

**The  $M_r$  43K major capsid protein of rice ragged stunt oryzavirus is a post-translationally processed product of a  $M_r$  67 348 polypeptide encoded by genome segment 8\***

N. M. Upadhyaya<sup>1</sup>, E. Zinkowsky<sup>1</sup>, Z. Li<sup>1</sup>, W. Kositratana<sup>2</sup>, and P. M. Waterhouse<sup>1</sup>

<sup>1</sup>CSIRO Division of Plant Industry, Canberra, Australia

<sup>2</sup>Department of Plant Pathology, Kasetsart University, Nakhon Pathom, Thailand

Accepted April 19, 1996

**Summary.** The nucleotide sequence of DNA complementary to rice ragged stunt oryzavirus (RRSV) genome segment 8 (S8) of an isolate from Thailand was determined. RRSV S8 is 1 914 bp in size and contains a single large open reading frame (ORF) spanning nucleotides 23 to 1 810 which is capable of encoding a protein of  $M_r$  67 348. The N-terminal amino acid sequence of a ~43K virion polypeptide matched to that inferred for an internal region of the S8 coding sequence. These data suggest that the 43K protein is encoded by S8 and is derived by a proteolytic cleavage. Predicted polypeptide sizes from this possible cleavage of S8 protein are 26K and 42K. Polyclonal antibodies raised against a maltose binding protein (MBP)-S8 fusion polypeptide (expressed in *Escherichia coli*) recognised four RRSV particle associated polypeptides of  $M_r$  67K, 46K, 43K and 26K and all except the 26K polypeptide were also highly immunoreactive to polyclonal antibodies raised against purified RRSV particles. Cleavage of the MBP-S8 fusion polypeptide with protease Factor X produced the expected 40K MBP and two polypeptides of apparent  $M_r$  46K and 26K. Antibodies to purified RRSV particles reacted strongly with the intact fusion protein and the 46K cleavage product but weakly to the 26K product. Furthermore, in vitro transcription and translation of the S8 coding region revealed a post-translational self cleavage of the 67K polypeptide to 46K and 26K products. These data indicate that S8 encodes a structural polypeptide, the majority of which is auto-catalytically cleaved to 26K and 46K proteins. The data also suggest that the 26K protein is the self cleaving protease and that the 46K product is further processed or undergoes stable conformational changes to a ~43K major capsid protein.

\*The nucleotide sequence data reported in this paper has been submitted to the Genome Sequence Database (GSDB) and has been assigned the accession number L46682.

## Introduction

Plant reoviruses are presently grouped into three genera namely, *Phytoreovirus*, *Fijivirus* and *Oryzavirus* under the family *Reoviridae* [2, 9, 18]. This classification is based on particle morphology, dsRNA profile, vector specificity and genus-specific conserved terminal nucleotide sequences. Rice ragged stunt virus (RRSV) belongs to the genus *Oryzavirus* [9]. RRSV has icosahedral particles (~65 nm in diameter) each containing 10 dsRNA segments [19]. All 10 dsRNA segments of RRSV have conserved terminal nucleotide sequences 5' GAUAAA--- and ---GUGC 3' [33] which are different from those of other plant reoviruses [9]. The nucleotide sequences of RRSV genome segments 5 and 9 have been determined and show no homology with analogous sequences of other plant reoviruses [13, 27, 29].

SDS-PAGE analysis of purified RRSV particles and *in vitro* translation products (derived from denatured dsRNA) before and after immunoprecipitation (with antibodies raised against purified RRSV) reveal that there are at least 13 RRSV encoded polypeptides [14, 15]. These studies indicate that there are five major and highly immunoreactive structural polypeptides ( $M_r$  33K, 39K, 43K, 70K and 120K), five minor and less immunoreactive structural polypeptides ( $M_r$  49K, 60K, 76K, 90K and 94K), and three non-structural polypeptides ( $M_r$  31K, 63K and 88K). As RRSV has only 10 dsRNA segments, the above observation means that some of the segments must encode more than one polypeptide or that some of the viral proteins are post-translationally-processed products from precursor polypeptides. Furthermore, each of the largest four genome segments possibly have the capacity to encode a polypeptide of >120K and yet only one product of >120K is visualized. Use of segment specific cDNA derived polypeptides, and antibodies raised against them, would greatly facilitate precise coding assignments. Using this approach we and others have shown that RRSV S9 encodes a 38K major structural protein [27, 29] and that RRSV S5 encodes a 90K minor structural protein [13].

In this paper we show that RRSV genome segment 8 encodes a major structural polypeptide, the majority of which is auto-catalytically cleaved into 26K and 46K proteins. Our data suggest that the 26K protein is a self-cleaving protease and that the 46K protein is further processed to a 43K major structural protein.

## Materials and methods

### *Plants, viruses and dsRNA*

RRSV infected rice plants (cultivar TN1) were from Thailand and were obtained by infesting healthy seedlings with viruliferous brown planthoppers according to Hibino and Kumura [8]. Virus purification was carried out according to Omura and others [19] and monitored by electron microscopy [7]. RRSV dsRNA was extracted and purified from infected rice plants according to Uyeda and others [28] with some modifications [27].

