Molecular analysis of rice dwarf phytoreovirus segment S11 corresponding to wound tumour phytoreovirus segment S12

Nobuhiro Suzuki,* Mariko Harada and Tomonobu Kusano

Laboratory of Plant Genetic Engineering, Biotechnology Institute, Akita Prefectural College of Agriculture, Ohgata, Akita 010-04, Japan

The complete nucleotide sequence of rice dwarf phytoreovirus (RDV) genome segment S11 was determined. S11 is 1067 nucleotides long. There is an inverted repeat of 10 bp adjacent to the conserved 5’-terminal hexanucleotide (5’ GGUAAA 3’) and 3’-terminal tetranucleotide (5’ UAGU 3’) sequences. A single large open reading frame found in the plus strand of S11 begins with the first AUG codon (bases 6 to 8) and extends for 567 bases. Evolutionary relatedness between RDV S11 and wound tumour phytoreovirus S12 based on amino acid sequence similarity (25.8%) was found. In addition to the first AUG triplet, RDV S11 possesses a second in-phase AUG triplet (positions 30 to 32) nearby, which conforms to the Kozak consensus sequence. Two forms of the protein were identified by using an in vitro transcription and translation system in which a tailored full-length cDNA was the initial template. The abolition of the first AUG codon by site-directed mutagenesis resulted in disappearance of the larger translation product. These results strongly suggest that the two products are translated from the first and second AUG codons. Whether the two proteins are expressed in vivo is at present unclear.

Introduction

Rice dwarf phytoreovirus (RDV) (Boccardo & Milne, 1984) has a genome consisting of 12 segmented dsRNAs, which are termed S1 to S12 in increasing order of electrophoretic mobility in acrylamide gels. Extensive sequence analyses of the RDV genome segments have been performed (Uyeda et al., 1987, 1989; Omura et al., 1988, 1989; Fukumoto et al., 1989; Hayashi & Minobe, 1990; Kano et al., 1990; Nakashima et al., 1990; Suzuki et al., 1989, 1990a, b; Yamada et al., 1990). Only the sequences of S1, S2, S11 and S12 remain to be determined. All sequenced segments are shown to have single long open reading frames (ORFs) and terminal sequence domains (Anzola et al., 1987), and share encoded amino acid sequence similarities, albeit to different extents, with the equivalent wound tumour phytoreovirus (WTV) segments (Anzola et al., 1989; Nakashima et al., 1990; Suzuki et al., 1990a).

In this study, we present the complete nucleotide sequence of RDV S11 and show amino acid sequence homology between gene products of RDV S11 and WTV S12. Moreover, we analysed the RDV S11-encoded translation product by using an in vitro transcription and translation system and site-directed mutagenesis.

Methods

cDNA library construction and screening. A cDNA library of RDV transcripts was constructed as described by Suzuki et al. (1990b) except that blunt-ended cDNA was cloned into the SmaI site of pUC18 (Norrander et al., 1983) without ligation to the EcoRI adaptor. The probe used in the first screening procedure was the 3'-32P-labelled S11 genomic RNA which was isolated from a 3.5% polyacrylamide gel (Suzuki et al., 1990b).

Nucleotide sequencing. Plasmids deleted at intervals of about 200 bp were produced from an almost full-length cDNA clone, p11RD1, by digesting with exonuclease III (Yanisch-Perron et al., 1985). The deleted cDNAs were sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977), using a Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical) in which 7-deazaguanosine triphosphate was utilized instead of dGTP (Mizusawa et al., 1986). The entire sequence of the cDNA was analysed in both directions. For RNA sequencing, the oligodeoxynucleotide (5’ GCATGCCAACGAATGACTC 3’) complementary to bases 42 to 60 of S11 mRNA was synthesized on a Biosearch Cyclone DNA synthesizer and used for primer extension on an RNA template in the presence of ddNTP (DeBorde et al., 1986). The nucleotide and amino acid sequence analyses were performed with the aid of SDG-GENETYX genetic information-processing programs (Software Development Company).
In vitro transcription and translation. A full-length cDNA of S11 was synthesized by the polymerase chain reaction (PCR) (Saiki et al., 1988). Clone p11RD1, lacking the extreme 5' 11 nucleotides, was excised from the pUC18 polylinker site by digesting with KpnI and BamHI and was submitted to 40 PCR cycles (1 min at 92 °C, 2 min at 50 °C and 4 min at 72 °C) followed by a 10 min final extension at 72 °C, using a Model TC-100 thermal cyclic reactor (Hoei Science Company). The reaction mixture (100 µl) was 10 mM-Tris-HCl pH 8.3, 50 mM-KCl, 1.5 mM-MgCl2, 0.001% gelatin, 100 ng of the excised p11RD1, 2.5 units of *Thermus aquaticus* DNA polymerase (Takara Shuzo) and 2 µg each of the synthesized primers, 5' TCCCCGCGTAAATGATGTTGGAA- CATTTACCCGTGGCATAGCCGG 3' and 5' CCGGTACCATCATATAGTGTTGGTGT 3' (sequences underlined are DNA was recloned into the pUC18 plasmid (Bullock et al., 1987), and then used for transformation of *Escherichia coli* strain XL1-Blue (Bullock et al., 1987).

