Sequence analysis of cymbidium ringspot virus satellite and defective interfering RNAs

L. Rubino, J. Burgyan, F. Grieco and M. Russo

Dipartimento di Patologia Vegetale, Università degli Studi and Centro di Studio del CNR sui Virus e le Virosi delle Colture Mediterranee, Bari, Italy and Research Institute for Plant Protection, Budapest, Hungary

The nucleotide sequences of cloned cDNA copies of satellite and defective interfering (DI) RNAs of cymbidium ringspot virus were determined. DI RNA is 499 nucleotides long and is composed of six stretches of sequence derived from CyRSV genomic RNA. Four of these stretches share common 5' sequences and the 5' and 3' ends are identical to those of genome RNA. Satellite RNA is 621 nucleotides long. In some regions the sequence is very similar (60% to 100% identical) to genomic RNA. A consensus sequence is proposed to be involved in the replication of genomic, DI and satellite RNAs.

Introduction

The genome of cymbidium ringspot tombusvirus (CyRSV) is one positive-sense, single-stranded RNA molecule, 4733 nucleotides long, which is encapsidated in icosahedral particles (Grieco et al., 1989a; Martelli et al., 1989). Two subgenomic RNAs (about 2.1 kb and 0.9 kb) are also encapsidated and they can be translated in vitro to give the virus coat protein (Mr about 41 000) and a polypeptide with an Mr of about 22000, respectively (Russo et al., 1988; Grieco et al., 1989b). Full-length genomic RNA encodes a protein which has homology with the putative replicase of other viruses (Grieco et al., 1989a; Kamer & Argos, 1984). Particles of some isolates of CyRSV may also contain either a satellite (sat) RNA (Gallitelli & Hull, 1985) or a defective interfering (DI) RNA (Burgyan et al., 1989). Satellite and DI RNAs were so defined by the results of hybridization experiments which showed no sequence homology and high sequence homology respectively with genomic RNA (Gallitelli & Hull, 1985; Burgyan et al., 1989). Satellite and DI RNAs have now been cloned and sequenced. Sequence analysis showed that DI RNA is completely derived from the helper virus genome, whereas satellite RNA contains only a few sequences present in genomic and DI RNAs.

Methods

Virus propagation and purification. Isolates of CyRSV containing either satellite (Gallitelli & Hull, 1985) or DI RNA (Burgyan et al., 1989) were propagated in Nicotiana clevelandii and purified as described by Gallitelli et al. (1985).

DNA sequencing. Plasmid DNA was prepared as in Hattori & Sakaki (1986) and sequenced with T7 DNA polymerase (Sequenase, US Biochemicals) as previously described (Grieco et al., 1989b).

All the sequences except 80 nucleotides and 10 nucleotides at the 5' ends of DI and sat RNA respectively were determined by sequencing DNA in both orientations.
this region and for DI RNA priming was with a denatured restriction
the reaction was primed by the same oligonucleotide used for cloning
reverse transcription (Zimmern & Kaesberg, 1978); for satellite RNA
fragment (77 bp) from a DI RNA cDNA clone (positions 92 to 169 in
comparing its position to that of ribonucleotide markers visualized
under u.v. light.

Amersham) (England & Uhlenbeck, 1978) and sequenced by the
partial enzymic method (Donis-Keller
labelled RNA was exhaustively digested with ribonuclease T2 (BRL)
and chromatographed on thin-layer polyethylenimine cellulose plates
(Wurst et al., 1977). An aliquot of the 3'-
labelled RNA was exhaustively digested with ribonuclease T2 (BRL)
and chromatographed on thin-layer polyethylenimine cellulose plates
(Wurst et al., 1978). The labelled nucleotide was identified by
comparing its position to that of ribonucleotide markers visualized under u.v. light.
The 5' region was sequenced by dideoxynucleotide-terminated
reverse transcription (Zimmern & Kaesberg, 1978); for satellite RNA
the reaction was primed by the same oligonucleotide used for cloning
this region and for DI RNA priming was with a denatured restriction
fragment (77 bp) from a DI RNA cDNA clone (positions 92 to 169 in
comparing its position to that of ribonucleotide markers visualized
under u.v. light.