*Bacterial strains and plasmids, enzymes and radioisotopes*

*E. coli* strains JPA101 and JM109 were used to maintain and amplify cloning vectors pUC119 [30], Bluescript II SK<sup>+</sup> (Stratagene Cloning Systems, California, USA) and fusion protein expression vector pMAL-c2 [16]. Unless otherwise stated, all enzymes and radioisotopes used in cDNA cloning and sequencing were purchased either from New England Biolabs (Beverly, MA, USA), Promega Corporation (Madison WI, USA), Applied Biosystems Inc. (USA) or Amersham International Plc (Amersham, UK) and used according to manufacturer's recommendations.

*DNA cloning*

cDNA of RRSV was produced using M-MLV H<sup>-</sup> reverse transcriptase (Superscript, BRL Life Technologies, Inc., Gaithersburg, MD, USA) according to the manufacturer's protocol with minor modifications [27]. S7/S8 specific cDNA clones were selected by colony hybridization using partially hydrolyzed (50 mM KOH for 8 min at 65 °C followed by neutralization and precipitation) 3' end-labelled S7 + S8 dsRNAs as probes [11]. Segment specificities of the cDNA clones were determined by hybridization of radioactively-labelled representative clones to a blot of gel purified [12] RRSV segments. Segment separation was achieved by PAGE (7.5%) with 30 mA constant current in 2 × TBE for ~70h at 4 °C.

*Sequencing and analyses*

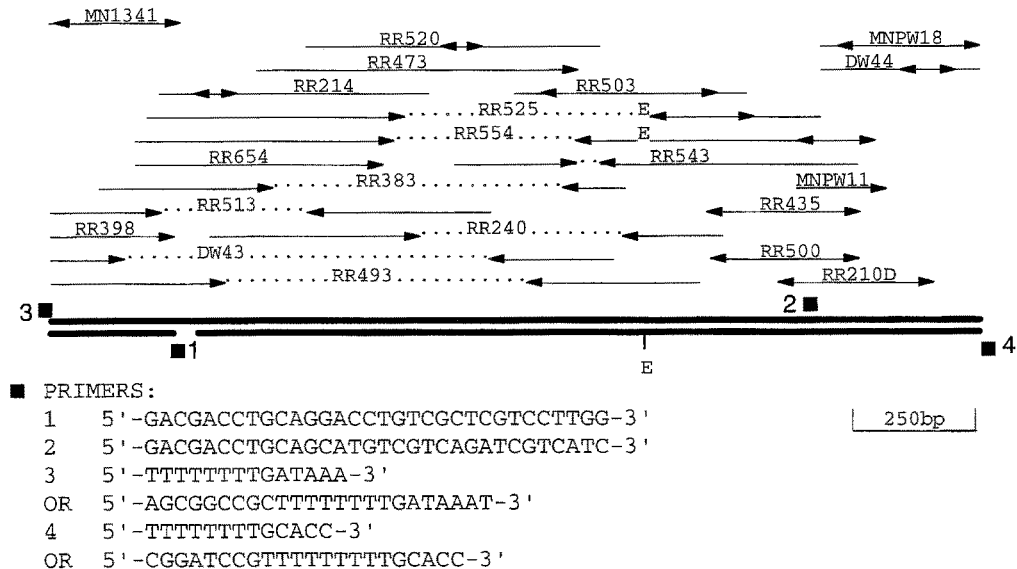
S8 specific double stranded cDNA clones were sequenced by the dideoxynucleotide chain termination method [22] using either <sup>32</sup>P-labelled (Life Technologies, Inc., BRL) primers or dye-labelled primers or terminators (Applied Biosystems). Terminal (5' and 3') sequences were obtained by synthesizing end specific cDNA using the RACE method [6] on heat denatured dsRNA. Internal primers were used to prime cDNA synthesis with reverse transcriptase (Superscript) to generate cDNA which were then polyadenylated using terminal transferase (Boehringer Mannheim). The RNA was removed using ribonuclease H and the cDNA amplified by polymerase chain reaction (PCR) using an internal primer and a primer consisting of the conserved RRSV terminal sequences, a polyT region and an appropriate restriction site (see Fig. 1 for details). Both strands were sequenced entirely using appropriate regions of cDNA clones, subclones or RACE clones as templates. Sequence assembly and analyses were carried out using the programs of the University of Wisconsin Genetics Computer Group (GCG) [5].

*Viral polypeptide isolation and amino acid sequencing*

Concentrated virus particle preparations were denatured and electrophoresed in SDS-polyacrylamide gels according to Hagiwara and others [7]. The viral polypeptides were then electro-blotted on to Selex 20 membrane (Schleicher and Schuell Inc., USA) using a trans-blot electrophoretic transfer cell (Bio-Rad Laboratories, Richmond, CA). The required band was excised after staining the membrane lightly with Coomassie Brilliant Blue, then micro-sequenced using a Model 477A amino acid sequencer (Applied Biosystems).

