Journal of General Virology (1994), **75**, 2803–2806. *Printed in Great Britain*

Nucleotide sequence of *Nilaparvata lugens* reovirus genome segment S8 coding for the major outer capsid protein

Nobuhiko Nakashima* and Hiroaki Noda

National Institute of Sericultural and Entomological Science, Owashi, Tsukuba, Ibaraki 305, Japan

The complete nucleotide sequence of genome segment 8 (S8) of *Nilaparvata lugens* reovirus (NLRV) was determined. It consisted of 1802 nucleotides containing a long open reading frame (562 amino acids), which was expressed in *Escherichia coli* as a fusion protein. The expressed S8 product, a 62K protein, was detected by Western blotting using IgG directed against intact NLRV particles. This result indicates that S8 encodes the major outer capsid protein of NLRV. The protein exhibited 18·6% amino acid sequence identity with the predicted translation product of S10 of rice black-streaked dwarf virus.

The family *Reoviridae* is composed of nine genera: *Orthoreovirus*, *Orbivirus*, *Coltivirus*, *Rotavirus*, *Aquareovirus*, *Cypovirus*, *Phytoreovirus*, *Fijivirus* and *Oryzavirus* (Holmes *et al.*, 1994). The reovirus genome consists of 10 to 12 segments of dsRNA; each genome segment has conserved terminal sequences specific to the genus (Kudo *et al.*, 1991). The members of *Phytoreovirus*, *Fijivirus* and *Oryzavirus* genera are phytopathogenic viruses transmitted by leafhoppers and planthoppers (Nault & Ammar, 1989). Members of the *Fijivirus* genus, such as maize rough dwarf virus (MRDV) and rice black-streaked dwarf virus (RBSDV), have 10 dsRNA genome segments and their conserved terminal sequences of the plus strand are 5′ AAGUUUUUU---UGUC 3′ (Azuhata *et al.*, 1992). MRDV and RBSDV are transmitted by planthoppers and cause disease in maize and rice plants, respectively.

Nilaparvata lugens reovirus (NLRV) was isolated from a healthy colony of the brown planthopper, *N. lugens* (Noda *et al.*, 1991*a*). NLRV has 10 dsRNA genome segments and the conserved terminal sequences in the plus strand are 5′ AGU---GUUGUC 3′. These sequences are similar to those of members of the *Fijivirus* genus but lack the 5′-terminal adenine present in the latter (Noda *et al.*, 1994). This suggests that NLRV is closely related to the members of the *Fijivirus* genus, although rice plants sucked by NLRV-infected planthoppers show no symptoms.

Among the members of the *Phytoreovirus* genus, the nucleotide sequences are well-investigated and are similar among corresponding segments. For example, the amino acid sequence identities between the major outer capsid proteins (MOCPs) range from 48 to 56% (Omura *et al.*, 1989; Xu *et al.*, 1989; Noda *et al.*, 1991*b*). In the *Fijivirus* genus, the nucleotide sequences of MRDV genome segment 6 (S6) (Marzachi *et al.*, 1991) and RBSDV S10 (Uyeda *et al.*, 1990), S8 and S7 (Azuhata *et al.*, 1993) have been reported, but those of their MOCPs are unknown. We previously reported the nucleotide sequence of NLRV S10 (Noda *et al.*, 1994). Here we report the sequence of S8 together with evidence that it encodes the major outer capsid protein.

The cDNA library of NLRV genome segments cloned into pBluescript II and the method for screening the clones containing a S8 segment by using the enhanced chemiluminescence oligonucleotide labelling and detection system (ECL system; Amersham), have been described (Noda *et al.*, 1994). Using the ECL hybridization assay, with genomic RNA of NLRV S8 as a probe, 10 clones were obtained. Two clones that had the full-length cDNA of S8 were used for sequencing. The nucleotide sequence of S8 was determined by the *Taq* dye primer cycle sequencing method according to Noda *et al.* (1994).

The complete sequence of S8 was 1802 nucleotides (nt) long (Fig. 1). The GC content of S8 RNA was 35·3%. A long open reading frame (ORF) encoding 562 amino acids was identified in one of the strands. There were no long ORFs found in the opposite strand. The ORF contained two possible in-frame start codons, the first one located at nt 7 to 9, and the second one at nt 13 to 15. The second AUG appears to be suitable as a

The nucleotide sequence data for NLRV S8 reported in this paper will appear in the DDBJ, EMBL and GenBank databases under the accession number D26127.

