Complete nucleotide sequence of RNA 3 of rice stripe virus: an ambisense coding strategy

Yafeng Zhu, Takahiko Hayakawa, Shigemitsu Toriyama and Mami Takahashi

Plantech Research Institute, c/o MKC Research Center, 1000, Kamoshida-cho, Midori-ku, Yokohama-city 227 and National Institute of Agro-Environmental Sciences, 1-1, Kannondai, 2-chome, Tsukuba-city, Ibaraki 305, Japan

The RNA 3 of rice stripe virus (RSV) isolate T was cloned and its nucleotide sequence was determined. The complete primary structure was found to consist of 2504 nucleotides. One putative open reading frame (ORF), between nucleotides 2412 and 1444 in the 5'-proximal region of the virus complementary-sense RNA, encoded a 322 amino acid protein with an Mr of 35134 that was identified as the coat protein. The other ORF, between nucleotides 66 and 701 in the 5'-proximal region of the virus-sense RNA, encoded a 211 amino acid protein with an Mr of 23 874. An intergenic non-coding region (742 nucleotides) between the two ORFs contains oligo(A)- and oligo(U)-rich regions that can be arranged into a 126 base pair stem configuration. These results suggest that RSV RNA 3 has an ambisense coding strategy.

Introduction

Rice stripe virus (RSV), the type member of the tenuivirus group of plant viruses, has filamentous particles which appear folded, branched and supercoiled in the electron microscope and which are transmitted by planthoppers (Gingery, 1988; Toriyama, 1986a).

Four species of ssRNA and four species of dsRNA, named RNA 1 to RNA 4 in order of decreasing Mr, were obtained from purified virus particles. The ssRNAs were predicted to be counterparts of the dsRNAs (Toriyama & Watanabe, 1989; Ishikawa et al., 1989). The virus particles contain a single coat protein (CP) with an estimated Mr of 32K (Koganezawa et al., 1975; Toriyama, 1982) and a minor polypeptide with an estimated Mr of 230K, which may be a replicase (Toriyama, 1986b). Hybridization experiments with oligonucleotide probes, which had sequences deduced to be CP gene-specific from amino acid sequence analysis of the RSV CP, showed that the CP was encoded in RNA 3 (Hayano et al., 1990). Takahashi et al. (1990) showed that the sequences of around 20 nucleotides at the 5' termini of each RNA are complementary to those of the 3' termini and suggested that the RNA molecules form panhandle structures.

The molecular properties of RSV, however, are not well known. In this communication, we report the cloning and sequence analysis of RNA 3 of RSV isolate T and discuss the possibility that the RNA 3 of RSV is an ambisense RNA.

Methods

Preparation of viral RNA. The virus particles were purified from wheat leaves infected with RSV isolate T as described previously (Toriyama & Watanabe, 1989). The RNA was prepared from the virus particles by phenol extraction in 10mM-Tris-HCl buffer pH 7.5 containing 0.5% SDS and 0.1% 8-hydroxyquinoline (Toriyama, 1986b).

Cloning of ssRNA. A cDNA library of RSV was prepared by the method of Gubler & Hoffman (1983). For the synthesis of first-strand cDNA, a 15-mer oligonucleotide (5'ACACAAAGUCUGGGU 3') complementary to the 3'-terminal sequence of the virus ssRNA determined by Takahashi et al. (1990), was used as a primer. The double-stranded cDNA was made blunt with T4 DNA polymerase and was inserted into the SmaI site of pUC19 (Takara Shuzo) according to Henikoff (1984). Cloning of ssRNA. A cDNA library of RSV was prepared by the method of Gubler & Hoffman (1983). For the synthesis of first-strand cDNA, a 15-mer oligonucleotide (5'ACACAAAGUCUGGGU 3') complementary to the 3'-terminal sequence of the virus ssRNA determined by Takahashi et al. (1990), was used as a primer. The double-stranded cDNA was made blunt with T4 DNA polymerase and was inserted into the Smal site of pUC19 (Yanisch-Perron et al., 1985). Transformants of Escherichia coli strain DH5 were selected by ampicillin according to Hanahan (1985).