Sequence assembly and analysis. The sequence was compiled and
analysed with the DNA Strider program (Marck, 1988). Alignments
were made using the program GLORIA of the University of Bari. The
most stable secondary structures for RNA sequences were generated

Results

Cloning and sequencing of DI RNA

Although several clones of DI RNA were obtained, none
of them was full-length. The largest clone derived from
DNA constructed by random priming was 383 bp long;
all others were smaller and contained within it. Clones
prepared by extension of a primer complementary to the
3'-terminal 17 nucleotides did not extend beyond 350
nucleotides. Heating the RNA at 65 °C for 1 min and
quickly cooling it prior to the synthesis of cDNA did not
lead to the production of longer clones. The uncloned
part of the sequence was determined by primer extension
directly on the RNA (Fig. 1a). As with genomic RNA
(Grieco et al., 1989a) it was not possible to label and
determine the 5' terminus. The 499 nucleotide sequence
shown in Fig. 2(a) is fully representative of the DI
species size-fractionated from total RNA preparations as
all sequenced clones align perfectly with each other and
with the part of the sequence derived directly from
RNA. It seems that, at least under the inoculum and
growing conditions of the present study, the DI RNA
analysed consists of a stable, single type of molecule,
rather than a population of different molecules. As
shown in Fig. 2(b) DI RNA is made up of sequences
prepared by extension of a primer complementary to the
3'-terminal 17 nucleotides did not extend beyond 350
nucleotides. Heating the RNA at 65 °C for 1 min and
quickly cooling it prior to the synthesis of cDNA did not
lead to the production of longer clones. The uncloned
part of the sequence was determined by primer extension
directly on the RNA (Fig. 1a). As with genomic RNA
(Grieco et al., 1989a) it was not possible to label and
determine the 5' terminus. The 499 nucleotide sequence
shown in Fig. 2(a) is fully representative of the DI
species size-fractionated from total RNA preparations as
all sequenced clones align perfectly with each other and
with the part of the sequence derived directly from
RNA. It seems that, at least under the inoculum and
growing conditions of the present study, the DI RNA
analysed consists of a stable, single type of molecule,
rather than a population of different molecules. As
shown in Fig. 2(b) DI RNA is made up of sequences
from six different regions of CyRSV genomic RNA. The
first, second and third regions correspond to 165
nucleotides from the 5' terminus of genomic RNA with two deletions of 15 and eight nucleotides, the fourth region, 158 nucleotides in length, corresponds to nucleotides 1348 to 1508 of genomic RNA and the fifth and sixth regions are derived from the 3' region of genomic RNA with an intervening deletion of 182 nucleotides. The entire sequence of CyRSV DI RNA is therefore identical to parts of CyRSV genomic RNA (Grieco et al., 1989a, b) with the exception of four single-base substitutions, five single-base deletions and two single-base insertions. Most of the sequence of DI RNA (74%) belongs to non-coding regions of genomic RNA; the exceptions are four nucleotides of the third block which correspond to those in the 33K protein cistron, the fourth block, which corresponds entirely to the region encoding the 92K readthrough polypeptide and 26 nucleotides of the fifth block that correspond to part of the gene for the 22K protein. In the fourth region there is the only open reading frame (ORF) of 95 nucleotides which encodes a 32 amino acid polypeptide.

Alignments of the first 15 nucleotides (Fig. 3) of each conserved block of DI RNA (written as genomic RNA sequences) with satellite and subgenomic (sg) RNAs. (G at 5' end of genomic, DI and satellite RNA is preceded by an unknown nucleotide, which is omitted in this presentation). Homologies are underlined.