0001-2302 © 1994 SGM

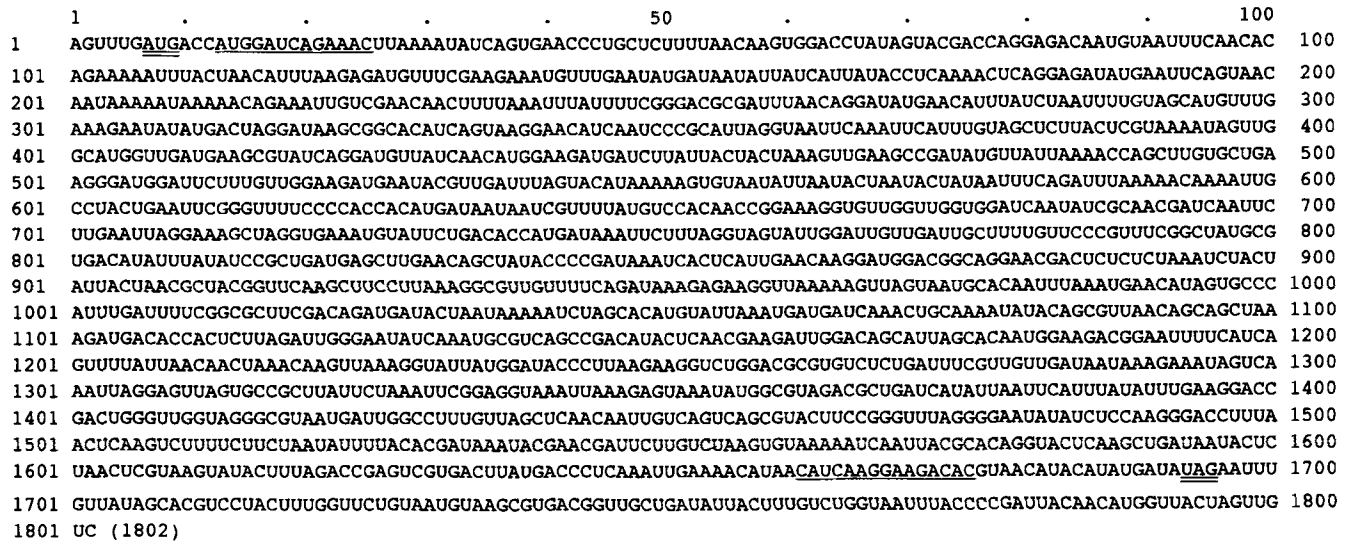

Fig. 1. The nucleotide sequence of genome segment S8 of NLRV. Double underlined residues indicate start and terminal codons. Single underlined region indicates the positions from which primers were derived, for use in PCR amplification.

functional initiator codon because of its flanking sequence, ACCAUGG, which is the optimal sequence for initiation by eukaryotic ribosomes (Kozak, 1986). However, the AUG existing within 10 bases from the 5′ end of mRNA is not well-recognized (Kozak, 1987). In-frame AUGs in the 5′-terminal region are also observed in S11 of rice dwarf virus (RDV; Suzuki *et al.*, 1991).

The M_r of the predicted protein encoded by NLRV S8 was 62·4K, which is close to that of the MOCP of NLRV, i.e. 64K. The 64K protein is the smallest structural protein of NLRV and exists in the outer capsid (Noda *et al.*, 1991*a*). To examine the assumption that S8 encodes the 64K protein, the S8 ORF was expressed in *Escherichia coli* (strain TB1) using an expression vector, pMAL-c2 (New England Biolabs). This vector is designed to produce a fusion protein with the maltose-binding protein (MBP). By insertion of the specific recognition sequence of factor Xa (a type of protease), between the MBP and a foreign protein, it is possible to separate the two. The cDNA of S8 ORF was amplified using PCR (Innis & Gelfand, 1990). The PCR primers, 5′ AGGATCCATGGATCAGAAAC 3′ (plus strand) and 5′ ACTGCAGGTGTCTA*CCTTGATG 3′ (minus strand) were synthesized with a Model 392 DNA/RNA synthesizer (Applied Biosystems). The position of each primer is shown in Fig. 1. Since the flanking sequence of the terminal codon was AT-rich, the minus-strand primer was derived from the upstream region. To stop synthesis of the fusion protein, an artificial stop codon was introduced in the primer of the minus strand by a point mutation at residue 1669 (shown by an asterisk in the above primer sequence of the minus strand). The 5′ regions of the plus and minus strand primers contain restriction sites, of *Bam*HI and *Pst*I respectively, which are underlined in the primer sequences shown above; these are to facilitate insertion of the amplified ORF sequence to the polycloning site of pMAL-c2. Insertion of the amplified DNA in the obtained transformants was confirmed by restriction enzyme analysis. The production and extraction of the MBP–S8 fusion protein were undertaken according to the manufacturer's recommendation. After 3 h, the induced cells were collected and suspended in column buffer (10 mM-Tris–HCl pH 7·4, 200 mM-NaCl, 1 mM-EDTA). The cells were broken by sonication and centrifuged to remove the debris. The supernatant was then subjected to affinity column chromatography utilizing the binding ability of amylose resin to MBP. The eluate was concentrated by ultrafiltration using Suprec-02 (Takara Shuzo), and then treated with factor Xa to separate the S8 product from the MBP. The reaction mixture was analysed directly by SDS–PAGE and Western blotting. For the latter, proteins were transferred from the polyacrylamide gel to a nylon membrane (Hybond-N^+; Amersham) using a semi-dry transfer apparatus (Bio-Rad). IgG directed against intact NLRV particles and peroxidase-labelled goat anti-mouse IgG (Bio-Rad) were used for detection of the expressed S8 product.

