CITRUS EXOCORTIS VIROID: NUCLEOTIDE SEQUENCE AND SECONDARY STRUCTURE OF AN AUSTRALIAN ISOLATE

Jane E. VISVADER, Allan R. GOULD, George E. BRUENING* and Robert H. SYMONS†
Department of Biochemistry, University of Adelaide, Adelaide, SA 5001, Australia

Received 10 December 1981

1. Introduction

Exocortis or 'scaly butt disease' of citrus is distributed world-wide [1]. The causative agent is citrus exocortis viroid (CEV) [2,3] which is a member of that unique group of plant pathogens, the viroids, of which only 8 have been described [1,4,5]. Like other members of the group, CEV consists of a single-stranded covalently closed circular RNA molecule which is highly base-paired, rod-like, infectious and non-encapsidated [1,4].

The primary sequence and proposed secondary structure of only 3 viroids have been reported so far: potato spindle tuber viroid (PSTV) with 359 residues [6]; chrysanthemum stunt viroid (CSV) with 356 residues [7]; and avocado sunblotch viroid (ASBV) with 247 residues [8]. We report here the sequence and proposed secondary structure of the 371 residues of an Australian isolate of CEV and discuss the significance of the extensive sequence homology which exists between PSTV, CSV and CEV.

2. Materials and methods

CEV (our strain A) was originally obtained from Dr R. van Velsen, South Australian Department of Agriculture, and propagated in chrysanthemum (*Chrysanthemum morifolium* cv. Bonnie Jean and Velvet Ridge), cuttings of which were kindly provided by Dr R. K. Horst, Cornell University. Linear and circular CEV were purified as for CSV [9] and ASBV [5]. [γ-^{32}P]ATP, [α-^{32}P]dCTP and [α-^{32}P]-dGTP were prepared as in [8,10].

The sequencing of RNA fragments obtained by partial RNase digestion of viroids followed methods described fully in [7,8].

CEV was cloned as double-strand DNA under C_1 containment conditions in the bacteriophage vector M13 mp83 (kindly provided by Dr J. Messing) by an approach to be detailed elsewhere. Briefly, linear CEV was polyadenylated with poly(A) polymerase [11] and transcribed into double-strand DNA using reverse transcriptase and oligo$(dT)_{10}$ as the primer for first-strand cDNA synthesis and self-priming for second-strand synthesis (Dr D. J. Kemp, personal communication). The product was cleaved with *Alu*I restriction endonuclease, the fragments purified by polyacrylamide gel electrophoresis and blunt-end ligated into the unique *Sma*I site of the replicative form of phage M13 mp83 using T4 DNA ligase [12, 13]. Phage and cells grown from the resultant white plaques obtained after transfection were screened for CEV inserts on nitrocellulose filters [14,15] using [5′-^{32}P]CEV fragments generated by limited RNase U_2 digestion of circular CEV [7]. Cloned CEV inserts were sequenced either by the Maxam and Gilbert method [16] or by use of the M13 specific primer $GTA_4CGACG_2C_2AGT$ (Collaborative Res. Inc.) [12]. A restriction enzyme fragment derived from the replicative form of the CEV insert was used as a primer for the sequencing of CEV by the dideoxy-nucleotide chain termination procedure [8,17].

3. Results

3.1. *Sequencing procedure*

The approach used successfully for the sequencing of CSV [7] and ASBV [8] was also used here. In brief, purified circular CEV was partially digested with RNases T_1, U_2 and A and the exposed 5′-ends

* Permanent address: Department of Biochemistry and Biophysics, University of California, Davis, CA 95616, USA

† To whom correspondence should be addressed

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$02.75

labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. The 5′-labelled fragments were fractionated by size on an 80 cm long denaturing polyacrylamide slab gel, 20–30 labelled bands from each gel were eluted and sequenced using the partial enzymic cleavage method detailed in [7,8]. Almost the entire sequence of CEV was obtained from the large number of overlapping fragments sequenced on 7 M urea gels [7,8]. Four regions of band compression were resolved by use of gels containing 98% formamide [7]. However, one region of band compression (residue no. 186,187, fig.1) could not be resolved due to its location ~100 residues from the labelled 5′-terminus of the fragment containing it; this length exceeded the resolving capacity of the gel system used.