*Fusion proteins, antibodies and Western blot analysis*

The complete ORF of S8 was inserted into the pMAL-c2 vector, and the fusion protein produced and purified as described previously [16]. Where necessary, fusion protein was also purified from polyacrylamide gel (SDS-PAGE) by excising the protein band and eluting the

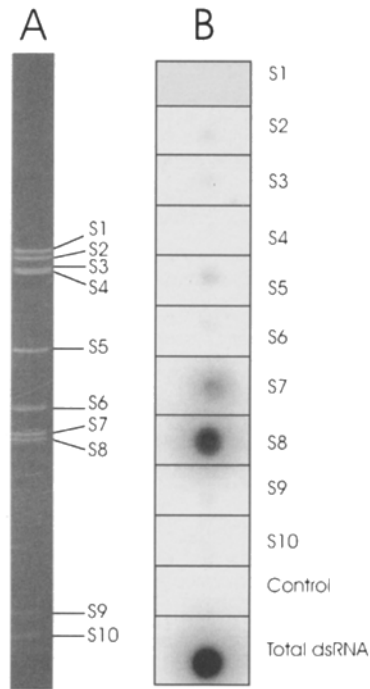


**Fig. 1.** cDNA cloning and sequencing strategy for RRSV genome S8. 17 cDNA-derived clones selected from the RRSV cDNA library were sequenced from both directions using M13 universal forward and reverse primers (indicated by arrows). *EcoRI* subclones (E) from RR525 and RR554 were also used. Internal primers (with restriction site overhangs) complementary to (-)ve strand sequence nt 259 to 240 (primer 1) and (+)ve strand sequence nt 1546 to 1565 (primer 2) were used in first strand synthesis by reverse transcription. After RNase H digestion and A-tailing of the first strand, PCR was performed with the respective internal primer (primer 1 or 2) and poly-T primers with *NotI* (primer) or *BamHI* overhang (primer 4) and/or end specific sequence anchors, respectively. Following cloning into plasmid Bluescript SK<sup>+</sup>, one 5' end clone (MN 134) and three 3' end clones (MNPW11, MNPW18 and DW44) were used to obtain terminal sequences. Both the strands were sequenced to near completion (indicated by bold lines)

protein with 50 mM ammonium carbonate and 0.1% SDS. The immunization procedure to raise antibodies in rabbits against the purified fusion proteins was essentially as described by Waterhouse and Murant [31]. Antibodies were purified and conjugated with alkaline phosphatase as described by Clark and Adams [4] and used in Western blot analysis. A 1:2000 dilution of alkaline phosphatase-antibody conjugate was used. The chromogenic substrates para-nitroblue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) [24] were used in Western blot analysis. The polyclonal antiserum raised against purified virus particles was a gift from Dr. Koganezawa, International Rice Research Institute, Manila, Philippines.

#### *In vitro* transcription and translation of S8 ORF

The coding region sequence of RRSV S8 from the plasmid pMN101N-4 was subcloned to Bluescript II SK<sup>+</sup> as a *BamHI* fragment such that its transcription was controlled by the T3 promoter. Transcription and translation were carried out using TNT coupled wheat germ extract system (Promega Corporation) with <sup>35</sup>S-methionine labelling, according to manufac-



**Fig. 2.** Confirmation of S8 specificity of cDNA sequences. RRSV genome segments were separated by 7.5% PAGE in  $2 \times$  TBE for 64 h at  $4^\circ\text{C}$  and stained with ethidium bromide (A), eluted from crushed gel, heat denatured and dot blotted onto nylon membrane, (B) then hybridized with  $^{32}\text{P}$ -labelled S8 cDNA fragment (nt 8–1914) from plasmid pMN101N-4 and autoradiographed using PhosphorImager (Molecular Dynamics, CA). In A the genomic segments of RRSV are labelled S1–S10 and B shows separated segments dot-blotted onto nylon membrane grid and labelled accordingly along with total dsRNA and no RNA dot-blot control

turer's instruction. Products were analysed by SDS-PAGE (10% separating and 6% stacking) and autoradiographed using a PhosphorImager (Molecular Dynamics).

## Results

### *cDNA cloning and nucleotide sequencing of RRSV genome segment S8*

From a library of 1424 cDNA clones prepared from RRSV total genomic RNA, 28 hybridised with S7/S8 dsRNA. Sequencing of these clones using forward and reverse primers produced two large contiguous sequences of 1926 bp (11 overlapping clones) and 1811 bp (17 overlapping clones). The 1811 bp contiguous sequence (Fig. 1) was confirmed to be derived from S8 by dot-blot hybridization (Fig. 2). This sequence did not have the previously reported [33] 5' or 3' terminal sequences. However, sequencing of RACE clones revealed that 8 and 95 bp at the 5' and 3' ends, respectively, were not represented in the 1811 bp contiguous sequence. The terminal sequences of these RACE clones have the previously reported [33] terminal sequences (5'-GATAAAT- and -GUGC-3'). The complete S8 nucleotide sequence, together with the deduced amino acid sequence, is presented in Fig. 3.

