTMoV RNA 2 nor SNMV RNA 2 supports the replication of the heterologous RNA 1 species (4), the biological specificities of the RNA 2 species must be determined by differences in primary and/or secondary structures. The only extensive region of sequence differences between VTMoV and SNMV RNA 2 lies around VTMoV RNA 2 residues 333 - 347 and SNMV RNA 2 residues 339 - 359. Hence, this region may be involved in determining the specificity of the relationship between the RNA 1 and RNA 2 species, although the involvement of other structural differences cannot be excluded.

An unexpected complication during the sequence determination of the RNA 2 molecules was the sequence heterogeneity in VTMoV RNA 2 which consisted of two species existing in approximately equimolar amounts and differing by the presence or absence of a U residue at position 108 (Fig. 2). The two RNAs may have arisen either by a single mutation in a parent molecule followed by independent replication of the resultant two RNA species, or a mixture may have been produced during each cycle of replication by transcriptional and/or processing events. In each case, it is possible that one of the two RNA species may be non-funcitonal. Sequence heterogeneity within RNA populations has been reported for RNA phage Qβ (33), vesicular stomatitis virus (34), satellite tobacco necrosis virus (35) and citrus exocortis viroid (36).

In the secondary structures of the viroids PSTV (23), CSV (7,36), CEV (20,36) and coconut cadang cadang viroid (CCCV) (Haseloff, Mohamed and Symons, unpublished), there is a central region of the rod-like structures which is highly conserved in both sequence and structure. ASBV does not share this common structure (19) except for the residues GAAACC (ASBV residues 45 - 50; ref. 19) which, as in the other viroids, are present on a single-strand loop in the central region of the native molecule. Interest-

ingly, VTMoV and SNMV RNA 2 also contain the sequence GAAAC (residues 86 - 90 in both molecules) which is also present in a single-strand region in the centre of the proposed secondary structures. However, there is no extensive sequence homology or complementarity between viroids PSTV, CSV, CEV, ASBV and CCCV and RNA 2 of VTMoV and SNMV.

It will be of considerable interest to determine the exact mechanisms by which VTMoV and SNMV RNA 2 support the replication of the homologous RNA 1 species as well as the molecular basis for the specificity between RNA 1 and RNA 2. It seems likely that the common structures of VTMoV RNA 2, SNMV RNA 2 and viroids may reflect some common function so that determining the role of the RNA 2 molecule in replication may contribute to our understanding of viroid replication.

ACKNOWLEDGEMENTS

We thank Drs.A.R.Gould, R.I.B.Francki and J.W.Randles for materials and helpful discussions, and Miss J.L.Rosey for assistance. This work was supported by the Australian Research Grants Committee.

REFERENCES

1. Randles, J. W., Davies, C., Hatta, T., Gould, A.R. and Francki, R.I.B. (1981) *Virology* 108, 111-122.
2. Gould, A.R. and Hatta, T. (1981) *Virology* 109, 137-147.
3. Tien-Po, Davies, C., Hatta, T. and Francki, R.I.B. (1981) *FEBS Lett.* 132, 353-356.
4. Gould, A.R., Francki, R.I.B. and Randles, J.W. (1981) *Virology* 110, 420-426.
5. Diener, T.O. (1979) *Viroids and Viroid Diseases*, John Wiley and Sons, New York.
6. Gross, H.J. and Riesner, D. (1980) *Angew.Chem.Int.*Ed.Engl.19, 231-243.
7. Haseloff, J. and Symons, R.H. (1981) *Nucleic Acids Res.*9, 2741-2752.
8. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc.Natl.Acad.Sci. USA* 74, 5463-5467.
9. Efstratiadis, A., Vournakis, J.N., Donis-Keller, H., Chaconas, G., Dougall, D.K. and Kafatos, F.C. (1977) *Nucleic Acids Res.*4, 4165-4174.
10. Gould, A.R., Palukaitis, P., Symons, R.H. and Mossop, D.W. (1978) *Virology* 84, 443-455.
11. Symons, R.H. (1979) *Nucleic Acids Res.* 7, 825-837.
12. Gould, A.R. and Symons, R.H. (1982) Submitted for publication.
13. Sanger, F. and Coulson, A.R. (1978) *FEBS Lett.* 87, 107-110.
14. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
15. Goodman, H.M. and MacDonald, R.J. (1979) *Methods Enzymol.*68, 75-90.
16. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucleic Acids Res.*9, 309-321.
17. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J.Mol.Biol.* 143, 161-178.
18. Birnboim, H.C. and Doly, J. (1979) *Nucleic Acids Res.*7, 1513-1523.
19. Symons, R.H. (1981) *Nucleic Acids Res.* 9, 6527-6537.

20. Visvader, J.E., Gould, A.R., Bruening, G.E. and Symons, R.H. (1982) *FEBS Lett.* 137, 288-292.
21. Tinoco, I., Uhlenbeck, O.C. and Levine, M.D. (1971) *Nature* 230, 362-367.
22. Sänger, H.L., Klotz, G., Riesner, D., Gross, H.J. and Kleinschmidt, A.K. (1976) *Proc.Natl.Acad.Sci.USA* 73, 3852-3856.
23. Gross, H.J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H. and Sänger, H.L. (1978) *Nature* 273, 203-208.
24. Gould, A.R. (1981) *Virology* 108, 123-133.
25. Baralle, F.E. and Brownlee, G.G. (1978) *Nature* 274, 84-87.
26. Kozak, M. (1978) *Cell* 15, 1109-1123.
27. Kozak, M. (1979) *Nature* 280, 82-85.
28. Konarska, M., Filipowicz, W., Domdey, H. and Gross, H.J. (1981) *Eur.J. Biochem.* 114, 221-227.
29. Hollings, M., Stone, O.M., Barton, R.J. and Greber, R.S. (1979) Rep. Glasshouse Crops Res.Inst. 1978, 150-151.
30. Conjero, V. and Semancik, J.S. (1977) *Virology* 77, 221-232.
31. Davies, J.W., Kaesberg, P. and Diener, T.O. (1974) *Virology* 61, 281-286.
32. Semancik, J.S., Conjero, V. and Gerhart, J. (1977) *Virology* 80, 218-221.
33. Domingo, E., Sabo, D., Taniguchi, T. and Weissman, C. (1978) *Cell* 13, 735-744.
34. Holland, J.J., Grabau, E.A., Jones, C.L. and Semler, B.L. (1979) *Cell* 16, 495-504.
35. Donis-Keller, H., Browning, K.S. and Clark, J.M. (1981) *Virology* 110, 43-54.
36. Gross,H.J.,Krupp, G., Domdey, H., Raba, M., Jank, P., Lossow, C., Alberty, H., Ramm, K. and Sänger, H.L. (1982) *Eur.J.Biochem.* 121, 249-257.