The results of SDS–PAGE and Western blotting of the expressed proteins are shown in Fig. 2. The predicted M_r of the fusion protein is 105K because the sizes of MBP and the S8 product are 43K and 62K, respectively. After the treatment with factor Xa, the band corresponding to 105K (Fig. 2*a*, lane 1) was no longer visible and a 62K band (Fig. 2*a*, lane 2) was present. The bands, 105K and 62K, were also detected by Western blotting (Fig. 2*b*,

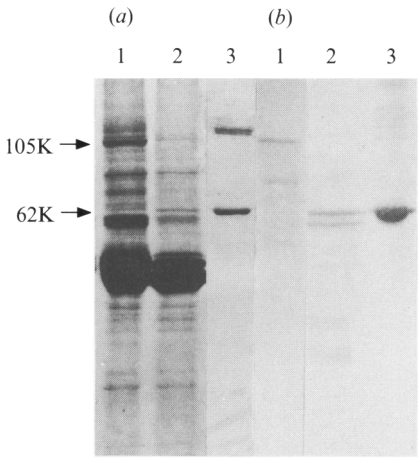

Fig. 2. SDS–PAGE (10%) (*a*) and Western blotting analysis (*b*) of MOCP of NLRV produced by *E. coli*. Lane 1, MBP–S8 fusion proteins after MBP affinity chromatography; lane 2, proteins separated from MBP after treatment with factor Xa; lane 3, structural proteins of purified NLRV particles. Antiserum against intact NLRV particles were used for the Western blotting.

lanes 1 and 2). The positive reaction of the S8 products with IgG directed against intact NLRV particles, together with the fact that the M_r of the S8 product was close to that of MOCP of NLRV, indicates that S8 encodes a structural protein. Other bands were also visible by Western blotting analysis, but these seem to represent degraded products because none were detected with M_r values above 105K and 62K (Fig. 2*b*, lanes 1 and 2, respectively).

Azuhata *et al.* (1993) reports that MRDV S6 corresponds to RBSDV S7 and MRDV S7 to RBSDV S8 in the two closely related members of the *Fijivirus* genus. However, the genome segment encoding a MOCP is unknown in the fijiviruses. In computer analysis using FASTA programs (Pearson, 1990), the amino acid sequence of the predicted protein encoded by NLRV S8 showed 18·6% identity with that of RBSDV S10 in a 531 amino acid overlap, and 18·4% identity with that of RBSDV S8 in an 87 amino acid overlap. Comparison with the MOCPs of the members of the genus *Phytoreovirus* showed 12·7% identity with that of RDV in a 166 amino acid overlap and 29% identity with that of wound tumour virus in a 62 amino acid overlap. However, there were no long identical predicted amino acid sequences found between MOCPs of NLRV and the three members of the genus *Phytoreovirus*. When we undertook dot matrix comparisons of the predicted amino acid sequence of the proteins encoded by NLRV S8 with that by RBSDV S8 and with those by phytoreoviruses, identical regions were far from diagonal lines (data not shown).