### *Characteristics of the S8 sequence*

A single open reading frame, spanning nt 23 to nt 1810 and predicted to encode a 596 aa polypeptide with a deduced  $M_r$  67 348 is present in the S8 sequence (Fig. 3). Comparison of S8 nucleotide and deduced amino acid sequences with

M N T K G F A Q K I Q R T F H L I H Q S H Y T 23  
**GATAAAT**CTCCCAGGACCAAGCATGAATACCAAGGGCTTCGCGCAGAAGATCCAGAGAACCTTCCACTTGATTCCACCAGAGTCACATACAC 90  
 P P P N K I S T H V E I F S E L L T Y V D Y E W Q R A A I G 53  
 ACCTCCACCAAAACAAATATCTACCCATGTTGAAATTTTCTCAGAAGTGTAAACATACGTTGACTATGAATGGCAGAGGGCAGCGATCGG 180  
 A S R W V E L A A E E N E I H I P F N T K D E R Q V N F L I 83  
 CGCTTCACGATGGGTTGAGTTAGCTGCTGAGGAAAACGAAATACACATCCCTTTTAATACCAAGGACGAGCGACAGGTCAACTTCTTAAT 270  
 A I I H V A I S L Y S N S G V E P F A D C D L E K I L N G D 113  
 CGCCATTATCCATGTTGCTATTTTCGTTGATTTGGAAGTCCAGGAGTAGAGCCTTTCGCTGATTTGTGATCTAGAGAAAATTTTGAATGGCGA 360  
 Y E S S K F D W N V V F D I V L P N D V R L K E K T L K A A 143  
 CTATGAGAGTACTAAGTTCGATGGAAATGTGGTGTTCGACATGCTTTTGCCAAATGATGTAAGATTGAAAGAAAAGACGCTTAAAGCAGC 450  
 V V E E E R V E P R P K R R E I T L G G Q R P L Q T R D V R 173  
 GGTGGTTGAGGAGAAGCGTGTGAACCTCGTCCAAAACGTAGAGAGATAACGCTTGGAGGACAGAGGCTTTTCAGACTAGGGATGTGCG 540  
 S D S D L I V L S K L E A Y D D V V E R P R Q N N T D I I I 203  
 TTCGGATTCGATTTGATGTTTTCGAGCAAACCTGAGGCGTATGATGATGTTTGAAGACCAAGGACAGAAATAACACCGCATTATCAT 630  
 H R L G L A K K M N P A S N S I L P A K I S R I I T E O V F G 233  
 CCAATCGACTGGTTAGCCAAAGATGAATCCTGCGAGTAATCAATCCACCGCCAAAATCTCGCGAATCATCACTGAACAAGTATTTGG 720  
 A G V N V D Y I A R P T E G F I S L S T A T Y T V A S Y F E 263  
 AGCTGAGTGAACGTGATTAATGCGCAGCAACTGAAGGATTTATATCTCTGTCGACCGCCACCTACACGGTAGCTTCTTACTTTGA 810  
 R G I A N Y D I N A V L D N I I D K T D Q L N T T D T V V E 293  
 ACGGGCATCGCCAAATATGACATCAATGCGCTCTGGACAACATCAATGATAAGACCGATCAACTGAAATACACCGATACCGCTCGTTGA 900  
 L P S I P P E D S S I E V A T P S H E T F F D I N T M I Y I 323  
 GCTACCATCCATTCGCCAGAGGACTCGTCAATGAAAGTGGCGACGCGCTCGCACGAGACGTTCTTTGATATCAACACCATGATATACAT 990  
 I M C C G S I T N P M I Q R L N G I V T R Y N T T N Y V V S 353  
 CATAATGTGCTGTGGAAGCATCACTAATCCATGATCCAGCGCCTAAACGGTATCGTCACCAGATACAACACGACCAATACGTTGGTAAG 1080  
 Y P D T D D G R K K A L A E R A V I T V D G K Y Y K C Y N D 383  
 TTATCCTGATACTGACGATGGAAGGAAGAAGGCATTTGGCTGAGCGTGCCTTATCACTGTGGCGGTAAGTACTATAAGTGTACAACGA 1170  
 I K A D T D K R R I L N P A V I K E V M I S L R H Y C G S V 413  
 TATCAAGGCTGACACAGATAAGCGAAGAAATCTCAACCCAGTGTCAATCAAGAGGTTATGATTAGTCTGCGTCACATATTGTTGGTAGTGT 1260  
 I H Y R E R M E A T H I S Q V F C L L M G I C Y G G L D T K 443  
 CATTCAATACAGAGAGAGGATGGAAGCGACTCATATAGTCAAGTGTTTTGTCTTCTGATGGGAATATGCTATGGTGGCTTAGATACCAA 1350  
 K I R C R W F E W P A V S N V P V T S S Y A E F K G S N S G 473  
 GAAGATCAGGTGTCGATGTTGAGTGGCCCGCGTATCTAACGTTCCAGTCACTTCTAGTTACGCGAGATTTAAAGGCAGTAACAGTGG 1440  
 L P P Y Q C R K F T P R T V M T S F A H W N L W A W M L D L 503  
 TCTGCCACCTTACCAATTCGCGCAAATTCACACCAGAAAGGTTATGACATCTTTTCCGCATTTGAATTTATGGGCGTGGATGCTTGATTT 1530  
 A I D K R M S S D R H L S I M K I Y V L N A F S H L V L N S 533  
 AGCAATGACAAACGCATGTCGTCAGATCGTCACTTAGTATTATGAAGATTTACGTTTGAACGCTTTCAGTCAATTTAGTTCTTAATTC 1620  
 S L E N S A N I L G P F P S F S G Y T A F T S T S I V R D V 563  
 CTCATTTAGAAAATAGCGCCAACATTTTAGGTCCTTTCCATCGTTTTCGGGTACACCGCTTTTACATCTACATCTATAGTTTCGTGATGT 1710  
 S S C Q R L A P F L L R M F S L V D Y V A A A A S S E A S G 593  
 TTCCTCGTGTCAACGCTAGCTCCGTTCTACTCCGCATGTTTAGTTTGTGATTATGTCGCGGCAGCAGCGAGTTCAGAAGCCAGCGG 1800  
 S G V \* 596  
 TTCGGTGTATCAGAGAGAGGGTTGCTTTAGGTGAGACGGCTTAGCMTTGCCTCATGAACCTGGTCACTGAGTGGCTCTCAGGCTGGC 1890  
 ACTGGCAGAGATAGAATGGGTGC 1914