The sequence of CEV was finalised and much of the pre-determined sequence confirmed after cloning CEV via double-strand DNA in the bacteriophage vector M13 mp83 [12,13]. For example, a cloned fragment (residues 252–362, fig.1) was sequenced by the dideoxynucleotide method with M13 universal primer [12] and by the Maxam and Gilbert method [16]. In addition, a restriction endonuclease fragment (residues 272–329) derived from the cloned insert by hydrolysis with *MspI* was used as a primer for the dideoxynucleotide sequencing on the circular CEV [8]. This latter method gave a further 143 residues of readable sequence from the end of the primer and resolved the final 2 residues (186,187) which showed band compression of the direct RNA sequencing gels. A total of 253 residues were confirmed and sequenced using cloned CEV inserts.

3.2. *Primary sequence and proposed secondary structure of CEV*

The complete base sequence of the 371 residues of CEV is given in fig.1; although CEV is circular, it is presented in linear form for convenience and for discussion of possible translation products (see below). A secondary structure model was constructed (fig.2) to provide maximum sequence and structural homology with the proposed secondary structures of PSTV [6] and CSV [7]. The numbering of CEV residues follows the convention established for PSTV [6].

The numbers of the 3 types of base pairs for the proposed CEV structure are compared with the corresponding values for CSV, PSTV and ASBV (table 1). Although the percentages of residues base-

```
CGGGAUCUUUCUUGAGGUUCCUGUGGUGCUCACCUGACCCUGCAGGCAGGAAAAGAAAAA
1         10        20        30        40        50        60

AGAGGCGGCGGGGGAAGAAGUCCUUCAGGGAUCCCCGGGGAAACCUGGAGGAAGUCGAGG
61        70        80        90        100       110       120

UCGGGGGGGACAGCUGCUUCGGUCGCCGCGGAUCACUGGCGUCCAGCGGAGAAACAGGAG
121       130       140       150       160       170       180

CUCGUCUCCUUCCUUUCGCUGCUGGCUCCACAUCCGAUCGUCGCUGAAGCGCCUCGCCCC
181       190       200       210       220       230       240

CUCGCCCGGAGCUUCUCUCUGGAGACUACCCGGUGGAAACAACUGAAGCUUCAACCCCAA
241       250       260       270       280       290       300

ACCGCUUUUCUUGUAUCUUCACUGCUCUCCGGGCGAGGGUGAAAGCCCUCGGAACCCUAG
301       310       320       330       340       350       360

AUUGGGUCCCU
361       370
```

Fig.1. Nucleotide sequence of CEV. The circular RNA is presented as a linear sequence with residue 1 corresponding to the left-hand end of the secondary structure model of fig.2. A bar beneath 3 residues indicates an initiation triplet and a bar above a sequence indicates a termination triplet.