Pearson (1990) reports that sequences sharing more than 20 to 25% identity over their entire length almost always share a common ancestor, and it is possible to

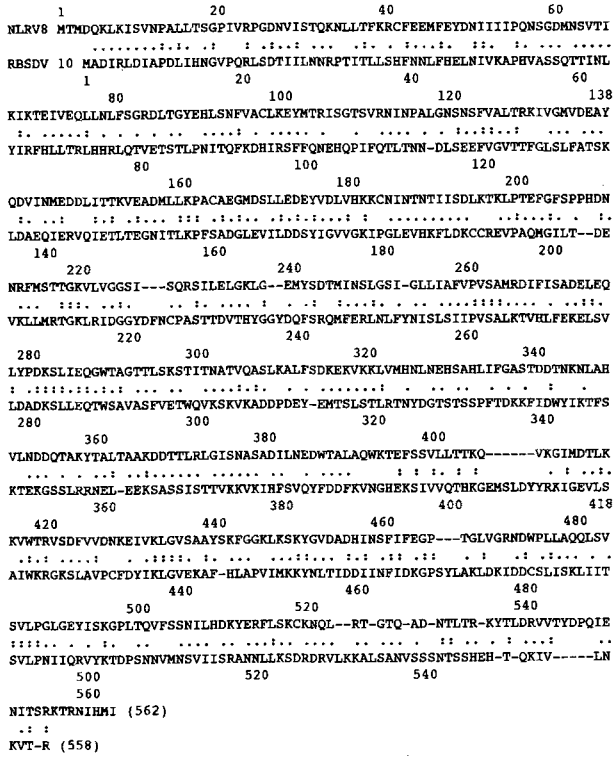

Fig. 3. Comparison of the amino acid sequences of the predicted proteins encoded by NLRV S8 and RBSDV S10 (Uyeda *et al.*, 1990). The sequences were aligned by the LFASTA program (Pearson, 1990) from the amino acid residues 7 to 522. The outer terminal sequences are manually aligned.

show convincingly that sequences that share as little as 15% identity over their entire length are homologous. Taking his conclusion into consideration, NLRV S8 may correspond to RBSDV S10. An alignment of the predicted amino acid sequences of the proteins encoded by NLRV S8 and RBSDV S10 revealed that there were some identical sequences within a distance of six amino acid residues (Fig. 3). For example, they existed at amino acid residues 260 to 263 (PVSA), 280 to 286 (DKSLXEQ), 432 to 435 (KLGV), and 484 to 487 (SVLP) in Fig. 3. The hydrophobicity plot (Fig. 4) showed some similar profiles between the predicted proteins encoded by NLRV S8 and RBSDV S10. The most hydrophobic region existed in the central part and the most hydrophilic peak was located in the terminal part of the sequences. Moreover, the size of the predicted protein encoded by NLRV S8, 62·4K, is very similar to that of RBSDV S10, 63·2K.

The predicted protein encoded by NLRV S10 reported previously (Noda *et al.*, 1994) showed homology with protein 1 encoded by MRDV S6. The conserved terminal sequences of the plus strand of NLRV are also similar to those of members of the *Fijivirus* genus. Fijiviruses have six uridine residues in the 5′-terminal region of the plus strand, and S3, S4, S8 and S9 segments of NLRV have

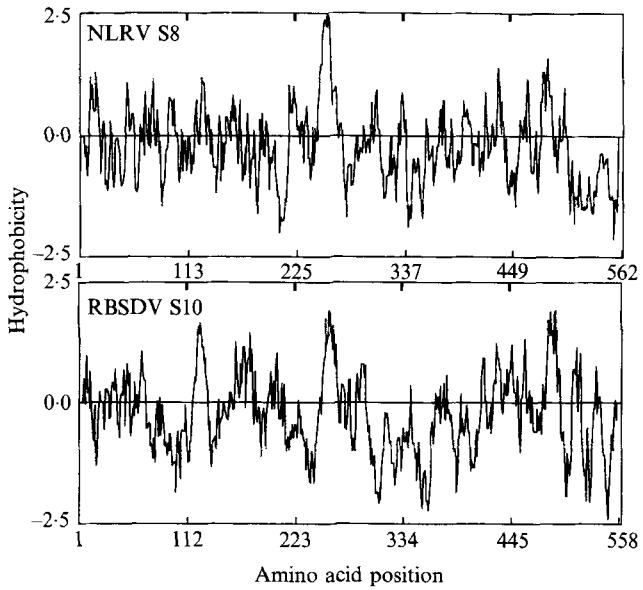

Fig. 4. Hydrophobicity plots of the predicted proteins encoded by NLRV S8 and RBSDV S10 (Uyeda *et al.*, 1990). They were generated by the software GENETYX using the program of Kyte & Doolittle (1982) with a sliding window of 15 residues.

three uridine residues; also some other segments are rich in uridine in this region. These facts support the hypothesis that the taxonomic position of NLRV is close to the *Fijivirus* genus in the family *Reoviridae*.