**Fig. 3.** Complete nucleotide sequence of RRSV S8 cDNA. The deduced amino acid sequence of the S8 reading frame is indicated above the nucleotide sequence. Genus specific terminal sequences (bold letters), amino acid sequence matched to N-terminal sequence of the 43K viral polypeptide (underlined), internal PCR primers used (arrowed underline indicating the direction of priming), and the predicted translation termination codon (\*) are indicated

those in the GenBank and EMBL databases revealed no significant homologies. A more detailed pairwise comparison of S8 nucleotide and deduced amino acid sequences using GAP algorithm [5] with available reovirus sequences, including those of the 12 genome segments of rice dwarf virus (RDV), also showed little sequence similarity (data not shown). Comparison of RRSV S8 nucleotide sequence individually with possible analogous segments including RDV S8 [21], rice gall dwarf virus (RGDV) S8 [17], rice black streaked-dwarf virus (RBSDV) S7 [1], wound tumor virus (WTV) S8 [32], bluetongue virus (BTV) S5 [20] and reovirus serotype 3 (REOS3) M2 [26] showed similarities no greater than 40% (Table 1).

#### *Function of S8-encoded protein*

In order to perform the coding assignments for the RRSV genome segments we attempted N-terminal amino acid sequencing of the major virion polypeptides. Many of these polypeptides appeared to have blocked N-termini and yielded no amino acid sequences. However, the amino terminal sequence of the 43K virion protein was determined as "RIITEQVFGA". The amino acid sequence deduced from the S8 nucleotide sequence aa 225 to aa 234 (Fig. 3) matched perfectly with this sequence, suggesting that the 43K protein is translated from S8. One possible explanation for this internal matching is that S8 polypeptide undergoes post-translational cleavage. Such a cleavage at aa position 225 would produce an N-terminal 26K protein and the C-terminal region encoded 42K protein.

In order to further investigate the S8 encoded protein, the entire coding region of RRSV S8 was constructed in plasmid pMN101N-4 from three overlapping clones (RR493, RR543 and MNPW18) using the unique restriction sites *ScaI* (nt 1149) and *XmnI* (nt 1591). The S8 coding region was subcloned as a *BamHI* fragment from pMN101N-4 into pMAL-c2 to give an in-frame fusion between the MBP ORF and that of S8. This construct produced a single fusion protein of expected  $M_r > 100K$  (MPB 40K + S8 67K) in bacteria (Fig. 4, lane a). Polyclonal antibodies (raised in rabbits) against this fusion protein detected four protein bands of  $\sim M_r$  67K, 46K, 43K and 26K in a Western blot of dissociated proteins from purified RRSV particles (Fig. 5, lane a). Of these four bands the 43K protein was the most abundant and immunoreactive. In a twin Western blot, three of these bands (67K, 46K and 43K) and three additional bands were recognised by the antiserum raised against purified RRSV particles (Fig. 5, lane b). All but the 46K band were also clearly visible in a Coomassie Blue stained SDS-PAGE gels (data not shown). Incubation of the MBP-S8 fusion protein with Factor X produced the 40K MBP and two polypeptides of  $\sim M_r$  46K and 26K (Fig. 4, lane b). Antibodies to RRSV particles recognised the intact fusion protein in the untreated sample and the 46K protein with Factor X treated sample (Fig. 4, lanes c and d). Taken together, these results demonstrate that the S8 encodes a 67K protein which undergoes post-translational cleavage to produce 46K and 26K polypeptides. In an in vitro transcription-translation

