MOLECULAR CLONING, SEQUENCING, FUNCTIONAL ANALYSIS AND EXPRESSION IN *E. COLI* OF MAJOR CORE PROTEIN GENE (S3) OF RICE DWARF VIRUS CHINESE ISOLATE

F. ZHANG[1], Y. LI*[2], Y. LIU[1], C. AN[1,2], Z. CHEN[1,2]

[1]National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, P.R. of China; [2]De Montfort University, The Gateway Leicester, LE1 9BH, UK

Received January 23, 1997; revised April 7, 1997

Summary. – The complete nucleotide sequence of major core protein gene (segment S3) of rice dwarf virus (RDV) Chinese isolate was determined after cDNA cloning from the viral genomic RNA. Sequence analysis showed that the cloned fragment is 3195 bp in length and containes a single open reading frame (ORF), encoding the major core protein (P3) which M_r of 114 K. The nucleotide and deduced amino acid sequences of S3 of this isolate share significant homology (94.1% and 97%, respectively) with those of S3 of the Japanese isolate. At the amino acid level, P3 of RDV Chinese isolate shares significant homology with P3 of rice gall dwarf virus (RGDV), significant regional homology with the rotavirus VP4 protein which forms spikes on the virus particles and has been identified as a protein involved in the activation of the rotavirus penetration, and homology with spheroidin of amsacta entomopoxvirus (SPH), which is the major protein of the occlusion body, with clp-like ATP-dependent protease binding subunit and with ATP-dependent protease ATP-binding subunit. Amino acid sequence analysis also showed that P3 contains RNA-dependent RNA polymerase (RDRP) motif-like elements such as DXXXD, SGXXXXXXN, GDD and ENXXXY. These results may suggest that P3 is a multifunctional protein which plays very important roles in the virus structure formation, virus replication and penetration processes. The full length cDNA sequence of RDV S3 and a partial one which covers nt 1004-3195 were cloned into bacterial expression vector pTrcHisB for expression. The full length cDNA sequence failed to be expressed in *E. coli*, but the partial sequence was successfully expressed there as confirmed by the Western blot analysis. Further analysis of RDV P3 is under way.

Key words: rice dwarf virus; major core protein; gene segment 3; RNA-dependent RNA polymerase

Introduction

RDV is a member of the genus *Phytoreovirus* of the family *Reoviridae* (Boccardo and Milne, 1984). The viral particle has no spikes on its surface (Li *et al.*, 1995). Its genome consists of 12 dsRNA segments designated S1-S12 according to their mobilities in polyacrylamide gel electrophoresis (PAGE). Each segment contains one single ORF except S12 which has three overlapping ORFs (Suzuki *et al.*, 1992a). The complete nucleotide sequences of the genome of the Japanese isolate (Uyeda, 1994) and of the Chinese isolate of RDV have been determined (Li *et al.*, 1995; Zhao *et al.*, 1996; Li *et al.*, 1994a; Liu *et al.*, 1994; Qu *et al.*, 1996; Qu *et al.*, 1995; Chu *et al.*, 1993; Xiao *et al.*, 1996, 1997; Li *et al.*, 1994b). The hexanucleotide 5'-GGC/UAAA-3' and tetranucleotide 5'-U/CGAU-3' are conserved in both strains.

The intact particles of RDV contain seven proteins, P1, P2, P3, P5, P7, P8 and P8', with respective sizes of 170,

*Corresponding author. Mailing address: College of Life Sciences, Peking University, Beijing 100871, P.R. of China

Abbreviations: AMV = avian myeloblastosis virus; nt = nucleotide; ORF = open reading frame; PAGE = polyacrylamide gel electrophoresis; PCR = polymerase chain reaction; RDV = rice dwarf virus; RGDV = rice gall dwarf virus; RDRP = RNA-dependent RNA polymerase; SPH = spheroidin; WTV = wound tumour virus

130, 110, 89, 58, 46 and 42 K. These proteins have been confirmed as the products of the genome segments S1, S2, S3, S5, S7 and S8 (Suzuki *et al.*, 1994) and localized in the virion core, surface, major core, core, minor core and outer coat (2x), respectively (Omura *et al.*, 1989). The S2-coded 130 K protein (P2) is essential for virus attachment and/or penetration into vector cells. P2 does not occur in virus preparations purified either with chloroform or carbon tetrachloride (Omura *et al.*, 1994). P8' is a postcleavage product of P8 which has the same N-terminus as P8 (Mao, unpublished results).

