Nucleotide sequence and gene organization of the 3'-terminal region of chrysanthemum virus B genomic RNA

K. Levay and S. Zavriev*

All-Union Institute of Agricultural Biotechnology, Moscow 127253, U.S.S.R.

DNA clones complementary to the 3'-terminal 3426 nucleotides of the genomic RNA of the carlaviruses chrysanthemum virus B (CVB) have been sequenced. The sequence contains six open reading frames (ORFs) which encode putative proteins (in the 5' → 3' direction) of Mr 25749 (ORF2), Mr 11435 (ORF3), Mr 6984 (ORF4), the triple gene block proteins, a protein Mr 34638 (ORF5); the coat protein, and a protein of Mr 12609 (ORF6). The latter protein is basic and contains a putative zinc finger motif. The 5'-proximal ORF1 encodes a product with substantial homology to the C-terminal portions of the putative RNA replicases of two other carlaviruses, potato viruses M and S. The analysis of the minus-sense sequence shows one ORF which encodes a polypeptide of Mr 16817. The sequenced portion of the CVB genomic RNA contains three short internal non-coding regions, two of which are typical for carlaviruses (those between ORFs 1 and 2, and between ORFs 4 and 5), and one (between ORFs 5 and 6) is unusual. There is significant similarity in the amino acid sequences of the CVB RNA-encoded proteins and the corresponding proteins of other carlaviruses.

Introduction

Chrysanthemum virus B (CVB) is a member of the carlaviruses group (Wetter & Milne, 1981) and is transmitted by aphids in a non-persistent manner and by inoculation with sap. CVB is widespread throughout the world in cultivated varieties of chrysanthemum. The virus particles are slightly flexed, rod-shaped, and 685 nm long and 12 nm in diameter. They consist of multiple copies of a coat protein (Mr 37000) encapsidating the viral RNA (Hollings & Stone, 1972).

Recently, there have been reports of the complete nucleotide sequence of the genomic RNA of carlavirus potato virus M (PVM) (Zavriev et al., 1991) and the 3'-terminal regions of other carlaviruses genomic RNAs such as those of potato virus S (PVS), lily symptomless virus (LSV), helenium virus S (HelVS) and carnation latent virus (CLV) (MacKenzie et al., 1989; Memelink et al., 1990; Foster et al., 1990; Haylor et al., 1990). These carlaviruses have an identical genome organization, which is similar to that of potexviruses.

In this report we present the sequence of 3426 nucleotides from the 3' terminus of the CVB RNA, and compare its genomic organization and the deduced amino acid sequences of the encoded proteins with those of other carlaviruses.

Methods

Virus purification and preparation of viral RNA. The CVB isolate was provided by Dr J. Richter (Institute of Phytopathology, Aschersleben, Germany). The virus was purified from Nicotiana clevelandii plants 3 weeks after inoculation as described by Hollings & Stone (1972). Viral RNA was isolated from the purified virus preparation using the phenol-SDS method (Proll et al., 1981) with minor modifications.

cDNA synthesis and sequencing. Double-stranded cDNA was synthesized essentially as described by Watson & Jackson (1986). The first strand of the cDNA was primed with oligo(dT)$_3$. Second-strand synthesis was accomplished with Escherichia coli DNA polymerase I in the presence of RNase H and E. coli DNA ligase. Double-stranded cDNA was made blunt-ended by treatment with T4 DNA polymerase, then ligated into Smal-cut alkaline phosphatase-treated plasmid pGEM-7z(f+) (Promega) and transformed into competent E. coli XL-1B cells. Ampicillin-resistant transformants were screened for clones containing the 3'-terminal sequence of CVB RNA by colony hybridization (Maniatis et al., 1982) with a $^{32}$P-labelled, oligo(dT)$_3$-primed CVB-specific cDNA probe.

The insert size of the selected recombinant cDNA clones was determined by restriction analysis of plasmid DNA. Finally, six clones containing long overlapping cDNA inserts (see Fig. 1) were isolated and used for sequencing. A nested series of exonuclease III deletions were generated from the selected cDNA clones according to the Erase-a-Base system (Promega). Didodeoxynucleotide sequencing of single- and double-stranded DNA templates was carried out using [α-$^{32}$P]dATP (Institute of Applied Chemistry, Leningrad) and modified T7 DNA polymerase (Sequenase; U.S. Biochemicals).
**Results and Discussion**

**Isolation and sequencing of viral cDNA clones**

More than 30 clones containing recombinant plasmids with CVB-specific cDNA inserts were isolated. Almost all of them contained sequences from the 3' terminus of CVB RNA including a poly(A) tract of variable length (46 to 95 nucleotides). Seven overlapping clones were selected for sequencing. One of them (pCVB 272) contained the complete sequence presented in this paper [3426 nucleotides excluding the 3' poly(A) tract].

The relative locations of the cDNA clones used for sequencing are shown in Fig. 1. The cDNA was sequenced in both directions and about 60% of the sequence was determined by examining two or three clones covering each particular region. The only nucleotide variation was found at position 2058 (see legend to Fig. 2).