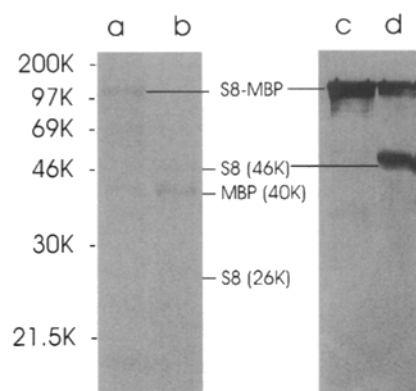
**Table 1.** Sequence comparisons between RRSV S8 and its predicted homologues in other reoviruses

	Nucleotide level (complete S8 sequence) % identity	Amino acid level (complete S8 sequence) % similarity	Amino acid level (S8 26K protein) % similarity	Amino acid level (S8 42K protein) % similarity	% identity
RDV S8	38.0	50	52.4	44.0	18.0
RGDV S8	39.0	47	46.0	44.3	18.0
RBSDV S7A <sup>a</sup>	39.0	51	47.9	46.0	22.5
RBSDV S7B <sup>a</sup>	39.3	55	44.4	53.0	23.0
WTV S8	39.2	50	48.3	46.0	19.6
BTV S5	38.7	46	48.4	45.9	21.4
REOS3 M2	39.5	46	47.0	51.0	24.3

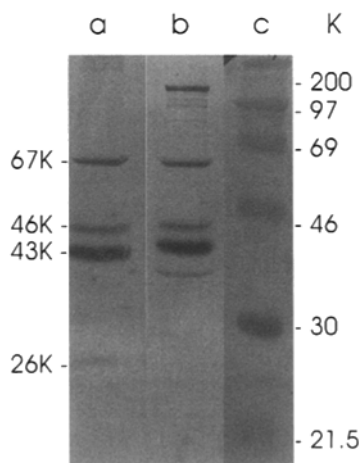
GCG program "GAP" [5] was used to calculate % identity of nucleotides (gap weight 5.00 and gap length weight 0.30) or % similarity and identity of inferred amino acids (gap weight 2.00, gap length weight 0.10)

<sup>a</sup>RBSDV S7 contains two non-overlapping ORFs

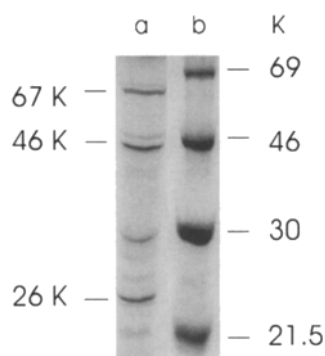




**Fig. 4.** Antigenicity of the MBP-S8 fusion protein and its cleavage (with Factor X) products. Electrophoretically-separated (12% SDS-PAGE and stained with Coomassie blue R-250) S8-MBP fusion protein uncleaved (*a*) and cleaved with Factor X (*b*), and the Western blot of electrophoretically-separated (12% SDS-PAGE) S8-MBP fusion protein uncleaved (*c*) and cleaved with Factor X (*d*), reacting with alkaline phosphatase conjugated immunoglobulins purified from antiserum raised against virus particles and detection by the use of BCIP and NBT as substrates



**Fig. 5.** Immunoblotting of RRSV proteins. Proteins from purified virus particles were separated by 10% SDS-PAGE and blotted onto nitrocellulose paper, and treated with alkaline phosphatase conjugated immunoglobulins (IgG) purified from either antiserum raised against MBP-S8 fusion protein (*a*) or purified RRSV particles (*b*) and assayed with BCIP and NBT substrates. *c* contains molecular weight markers. RRSV proteins were also blotted onto Selex-20 membrane and the band corresponding to the 43K protein excised and used for N-terminal amino acid sequencing



**Fig. 6.** In vitro transcription and translation from S8 coding region sequences. Autoradiograph (using PhosphorImager) of products of coupled transcription/translation from cloned (Bluescript SK<sup>+</sup>) S8 coding region sequences in the TNT Coupled Wheat Germ Extract System (using T3 RNA polymerase for transcription and <sup>35</sup>S-methionine for labelling in translation), separated by 10% SDS-PAGE and electro-blotted onto nylon (Hybond N) *a* and *b* <sup>3</sup>H-labelled protein molecular weight markers

experiment using the S8 coding region sequences cleavage of the 67K protein to 46K and 26 was also detected (Fig. 6).