![Alignment](image)

Fig. 3 Alignment of the first 15 nucleotides of each conserved block of DI RNA (written as genomic RNA sequences) with satellite and subgenomic (sg) RNAs. (G at 5' end of genomic, DI and satellite RNA is preceded by an unknown nucleotide, which is omitted in this presentation). Homologies are underlined.

arrangements made for regions chosen randomly inside the deletions (not shown).

Alignment of CyRSV DI RNA with that of petunia asteroid mosaic virus (PAMV; also referred to as the cherry strain of tomato bushy stunt virus, Hillman et al., 1987) shows that the RNAs have the same basic construction, i.e. the blocks in DI RNA are derived from roughly the same regions of genomic RNA and that DI sequences are 80 to 90% identical.

Cloning and sequencing satellite RNA

Clones selected for sequencing are shown in Fig. 1(b). Clone pCS2/47 was divided into two subclones. The 3' end sequence was confirmed by partial enzymic digestion of pCp-labelled satellite RNA. Finally, the sequence of the 5' region was obtained by analysis of the clones obtained by primer extension and by direct dideoxynucleotide sequencing of the RNA. Primer extension in the absence of dideoxynucleotides confirmed the size of the molecule (621 nucleotides). It was not possible to carry out enzymic sequencing of the 5' end, as attempts to label this terminus were unsuccessful after dephosphorylation either without (Burgyan & Russo, 1988), or with prior treatment with tobacco acid pyrophosphatase (Promega). For the same reason it was not possible to identify the 5'-terminal nucleotide. The sequences obtained from each DNA clone agreed with one another and, where they overlapped, with the sequence obtained from direct RNA sequencing.

The primary structure of CyRSV sat RNA is shown in Fig. 4. Three major ORFs are present: ORF 1 starts at...
Fig. 5. Alignments of sequences of satellite (upper row) and genomic (lower row) RNA. Numbers refer to nucleotide positions in the respective RNA. In alignment 3 gaps have been introduced to optimize alignment.

nucleotide 25 and encodes 63 amino acids, ORF 2 starts at 333 and encodes 42 amino acids, ORF 3 starts at 376 and encodes 57 amino acids. None of these proteins was detected in vitro in previous translation experiments in rabbit reticulocyte lysates or in vivo in infected cowpea protoplasts in the presence of [3S]methionine (Burgyan et al., 1986) nor in in vitro translation experiments in the presence of [3H]leucine (not shown).

When the sequence of CyRSV satellite RNA was aligned with the sequence of genomic RNA (Fig. 5), two major regions of homology were found: nine out of the first 14 nucleotides at the 5' end were the same in the two sequences (Fig. 3) and 49 out of 53 nucleotides were identical between positions 187 and 239 of satellite RNA and nucleotides 99 and 153 of genomic RNA (Fig. 5). The same alignments were obtained when comparing satellite RNA with the DI RNA (nucleotides 2 to 14 and 77 to 131, respectively). In addition there were other small scattered regions of homology, as indicated in Fig. 5.

The secondary structure with the lowest free energy (−52.2 kJ/mol) as determined by computer analysis is shown in Fig. 6. The two separate arms are reminiscent of the folding of turnip crinkle carmovirus RNA C (Simon & Howell, 1986) and make this structure clearly different from the secondary structures proposed for other satellites, like that of cucumber mosaic cucumovirus (Gordon & Symons, 1983) or the small nepovirus satellites (Kaper et al., 1988).

Fig. 6 Most stable secondary structure of CyRSV satellite RNA at 25 °C.