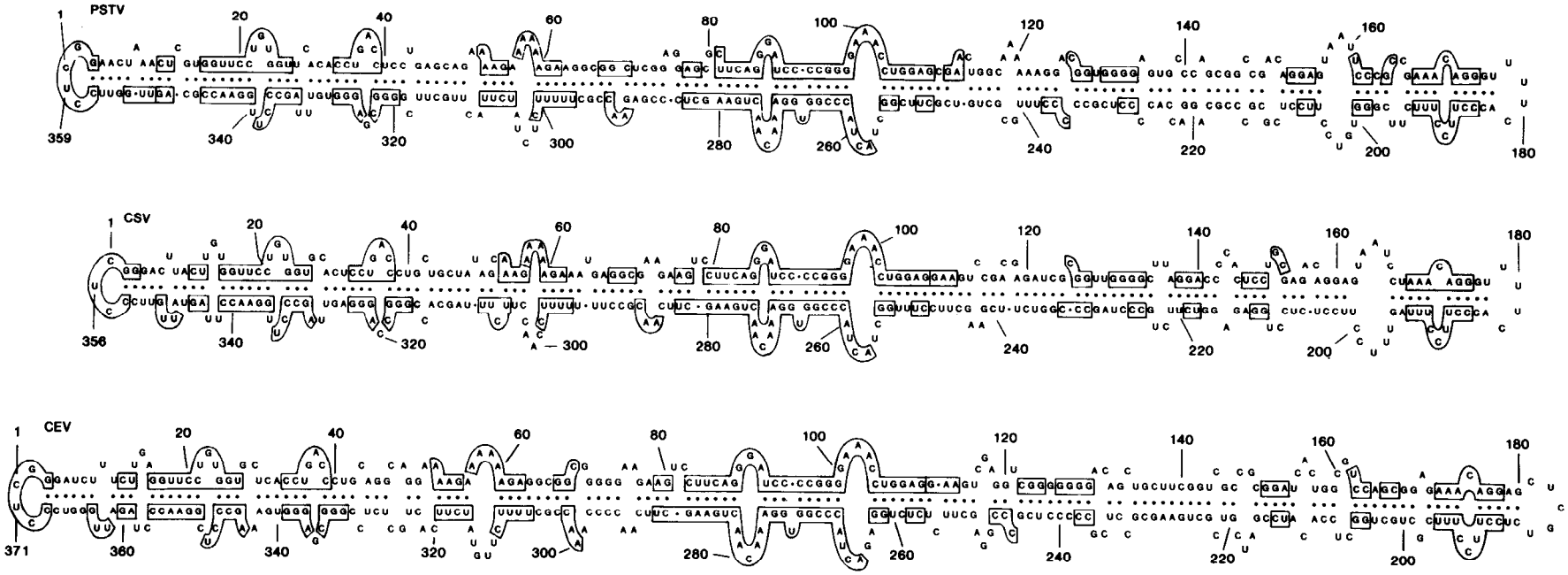

Fig.2. Proposed secondary structure of CEV together with those of PSTV [6] and CSV [7]. The boxed areas contain residues in CEV which are homologous in the other 2 viroids.

Table 1
Properties of proposed secondary structures for CEV, CSV, PSTV and ASBV

Viroid	No. of residues	No. of base pairs			G:C base pairs as % of total	Residues base paired	Ref.
		A:U	G:C	G:U			
CEV	371	34	72	18	58	67	This work
CSV	356	44	64	16	52	70	[7]
PSTV	359	37	73	16	58	70	[6]
ASBV	247	43	28	12	34	67	[8]

paired in the 4 structures are approximately the same, G:C base pairs make up a much lower proportion of the total in ASBV than in the other 3 viroids.

The homologous sequences shared by the 3 viroids CEV, CSV and PSTV are boxed in the structures of fig.2; 49% of CEV sequences are homologous with CSV and PSTV. The most notable feature of the homologous regions are the 2 long conserved sequences in the centre of each molecule; 28 residues on the top (residue no. 83–110 in CEV) and 26 residues on the bottom (residue no. 265–290 in CEV).

3.3. *Possible polypeptide translation products of CEV*

Although CEV does not act as a messenger RNA in cell-free translation systems [18] and there is no evidence for the in vivo synthesis of viroid-specific proteins [19], it is feasible that sub-genomic linear fragments of CEV may act as mRNAs. Hence, possible start and stop codons have been marked on the linear sequence of fig.1, while the positions of the start and stop codons in both the plus strand of CEV (the infectious strand) and the complementary minus strand are listed in table 2 together with potential polypeptide products.

Of the 10 start codons, only one of these is AUG and this occurs in the minus strand. There are 4 potential polypeptide products coded for by the plus strand (varying in length from 4–37 residues) and 5 potential products coded for by the minus strand (15–87 residues). UGA is the major stop codon while UAA is not used at all (table 2).

Table 2
Start and stop codons in plus and minus strands of CEV and potential polypeptide translation products

	Residue number at 1st residue					Polypeptide product (amino acid residues)
	Start codons		Stop codons			
	AUG	GUG	UAA	UAG	UGA	
Plus strand		23			35	4
		26			35	3
		273			13	37
		339			13	15
					225	–
					284	–
					340	–
				358		–
Minus strand	213				87	42
		322			268	18
		211			321	87
		156			87	23
		33		359		15

The residue numbers in the plus strand (infectious viroid) are retained in the minus strand; hence, residue numbers in the minus strand run in the 3′ to 5′ direction