### Discussion

We have determined the nucleotide sequence of RRSV genome segment 8 and identified a single large ORF predicted to encode a protein of  $M_r$  67 348. By S8 protein expression and immunological studies, virion polypeptide sequencing and sequence interpretation we have shown that S8 encodes a 67K precursor polypeptide from which polypeptides of 46K, 43K and 26K are produced and that the 43K protein is a major structural protein. We speculate that this 43K protein is a processed or conformationally changed derivative of the 46K protein and that the 26K protein is a self-cleaving protease.

Neither the predicted amino acid sequence nor the nucleotide sequence of RRSV S8 has any convincing similarity with the available sequences of other reoviruses. This is not unexpected as there appears to be little sequence similarity between reoviruses of different genera [9] and RRSV is the only member of the *Oryzavirus* genus for which there is sequence data available.

Our data demonstrate that S8 encodes a 67K protein which is cleaved to generate a 26K product from its N-terminal region and a 46K protein from its C-terminal region. Most probably this occurs by a self-cleavage mechanism, although no such mechanism has been previously reported in any animal or plant reoviruses. However, the dsRNA hypovirulence associated virus of chestnut blight fungus [23] possesses at least two autocatalytic proteases. Both these proteases (p29 and p48) are in the N-terminal regions of their respective polyproteins and have specific cleavage domains. By analogy the RRSV 26K product could be a self-cleaving protease, although we were unable to identify

any known catalytic domains, normally found in other self-cleaving proteases. It is interesting to note that the addition of the MBP to its amino terminal prevented the cleavage of the 67K protein. Lu and others [15] assigned a 70K virion polypeptide as the only product of RRSV S8 which is probably analogous to the 67K protein. However, they might have overlooked the processed forms (46K, 43K and 26K products) as degradation products.

We have demonstrated that a 43K major structural protein is encoded by S8. This protein is probably analogous to the outer capsid proteins encoded by RDV S8 (46K) [19] and RBSDV S7 (40K) [1], although neither of these proteins is produced by cleavage from a larger polypeptide. We suggest that RRSV 43K protein is either an alternative conformational form of the 46K polypeptide which migrates anomalously in gels or is a further processed form of the 46K polypeptide. This conformational shift or processing appears to only occur *in vivo* as the 43K form is not produced in *in vitro* cleavage experiments. This heterogeneity of a RRSV structural protein is analogous to the capsid protein heterogeneity reported in rotaviruses [3], animal reoviruses [10, 26] and phytoreoviruses [25]. In these cases, the mechanisms responsible are cleavage by host encoded proteases [10, 26], use of alternative initiation codons [3], or undetermined processes [25].

As discussed earlier, all four protein products of S8 (67K, 46K, 43K and 26K) were detected in virus particle preparations of RRSV by an antiserum raised against a S8-MBP fusion protein. The detection of the 67K protein suggests that it is assembled into the particle and once assembled, is unable to self-process. It is also possible that the small amount of the 26K protein detected in the virus particle preparation is not assembled into the particle in this form but is generated from its 67K progenitor in disassembling particles during the last stages of purification and storage.

We are currently investigating the nature of the putative 26K self-cleaving protease and its target sequence. It will be interesting to see whether further research shows that such autocatalytic cleavage is a general but undetected phenomenon among reoviruses or is a unique feature of RRSV.

### Acknowledgements

Financial support from the Rockefeller Foundation is gratefully acknowledged. We would like to thank Dr. Wayne Gerlach for his support. Thanks to Drs. David Dall, Brian Surin, Paul Keese, Rhonda Perriman and Michael Graham for critical reading of the manuscript. RRSV infected materials were imported into Australia under Australian quarantine permit (28546, BM 1243) and assistance of Mr. Lex Govaars, Australian quarantine officer, is gratefully acknowledged.

### References

1. Azuhata F, Uyeda I, Kimura I, Shikata E (1993) Close similarity genome structures of rice black-streaked dwarf and maize rough dwarf viruses. *J Gen Virol* 74: 1 227–1 232
2. Boccardo G, Milne RG (1984) Plant reovirus group. CMI/AAB Descriptions of Plant Viruses, No. 294