Most of the reoviruses associated with planthoppers are phytopathogenic, and show various symptoms to their host plants. However, NLRV does not cause any symptoms in rice plants. To reveal the difference between the phytopathogenic reoviruses and NLRV, similarity analyses of the nucleotide and amino acid sequences among the corresponding segments would be required.

We would like to thank Ms M. Koizumi for database access to DDBJ. This work was supported by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan, Bio-Media Program, BMP 93-V-(1)-4.

References

AZUHATA, F., UYEDA, I. & SHIKATA, E. (1992). Conserved terminal nucleotide sequences in the genome of rice black streaked dwarf virus. *Journal of General Virology* **73**, 1593–1595.

AZUHATA, F., UYEDA, I., KIMURA, I. & SHIKATA, E. (1993). Close similarity between genome structures of rice black-streaked dwarf and maize rough dwarf viruses. *Journal of General Virology* **74**, 1227–1232.

HOLMES, I. H., BOCCARDO, G., ESTES, M. K., FURUICHI, M. K., HOSHINO, Y., JOKLIK, W. K., MCCRAE, M., MERTENS, P. P. C., MILNE, R. G., SHIKATA, E. & WINTON, J. R. (1994). Family *Reoviridae*. In *Classification and Nomenclature of Viruses. Archives of Virology* supplement (in press).

INNIS, M. A. & GELFAND, D. H. (1990). Optimization of PCRs. In *PCR Protocols, a Guide to Methods and Applications*, pp. 3–12. Edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White. San Diego: Academic Press.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

KOZAK, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* **15**, 8125–8148.

KUDO, H., UYEDA, I. & SHIKATA, E. (1991). Viruses in the Phytoreovirus genus of the *Reoviridiae* family have the same conserved terminal sequences. *Journal of General Virology* **72**, 2857–2866.

KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.

MARZACHI, C., BOCCARDO, G. & NUSS, D. L. (1991). Cloning of the maize rough dwarf virus genome: molecular confirmation of the plant-reovirus classification scheme and identification of two large nonoverlapping coding domains within a single genomic segment. *Virology* **180**, 518–526.

NAULT, L. R. & AMMAR, E. D. (1989). Leafhopper and planthopper transmission of plant viruses. *Annual Review of Entomology* **34**, 503–529.

NODA, H., ISHIKAWA, K., HIBINO, H. & OMURA, T. (1991*a*). A reovirus in the brown planthopper, *Nilaparvata lugens*. *Journal of General Virology* **72**, 2425–2430.

NODA, H., ISHIKAWA, K., HIBINO, H., KATO, H. & OMURA, T. (1991*b*). Nucleotide sequences of genome segments S8, encoding a capsid protein, and S10, encoding a 36K protein, of rice gall dwarf virus. *Journal of General Virology* **72**, 2837–2842.

NODA, H., NAKASHIMA, N. & OMURA, T. (1994). Cloning of the *Nilaparvata lugens* reovirus genome: conserved terminal nucleotide sequences and nucleotide sequence of genome segment S10. *Journal of General Virology* **75**, 221–225.

OMURA, T., ISHIKAWA, K., HIRANO, H., UGAKI, M., MINOBE, Y., TSUCHIZAKI, T. & KATO, H. (1989). The outer capsid protein of rice dwarf virus is encoded by genome segment S8. *Journal of General Virology* **70**, 2759–2764.

PEARSON, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* **183**, 63–98.

SUZUKI, N., HARADA, M. & KUSANO, T. (1991). Molecular analysis of rice dwarf phytoreovirus segment S11 corresponding to wound tumour phytoreovirus segment S12. *Journal of General Virology* **72**, 2233–2237.

UYEDA, I., AZUHATA, F. & SHIKATA, E. (1990). Nucleotide sequence of rice black-streaked dwarf virus genome segment 10. *Proceedings of Japan Academy* Series B **66**, 37–40.

XU, Z., ANZOLA, J. V., NALIN, C. M. & NUSS, D. L. (1989). The 3′-terminal sequence of a wound tumor virus transcript can influence conformational and functional properties associated with the 5′-terminus. *Virology* **170**, 511–522.

(Received 13 December 1993; Accepted 25 May 1994)