DNA sequence analysis. Deleted mutants of the cloned cDNA were obtained with a Takara kilosequence deletion kit (Takara Shuzo) according to Henikoff (1984). The cDNA sequences were determined by the dideoxynucleotide chain termination method using both a Takara M13 sequencing kit (Takara Shuzo) according to Sanger et al. (1977) and a T7 sequencing kit (Pharmacia) according to Smith et al. (1986). Computer analysis was performed using Genetyx (Software Development Co., Tokyo, Japan).

Northern blot analysis. The RSV genome RNAs purified from virus components M and B and the cucumber mosaic virus (CMV) genome RNAs (Mr markers) were treated with 50% formamide and 2.2 M-formaldehyde (Lehrach et al., 1977) in MOPS buffer (20 mM-MOPS, 5 mM-sodium acetate, pH 7, 1 mM-EDTA) and separated on 1%
agarose gels containing 20 mM-MOPS buffer and 2-2 M-formaldehyde. The RNAs were transferred onto a Hybond-N membrane (Amersham) using a vacuum blotting system (Vacugene, LKB-Pharmacia) and were hybridized with mixed probes containing cDNAs of RSV RNA 3 obtained in the present study and CMV RNA 1 (Hayakawa et al., 1989a) and RNA 3 (Hayakawa et al., 1989a). The probes were labelled with $^{32}$PdCTP (111 TBq/mmol) using a multiprimer labelling kit (Amersham).

**Preparation of coat protein.** Virus particles were treated with 0.3% SDS containing 2% 2-mercaptoethanol on ice for 30 min, and were applied to a Superose 12 column (Pharmacia) equilibrated with 50 mM-Tris-HCl buffer pH 6.8 containing 0.1% SDS and 4% 2-mercaptoethanol. Proteins were eluted with the same buffer. Two main peaks were observed by monitoring at 260 nm. The CP obtained from the second peak fraction was completely separated from the 230K protein of RSV.

**Amino acid analysis of CP.** The purified CP was digested by a lysylendopeptidase (Masaki et al., 1981) (Wako Pure Chemical Industries) in 50 mM-Tris-HCl buffer pH 9.0 containing 5 M-urea at 37°C for 5 h. The polypeptide fragments obtained were isolated by HPLC with a reverse sphere column (~ Bondasphere C18 5~t-300A, Waters). The fragments were hydrolysed with 6 M-HCl containing 1% (v/v) phenol at 105°C for 24 h and then labelled with phenylisothiocyanate (Wako Pure Chemical Industries) at room temperature for 20 min. Amino acid compositions were determined by HPLC with the PICO-TAG system (Waters). The amino acid sequences of the fragments were determined by a liquid-gas phase amino acid sequencer (Applied Biosystems).

### Results and Discussion

Using a 15-mer oligonucleotide primer, which anneals with the 3'-terminal region of RSV ssRNAs 2, 3 and 4, we obtained a cDNA clone (1-0 kb), the 3' end of which had a nucleotide sequence corresponding to that of RNA 3 determined by Takahashi et al. (1990). To obtain cDNA covering an extended region of RNA 3, we carried out primer extension and obtained a new clone, which was more than 1500 nucleotides in length. Twenty-four nucleotides from the 5' end of this clone overlapped and were identical to the sequence determined by direct sequencing (Takahashi et al., 1990). The clones, including their subclones deleted using the Takara kilosequence deletion kit, were sequenced in both orientations. Other clones were also partially sequenced for confirmation. The complete sequence of the 2504 nucleotides of RSV RNA 3 is presented in Fig. 1.