The primer 5' GCCGGCGTAATGATGTTGGAA- CATTTACCCGTGGCATAGCCGG 3' was used in PCR for site-directed mutagenesis (the base T indicated by an asterisk was introduced in place of A in order to abolish the first AUG codon). The cloning procedure was the same as that stated above. For in vitro transcription under control of the T3 or T7 promoter, 3 µg of the DNA was digested with SacI or KpnI and then incubated for 30 min in 40 mM-Tris–HCl pH 7.5, 50 mM-NaCl, 8 mM-MgCl2, 2 mM-spermidine, 400 µM-UTP, 400 µM-CTP, 400 µM-ATP, 120 µM-gTP, 30 mM-DTT and 10 units of either T3 or T7 RNA polymerase. A sample (3 µg) of the transcript obtained was subjected to in vitro translation in wheatgerm extracts, SDS–PAGE and fluorography as described by Shirako & Ehara (1986).

**Results and Discussion**

**Nucleotide sequence of RDV S11**

About 300 cDNA clones were randomly selected from the library constructed. Recombinant plasmids were extracted by the rapid small-scale procedure (Maniatis et al., 1982) and then probed with labelled S11 genomic RNAs. As a result, many plasmids hybridized and most of them contained inserts of approximately 1 kb. Of these, the longest cDNA clone (p11RD1) covering bases 12 to 1067 was subjected to sequence analysis. The physical map of p11RD1 and the sequencing strategy are shown in Fig. 1. It was ascertained as stated by Suzuki et al. (1990a) that p11RD1 covered the 3' terminus of S11 mRNA. The nucleotide sequence present at the 5' terminus (bases 1 to 11) was determined directly by primer extension. The complete nucleotide sequence of S11 (1067 nucleotides) is shown in Fig. 2(a). From the SDS–PAGE pattern of S11 and S12 (Suzuki et al., 1990a), it was presumed that S11 was larger than S12. However, sequence analysis indicated that S11 is longer by only one base (N. Suzuki, M. Harada, M. Sugawara & T. Kusano, unpublished results) which is consistent with the observation that S11 and S12 comigrate under denaturing conditions (Suzuki et al., 1990a).

S11 contains the terminal sequence domain consisting of the conserved terminal sequences (5' GGUA AAA... UGAU 3') and a segment-specific inverted repeat (10 bp) adjacent to the terminal sequences. This domain is thought to play an important role in sorting and particle assembly (Anzola et al., 1987).

**S11 product assignment**

S11 has two nearby in-phase AUG codons. The longer ORF extends from bases 6 to 572 (Fig. 2a), encoding 189 amino acids with an *M* of 20759. The sequence (UXXAUAGA) flanking the first AUG codon (bases 6 to 8) is not the favoured initiation sequence which must have a purine (G or A) at positions −3 and +4 (Kozak, 1981). The second in-phase AUG codon (bases 30 to 32) does conform to the Kozak consensus sequence. In order to identify the presumed S11-encoded protein (P11), we tried to detect the translation products of the *in vitro*-synthesized transcript from p11RD1 which had been recloned into pBluescript SK(+) without the first AUG codon. However, its translation efficiency was very low, presumably because it lacked the 5'-terminal 11 nucleotides. Thus, we used a tailored full-length cDNA, the transcript of which would have 26 and 1 additional nucleotides at the 5' and 3' termini, respectively. Consequently, synthesis of two polypeptides with *M* of about 23K (P11a) and 24K (P11b) was directed by the plus strand transcript of the full-length cDNA (Fig. 3a, lane 1). No major polypeptide was detected from the minus strand transcript (Fig. 3a, lane 2) even when the gel concentration was increased, although there was an ORF (bases 449 to 835) on the S11 minus strand. The two products appeared to be translated from the first and second AUG codons, just as in the case of the simian rotavirus SA11 genome segment 9 (Chan et al., 1986). In order to examine this, we obtained a transcript lacking the first AUG using site-directed mutagenesis and *in vitro* transcription, and subjected it to *in vitro* translation. Only the smaller product, P11a, was detected, showing that the larger polypeptide was translated from the first AUG codon (Fig. 3a, lane 3). P11a, synthesized by the
Molecular analysis of RDV S11