3. Chan WK, Penaranda ME, Crawford SE, Estes MK (1986) Two glycoproteins are produced from the rotavirus neutralization gene. *Virology* 151: 243–252
4. Clark MF, Adams AN (1977) Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J Gen Virol* 34: 475–483
5. Devereux J, Haerberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12: 387–396
6. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85: 8998–9002
7. Hagiwara K, Minobe Y, Nozu Y, Hibino H, Kimura I, Omura Y (1986) Component proteins and structure of rice ragged stunt virus. *J Gen Virol* 67: 1711–1715
8. Hibino H, Kimura I (1982) Detection of rice ragged stunt virus in insect vectors by enzyme-linked immunosorbent assay. *Phytopathology* 72: 656–659
9. Holmes IH, Boccardo G, Estes MK, Furuichi MK, Hoshino Y, Joklik WK, McCrae M, Mertens PPC, Milne RG, Shikata E, Winton JR (1994) Family *Reoviridae* In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) *Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses*. Springer, Wien New York, pp 208–239 (*Arch Virol [Suppl]* 10)
10. Jayasuria AK, Nibert ML, Fields BN (1988) Complete nucleotide sequence of the M2 gene segment of reovirus type 3 Dearing and analysis of its protein product ml. *Virology* 163: 591–602
11. Kuchino Y, Nishimura S, Smith RE, Furuichi Y (1982) Homologous terminal sequences in the double-stranded RNA genome segments of cytoplasmic polyhedrosis virus of the silkworm *Bombyx mori*. *J Virol* 44: 538–543
12. Kudo H, Uyeda I, Shikata E (1991) Viruses in the phyto-reovirus genus of the *Reoviridae* family have the same conserved terminal sequences. *J Gen Virol* 72: 2857–2866
13. Li Z, Upadhyaya NM, Gibbs A, Gerlach WL, Waterhouse PM (1996) Sequence and functional analysis of rice ragged stunt virus genome segment 5. *J Gen Virol* (in press)
14. Lu HH, Gong Z, Cao T (1988) Studies on the RNA polymerase activity associated with rice ragged stunt virus. *Sci Sinica Ser B* 31: 573–575
15. Lu H, Gong Z, Cao T (1990) Studies on the genomic coding assignments of rice ragged stunt virus. *Chin J Virol* 6: 167–195
16. Maina CV, Riggs PD, Grandea AG, Slatko BE, Moran LS, Tagliamonte JA, McReynolds LA, Guan C (1988) A vector to express and purify foreign proteins in *Escherichia coli* by fusion to, and separation from, maltose binding protein. *Gene* 74: 365–373
17. Noda H, Ishikawa K, Hibino H, Kato H, Omura T (1991) Nucleotide sequence of genome segments S8, encoding a capsid protein, and S10, encoding a 36K protein, of rice gall dwarf virus. *J Gen Virol* 72: 2837–2842
18. Nuss DL, Dall DJ (1990) Structural and functional properties of plant reovirus genomes. *Adv Virus Res* 38: 249–306
19. Omura T, Minobe Y, Kimura I, Hibino H, Tsuschizaki T, Saito Y (1983) Improved purification procedure for RNA segments of rice ragged stunt virus. *Ann Phytopathol Soc Japan* 49: 670–675
20. Oldfield S, Hirasawa T, Roy P (1991) Sequence conservation of the outer capsid protein, VP5, of blue tongue virus: a contrasting feature to the outer capsid protein VP2. *J Gen Virol* 72: 449–451

21. 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. *J Gen Virol* 70: 2759–2764
22. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467
23. Shapira R, Nuss DL (1991) Gene expression by a hypvirulence-associated virus of the chestnut blight fungus involves two papain-like protease activities. *J Biol Chem* 266: 19419–19425
24. Shirako Y, Ehara Y (1986) Rapid diagnosis of Chinese yam mosaic virus infection by electro-blot immuno-assay. *Ann Phytopathol Soc Japan* 52: 453–459
25. Suzuki N, Sugawara M (1991) Outer capsid protein heterogeneity of rice dwarf phyto-reovirus. *J Gen Virol* 72: 2239–2342
26. Tarlow O, McCorquodale JG, McCrae MA (1988) Molecular cloning of the gene (M2) of serotype 1 and 3 of mammalian reovirus. *Virology* 164: 141–146
27. Upadhyaya NM, Yang M, Kositratana W, Ghosh A, Waterhouse PM (1995) Molecular analysis of rice ragged stunt oryzavirus Segment S9 and sequence conservation among isolates from Thailand and India. *Arch Virol* 140: 1945–1956
28. Uyeda I, Lee S, Yoshimoto H, Shikata E (1990) Size of dsRNA segments of rice dwarf and rice ragged stunt viruses determined by electron microscopy. *J Electron Microscop (Tokyo)* 39: 54–58
29. Uyeda I, Suda N, Lee SY, Yan J, Hataya T, Kimura I, Shikata E (1993) Rice ragged stunt *Oryzavirus* genome segment 9 encodes a 38 600 M<sub>r</sub> structural protein. *J Gen Virol* 76: 975–978
30. Vieira J, Messing J (1987) Production of single-stranded plasmid DNA. *Methods Enzymol* 153: 3–11
31. Waterhouse PM, Murrant AF (1982) Purification of carrot red leaf virus and evidence from four serological tests for its relationship to luteoviruses. *Ann Appl Biol* 79: 101–204
32. Xu Z, Anzola JV, Nuss DL (1989) Assignment of wound tumor virus nonstructural polypeptides to cognate dsRNA genome segments by *in vitro* expression of tailored full-length cDNA clones. *Virology* 168: 73–78
33. Yan J, Kudo H, Uyeda I, Lee S, Shikata E (1992) Conserved terminal sequences of rice ragged stunt virus genomic RNA. *J Gen Virol* 73: 785–789

Authors' address: Dr. N. M. Upadhyaya, CSIRO Division of Plant Industry, GPO Box 1600 Canberra, ACT 2601, Australia.

Received January 8, 1996