RNA thus resembles mouse hepatitis virus (Coronaviridae) DI RNA (Makino et al., 1988) more than it does Sindbis virus (Togaviridae) DI RNA (Tsang et al., 1985). CyRSV and PAMV DI RNA also differ in the constitution of the junction points; four blocks in CyRSV DI RNA have a similar start but no initial consensus is detectable in PAMV DI RNA. Therefore, notwithstanding the overall similarity between the two sequences, the derivation of the DIs from the respective genomic RNAs may have different explanations. A 'copy-choice' mechanism has been suggested for the formation of DI RNA (Lazzarini et al., 1981). The presence of similar sequences at the beginning of each block of CyRSV DI RNA may be of importance as they may represent possible recognition sites of the replicase either as starting or termination points. No involvement of particular secondary structures at rearrangement sites has been detected in the present work.

The DI RNA examined in the present study has a unique size and sequence. However, this does not imply that only one DI RNA type is generated; perhaps a population of molecules is formed, but only one may be favoured for replication and/or have minimum require-

Discussion

Analysis of the primary structure of CyRSV DI RNA shows that the 5' end is completely derived from the 5' end of the helper virus genomic RNA. This differs from PAMV DI RNA, which begins 13 nucleotides inside PAMV genomic RNA (Hillman et al., 1987). CyRSV DI
ments for encapsidation. It should be kept in mind that because inoculations are made with purified virions, DI RNA molecules that have sizes and sequences suitable for packaging will be selected even if not very efficiently. The dynamics of this positive selection will be the subject of further investigations, as it may help to understand which of the conserved blocks are strictly necessary and which could be lost without impairing replication.

The common 5' end sequence in satellite, genomic and DI RNA, which is also repeated in some blocks of DI RNA, suggests an involvement with the common replicase. It is worth noting that the RNA species of the CyRSV system that have a high replication rate (i.e. genomic, DI and satellite RNA) have a 5' consensus NGAAAC****C/T*GG, whereas subgenomic RNAs, for which the possibility of autonomous replication has been suggested (Russo et al., 1988) but which do not replicate at a high rate, only have the initial part of this consensus (GA or GAA). All these species are encapsidated, but not with the same efficiency; DI RNA is the least efficiently encapsidated both in CyRSV (Burgyan et al., 1989) and PAMV (Hillman et al., 1987). The signal for efficient encapsidation may be localized in a sequence present in genomic, subgenomic and satellite RNAs, but not in DI RNA. The small stretches of nucleotides shown in Fig. 5 are candidates for this function.

Another region of homology between satellite, DI and genomic RNA is the 53 nucleotide stretch that also occurs in other sequenced tombusviruses, i.e. cucumber necrosis virus (Rochon & Tremaine, 1989) and PAMV (Hillman et al., 1987). The lack of hybridization between satellite and genomic RNA found in previous work (Gallitelli & Hull, 1985) was probably due to the probe used being randomly primed cDNA. In further tests we have found that nick-translated cDNA clones of satellite RNA hybridize weakly, but distinctly, with DI and genomic RNAs (not shown).

Little or no appreciable homology to helper virus genome is a characteristic of plant virus satellite RNAs (Francki, 1985; Kaper & Collmer, 1988). CyRSV satellite RNA has a major stretch of sequence almost identical to a part of the virus genome RNA that represents about 10% of the whole sequence. This situation is comparable, but to a limited extent, to that of turnip crinkle virus RNA C, which has a distinct, conspicious domain homologous to part of the helper genome RNA (Simon & Howell, 1986).

Both DI and satellite RNAs are capable of interfering with symptom expression in infected plants. For DI RNA, the reduced synthesis of genomic RNA has been suggested to be the principal cause of decreased symptom severity (Burgyan et al., 1989; Hillman et al., 1987). DI RNA may compete successfully with genomic RNA for replicase because of its smaller size and because it does not participate in other processes such as translation and formation of subgenomic RNAs. The same considerations may apply to the interference mediated by satellite RNA, the replication of which reduces the amount of genomic RNA (see for instance Fig. 1 in Burgyan & Russo, 1988). However, although there is no evidence for the functionality of the ORFs, it cannot be excluded that ORFs operate in vivo producing proteins with a role in the interference process.

References


(Received 6 February 1990; Accepted 9 April 1990)