Fig. 3. (a) Fluorogram of SDS-PAGE of translation products in wheatgerm extracts directed by transcripts synthesized in vitro. Samples applied to a 10% polyacrylamide gel were products of the plus (lane 1) and minus (lane 2) strand RNAs from the tailored full-length cDNA of S11, the plus (lane 3) and minus (lane 4) strand RNAs from the authentic RDV S11 mRNA, and endogenous RNA (lane 5) as a negative control. The protein size markers used were: phosphorylase b (97.4K), BSA (66.0K), ovalbumin (45.0K), carbonic anhydrase (29.0K) and soybean trypsin inhibitor (20.1K). (b) Genetic organization of RDV S11. P11a and P11b shown in schematic form. (1) The transcript from the full-length cDNA; (2) the authentic RDV S11 mRNA; (3) the transcript from the mutated cDNA. The 5' cap structures are not shown.

tailored full-length transcript (Fig. 3a, lane 1), was considered to be translated by ribosomes that scanned past the first AUG triplet. The mapping of P11a and P11b is illustrated in Fig. 3(b). Whether the two products are expressed in vivo is unknown. As described by Kozak (1987), an AUG codon within 10 nucleotides of the cap site is not recognized efficiently by the 40S ribosomal subunit. Therefore, the first AUG codon, if utilized at all, might be used less efficiently since it is located between bases 6 and 8 of the authentic S11 mRNA (Fig. 3b).

The Mr of P11a (19,988) and P11b (20,759) as calculated from the nucleotide sequence are smaller than the 23K and 24K expected from electrophoretic mobilities which is often observed in SDS–PAGE (Sano et al., 1978; Nuss & Dall, 1990).
Relatedness between RDV S11 and WTV S12

The previous sequence analyses (Anzola et al., 1989; Dall et al., 1989; Nakashima et al., 1990; Suzuki et al., 1990a) showed the corresponding segment assignment of the two viruses to be: RDV S4—WTV S4, RDV S5—WTV S5, RDV S6—WTV S6, RDV S7—WTV S7, RDV S8—WTV S8, RDV S9—WTV S11 and RDV S10—WTV S10. Here, we show that RDV S11 corresponds to WTV S12.

RDV S11 and WTV S12 have similar sized 5' non-coding (29 nucleotides for RDV and 34 nucleotides for WTV) and coding regions (543 nucleotides for RDV and 534 nucleotides for WTV), assuming that P11 is translated from the second AUG codon, but different sized 3' non-coding regions, as seen in equivalent segments of the two viruses (Anzola et al., 1989; Suzuki et al., 1990a). The 3' non-coding region of RDV S11 is extremely long (495 nucleotides) and corresponds to 46-4% of its genome size, which is uncommon for viral mRNAs, whereas that of WTV S12 is 283 nucleotides long (33-3%). The length of the 3' non-coding region of the other segments ranges from 3-1% to 17-7% for RDV and from 5-6% to 14-9% for WTV (Nuss & Dall, 1990). Nucleotide sequence identity is 46-6% and amino acid identity is 25-8% with only one gap (Fig. 2b). The N-terminal halves of the proteins show greater homology. RDV S11 is considered to encode a non-structural protein based on the fact that no polypeptide with an Mr similar to that of P11 is detected from the purified RDV preparation, and that RDV S11 corresponds to WTV S12 which has been shown to encode a non-structural protein (Xu et al., 1989).

In addition to the long ORF discussed above, WTV S12 has a short ORF consisting of 40 codons located 102 nucleotides downstream from the large ORF (Nuss & Dall, 1990). RDV S12 also has a short ORF, consisting of 27 codons flanking the large ORF, although amino acid sequence identity with a score higher than 32 (Schwartz & Dayhoff, 1978) cannot be found between the two short ORFs. At present it is unknown whether these short ORFs are expressed in vivo.

In the present study, we have presented the complete nucleotide sequence of RDV S11 mRNA, which encodes a putative non-structural protein of 181 or 189 amino acids. We have also suggested that RDV S11 and WTV S12 have the same origin in terms of molecular evolution.

The authors are grateful to Drs Donald L. Nuss, Hiroshi Sano and Kripamoy Agusan for their fruitful discussion, Dr Marilyn Kozak for information on 3' non-coding regions of viral mRNAs, Mr Masahiro Yamanaka for his gift of RDV-infected rice plants and Ms Kouko Futada for typing the manuscript. This work was partly supported by Yamashita Tarō Memorial Foundation for Education and Science, Inoue Foundation for Science and a Grant-in-Aid for Scientific Research (02856013) from the Ministry of Education, Science and Culture, Japan.

References


Molecular analysis of RDV S11


(Received 15 February 1991; Accepted 5 June 1991)


(Received 15 February 1991; Accepted 5 June 1991)