**Organization of the 3'-terminal region of CVB RNA**

The nucleotide sequence revealed six potential ORFs in the 3' region of the positive RNA strand of CVB (Fig. 1 and 2). The long ORF1 at the beginning of the sequence ends at position 787 and encodes a polypeptide of at least $M_r$ 30255. The next three ORFs code for the proteins of the triple gene block which is found in all carla- and potexviruses except lily virus X (Memelink et al., 1990). ORF2, which extends from positions 821 to 1516, encodes a 231 residue polypeptide of $M_r$ 25749 (25K), ORF3 (positions 1494 to 1814) encodes a 106 residue polypeptide of $M_r$ 11435 (12K), ORF4 (positions 1805 to 1999) encodes a 64 residue polypeptide of $M_r$ 6984 (7K), ORF5 extending from positions 2054 to 3001 encodes a 315 residue polypeptide of $M_r$ 34638 (35K) and ORF6 extending from positions 3017 to 3340 encodes a 107 residue polypeptide of $M_r$ 12609 (12.6K).

Analysis of the negative RNA strand of the sequence reveals only one relatively long ORF (ORF7, Fig. 1) which encodes a putative 164 residue polypeptide of $M_r$ 16718 (17K). This ORF extends from the AUG codon at position 901 to the UAA stop codon at position 1395 of the negative strand.

The 3'-terminal non-coding region of the CVB RNA displays over 60% similarity to the 3' regions of five other carlaviruses, which confirms the reported conservation of this region in the carlaviruse group (Haylor et al., 1990).

**Similarities between CVB and other carlaviruses**

The results of the comparison of the CVB-encoded proteins with those of carlaviruses PVM, PVS, LSV, HelVS and CLV are given in Table 1.

The amino acid sequence of the ORF nearest the 5' end shows significant similarity with the C-terminal portions of the large proteins of carla- and potexviruses and apple chlorotic leaf spot closterovirus (data not shown). In particular, the CVB ORF1 polypeptide contains the sequence motif GDD (underlined in Fig. 2) which is conserved in the putative RNA replicases of most plant and animal positive RNA viruses (Poch et al., 1989). It is proposed therefore that ORF1 corresponds to the C-terminal part of the CVB RNA replicase as previously suggested for two other carlaviruses, PVS and PVM (MacKenzie et al., 1989; Morozov et al., 1990a).

ORF2 of CVB encodes a 25K protein analogous to those of carla- and potexviruses, which contains the consensus sequence of NTP-dependent DNA helicases (Gorbalenya et al., 1988) including the sequence GKS/T (underlined in Fig. 2). The hydrophobicity of the CVB proteins was analysed using the method of Kyte & Doolittle (1982). There are long hydrophobic stretches in the sequences of the CVB 12K and 7K proteins which are typical of the corresponding proteins of carla- and potexviruses and small non-virion proteins of some other...
**Sequence analysis of the putative protein encoded by the CVB negative-sense RNA**

The ORF for the putative 17K protein is located on the negative RNA strand within the region complementary to the coat protein gene (see Fig. 1 and Fig. 3(a)). Earlier we reported that the minus strand of PVM RNA contains an ORF at a similar location which encodes a putative 19K protein (Rupasov et al., 1990). The analysis of other known carlaviruses minus strand RNA sequences reveals that only HelVS RNA has a similar ORF which encodes a 12.5K protein (Fig. 3). A comparison of the amino acid sequences of these putative polypeptides shows that apart from the C-terminal third they are quite similar (Fig. 3b).

It should be noted that the minus strands of two potexvirus RNAs, those of white clover mosaic virus (Forster et al., 1988) and papaya mosaic virus (Sit et al., 1989), also contain ORFs similar in size and location to those in the above-mentioned carlaviruses.

**Putative CVB RNA nucleotide sequences essential for subgenomic RNA synthesis**

Alignment of the regions 5' to the initiation codons of the CVB 25K protein and the coat protein genes reveals pronounced homology (Fig. 4), which is similar to that observed earlier for PVM (Rupasov et al., 1989). An interesting peculiarity of all carlavirus genomic RNAs is the presence of the consensus sequence C/UUUAGGU (see Fig. 4) upstream of the AUG codons of the 25K protein and the coat protein genes. In all carlavirus RNA sequences known to date (MacKenzie et al., 1989; Memelink et al., 1990; Foster et al., 1990; Zavriev et al., 1991), this consensus sequence is found in only the two regions indicated above and may be essential for the initiation of carlavirus subgenomic RNA synthesis.

The most notable feature distinguishing CVB from other carlaviruses is the existence of a 17 nucleotide non-coding region between the coat protein and the 12.6K protein genes (see Fig. 2). This non-coding region and the
adjacent region downstream of the AUG codon of the 12.6K protein gene contain sequence motifs found in the vicinity of the AUG codon of the 25K protein gene (Fig. 4).

We would like to thank Drs S. Morozov, K. Kanyuka and A. Galkin for helpful discussions.

References


(Received 18 April 1991; Accepted 3 July 1991)