4. Discussion

The sequence of 371 residues for CEV represents the largest sequence so far reported for a viroid (table 1). Its primary sequence and proposed secondary structure show striking regions of homology with the sequences and proposed secondary structures of PSTV and CSV (fig.2). Although 49% of CEV residues are homologous with PSTV and CSV in the structures of fig.2, there are additional sequences of CEV present in PSTV and not CSV and vice versa. Thus, 59% of CEV sequences are homologous with PSTV and 60% with CSV. The data are consistent with the suggestion [8] that there exists a group of viroids which share common sequences and secondary structures and which may have evolved from a common ancestral viroid. The highly conserved central region of PSTV, CSV and CEV (fig.2), which is also present in the 4 RNAs of cadang-cadang viroid (Haseloff, Mohamed and Symons, unpublished), is intriguing and may represent a region essential for the replication of these viroids. However, very little of this central region is conserved in the sequence of ASBV [8] which indicates that it should be allocated to a separate viroid group.

The approaches used for sequencing CEV were the same as those used for ASBV [8] and CSV [7]. Direct sequencing of 5′-^{32}P-labelled viroid fragments, prepared after partial RNase digestion, by the partial enzymic cleavage method rapidly yielded sequence data. However, a large number of fragments had to be sequenced to generate the necessary overlaps to complete the circular sequence since some fragments were poorly represented in the partial RNase digests. In addition, repeat digests were needed to resolve the occasional ambiguity in distinguishing C and U residues.

The difficulty in finding poorly represented regions of the viroid in partial RNase digests is the limiting factor in the rapid sequencing of viroids by the partial enzymic cleavage method, but it can be overcome by the additional approach reported here. Thus, sequence data derived by direct sequencing was used to predict restriction enzyme sites in double-strand DNA transcribed from CEV. DNA restriction fragments were then prepared and cloned into the replicative form of the bacteriophage M13 [12]. The cloned inserts were sequenced directly using a universal M13 DNA primer or after isolation from the double-strand replicative form by the chemical method of Maxam and Gilbert [12,16,17]. Furthermore, the cloned CEV inserts could be further restricted to generate specific DNA primers to prime dideoxynucleotide sequencing reactions on the viroid RNA to further extend the sequencing data. This additional approach using cloned viroid sequences thus allowed the complete sequence of CEV to be finalised and also provided further confirmation of 68% of the sequence data obtained by direct RNA sequencing.

Acknowledgements

The authors thank Miss J. L. Rosey for assistance, J. Haseloff for advice on RNA sequencing, Dr R. I. B. Francki for the use of glasshouse facilities, Dr R. van Velsen for providing an inoculum of CEV and Dr R. K. Horst for chrysanthemum cuttings. This work was supported by the Australian Research Grants Committee and the Rural Credits Development Fund of the Reserve Bank of Australia.

References

[1] Diener, T. O. (1979) Viroids and Viroid Diseases, Wiley, New York.
[2] Semancik, J. S. and Weathers, L. G. (1972) Nature New Biol. 237, 242–244.
[3] Sänger, H. L. (1972) Adv. Biosci. 8, 103–116.
[4] Gross, H. J. and Riesner, D. (1980) Angew. Chem. Int. Ed. Engl. 19, 231–243.
[5] Palukaitis, P., Hatta, T., Alexander, D. McE. and Symons, R. H. (1979) Virology 99, 145–151.
[6] Gross, H. J., Domdey, H., Lossow, C., Jank, P., Raba, M. and Alberty, H. (1978) Nature 272, 203–208.
[7] Haseloff, J. and Symons, R. H. (1981) Nucleic Acids Res. 9, 2741–2752.
[8] Symons, R. H. (1982) Nucleic Acids Res. in press.
[9] Palukaitis, P. and Symons, R. H. (1980) J. Gen. Virol. 46, 477–489.
[10] Symons, R. H. (1977) Nucleic Acids Res. 4, 4347–4355.
[11] Gould, A. R., Palukaitis, P., Symons, R. H. and Mossop, D. W. (1978) Virology 84, 443–455.
[12] Messing, J., Crea, R. and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309–321.
[13] Vieira, J. and Messing, J. (1981) personal communication.
[14] Grunstein, M. and Hogness, D. S. (1975) Proc. Natl. Acad. Sci. USA 72, 3962–3965.
[15] Thomas, P. S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201–5205.
[16] Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[17] Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[18] Semancik, J. S., Conejero, V. and Gerhart, J. (1977) Virology 80, 218–221.
[19] Conejero, V. and Semancik, J. S. (1977) Virology 77, 221–232.