The $M_r$ of RNA 3 was estimated to be $1.8 \times 10^6$ for dsRNA and $1.1 \times 10^6$ for ssRNA (Toriyama & Watanabe, 1989). A similar estimation has been reported by Ishikawa et al. (1989). From these $M_r$s, the nucleotide lengths of the dsRNA 3 and the ssRNA 3 can be calculated to be about 2600 and 3100 bases, respectively. These values, especially the one estimated from the $M_r$ of ssRNA, are higher than those determined from cDNA in the present study. To eliminate the possibility of an internal deletion of the clones obtained, we re-estimated the $M_r$ of RNA 3 using denatured RNA and internal CMV RNA markers. Northern blot analysis showed that the RSV RNA 3 band (Fig. 2, lane 1) was closer to the CMV RNA 3 band (2.2 kb) than to the RNA 1 band (3.3 kb) (Fig. 2, lane 2). RNA molecular size markers also indicated that the length of RSV RNA 3 was just above 2.35 kb. These observations show that the full length of the cDNA of RSV RNA 3 was obtained in the present study. Toriyama & Watanabe (1989) and Ishikawa et al. (1989) estimated the $M_r$ of RNAs using 1.5% or 0.8% agarose gels for both ssRNAs and dsRNAs and so the $M_r$ may have been overestimated due to the complicated secondary structure, such as panhandles and hairpins, of the viral genomes (discussed later). The dsRNAs appear not to form a complicated higher structure, so the estimated value may be close to that determined in the present work.

A long open reading frame (ORF) was present in the 5'-proximal region of the virus complementary-sense RNA (vcRNA) between nucleotides 2412 and 1444 that encoded a 322 amino acid protein with an $M_r$ of 35134 (35K) (Fig. 1, lower part). The deduced $M_r$ was similar to that of CP as estimated by polyacrylamide gel electrophoresis (32K) (Toriyama, 1986a).

To confirm that the 35K protein was the CP, we analysed the amino acid composition of the peptides produced from coat protein by digestion with lysylendopeptidase. This proteinase cuts the carboxyl bond of the lysine residue, producing peptides containing a lysine residue at the carboxyl end with the exception of the last fragment of the protein. The peptide fragments were separated by HPLC. Thirteen fragments, including the last fragment of the coat protein, showed good agreement in amino acid composition with fragments predicted to arise from the nucleotide sequence (Fig. 1, underlined). In 12 out of 13 fragments, the ratio of the nearest integrated values of each amino acid was completely identical. In the longest peptide fragment, which had 16 amino acid residues, the ratio of the nearest integrated values of 14 amino acids was identical. The sequence analysis of two peptides gave the best indication that the 35K polypeptide is CP (Fig. 1, double underlined). These results indicate that the ORF deduced in the vcRNA contains the CP coding region.

No other long ORF was found downstream of the CP coding region. The next longest ORF encodes only 53 amino acids (residues 1487 to 1648) in vcRNA. However, an ORF (residues 66 to 701) was detected in the 5'-proximal region of virus-sense RNA (vRNA) which encoded a 211 amino acid protein with an $M_r$ of 23874 (Fig. 1, upper part). Although the function of this protein is unknown, it may be one of the stripe disease-
specific proteins that are found in plant tissues infected with RSV (Toriyama, 1986a). Auperin et al. (1984) showed that the nucleoprotein of arenavirus is encoded in the 5'-proximal region of the sequence complementary to the viral S RNA, corresponding to the 3' half of the S RNA, and a second gene product is encoded in the 5'-proximal region of the S RNA. They proposed the term 'ambisense RNA' to describe this novel coding strategy. In the present study, the RNA 3 of RSV isolate T was found to encode
proteins in the 5'-proximal regions of each of the strands. This suggests that the RSV RNA 3 could also be an ambisense RNA.

There is a non-coding intergenic region of 742 nucleotides between the two ORFs, and this contains oligo(A)- and oligo(U)-rich regions. These can be arranged into a 126 bp stem configuration, 72.6% of which forms base pairs (Fig. 3). This secondary structure may be involved in the stabilization of ssRNA or in the termination of mRNA transcription, as has been suggested by Auperin et al. (1984). Recently, de Haan et al. (1990) reported the ambisense character of RNA from tomato spotted wilt virus. The intergenic region of the RNA had an A-U-rich sequence and the formation of a stable hairpin structure was predicted. These characteristics are very similar to those of RSV RNA 3.

This work was supported in part by a Grant-in-Aid (Biocosmos Program) from the Ministry of Agriculture, Forestry and Fisheries [BCP-90-I-C-(2)]. We are grateful to Dr J. Kondo of Mitsubishi Kasei Corporation for his cooperation regarding amino acid sequencing.

References


(Received 10 September 1990; Accepted 2 January 1991)