In this study, we designed three pairs of primers according to the sequence of RDV S3 (Suzuki *et al.*, 1990) and amplified the S3 cDNA of RDV Chinese isolate. Its full length sequence was determined and both the amino acid and nucleotide sequences were compared with those of RDV Japanese isolate. In addition, we compared the deduced amino acid sequence of RDV S3 with other known protein sequences through the Internet. The full length cDNA sequence and a partial one covering nt 1004-3195 were cloned into a bacterial expression vector, and the recombinant expression products and the total protein extracts from infected rice and from partial purified virus were subjected to Western blot analysis.

Materials and Methods

Virus purification and RNA extraction. RDV was isolated from infected rice originating from the Fujian Province of P.R. of China. Virus purification and RNA extraction were performed as described by Chu *et al.* (1993) and Li *et al.* (1994).

cDNA synthesis, polymerase chain reaction (PCR) and cloning. According to the sequence of S3 of RDV Japanese isolate, we designed and synthesized three pairs of primers (Table 1).

First strand cDNA was synthesized with avian myeloblastosis virus (AMV) reverse transcriptase according to the manufacturer's instructions (Promega) with 2 μg of RDV S3 RNA isolated from the viral genomic RNA. The PCR amplification was carried out as described by Li (1994b). One cycle corresponded to 40 secs at 94°C, 1 min at 53°C and 1.5 min at 72°C. A successful amplifications of the fragments of the expected size was confirmed by 0.8% agarose gel eletrophoresis.

Table 1. Primers used for PCR

Primer	nt position	Sequence
P3-1	1-23	5'-GGCAAAATCGAGCGAAGAATCCT-3'
P3-2	990-1015	5'-CTTGTTGCACTGCAGTAACCATTTG-3'
P3-3	934-955	5'-GCCCTGAAACAATTTAGCCTCCCG-3'
P3-4	2142-2165	5'-GACCGACACTATTTCTTCTATAAC-3'
P3-5	2085-2107	5'-TCAAATAAGCTTGGGAGGAATCC-3'
P3-6	3172-3195	5'-ATCAGGAACGAAACAAAACAATCC-3'

The PCR product was extracted with equal volume of phenol-chloroform (1:1), ethanol precipitated and treated with T4 DNA polymerase for blunt-end ligation. It was then inserted into *Eco*RV-linearized pBluescript plasmid yielding recombinant plasmid pBSRS3 which contained the full length sequence of RDV S3. The recombinant plasmids were transformed into *E. coli* strain DH5α.

DNA sequencing and sequence analysis. The RDV S3 cDNA inserts with different lengths were subcloned into pBluescript and their sequencing was carried out by the dideoxynucleotide chain termination method (Sambrook *et al.*, 1989). DNA sequence data were assembled and analyzed by DNASIS, and the amino acid sequence analysis was carried out by blaster through Internet.

Gene expression in E. coli and Western blot analysis. To express RDV S3 in *E. coli* strain DH5α, the RDV S3 coding region was subcloned into expression vector pTrcHisB. As tested by the Western blot analysis (Li *et al.*, 1996), the full length coding region was not expressed in *E. coli*. Then a partial sequence of S3 coding region covering nt 1004-3195 with a length of 2192 bp was cut out from pBSRS3 with *Pst*I and inserted into *Pst*I-linearized pTrcHisB. A total of 114 non-viral nucleotides including a 5'-ATG were added to the 5'-end of the S3 sequence. After transforming *E. coli* strain DH5α with the resulting recombinant plasmid pTrcBRS3P, the expression of the partial sequence of S3 was successful as confirmed by the Western blot analysis.

In addition, total proteins from healthy and RDV-infected rice leaves and partially purified RDV were also extracted as described by Zheng *et al.* (1997). In the Western blot analysis, a mouse anti-RDV P3 monoclonal antibody (Suzuki *et al.*, 1994) and an anti-mouse IgG (H+L) alkaline phosphatase conjugate (Promega) were used as primary and secondary antibodies, respectively.

Results and Discussion

cDNA synthesis, PCR amplification and cloning

After cDNA synthesis and PCR amplification, three products of expected size (approximately 1.0, 1.1 and 1.2 kb) were obtained. They were cloned into the *Eco*RV site of pBluescript plasmid. After screening and restriction mapping, three positive clones were obtained (pBRS3-1, pBRS3-2 and pBRS3-3). The full length RDV S3 clone pBSRS3 was then constructed by combining these three clones with appropriate restriction enzymes.

Nucleotide and amino acid sequence of RDV gene segment 3

DNA sequencing of cloned fragments was carried out by subcloning the appropriate restriction fragments into pBluescript plasmid according to the restriction map. Each subclone was sequenced from both strands. The complete nucleotide sequence of RDV gene segment 3 was 3195 nt long, contained one ORF (nt 3057) and encoded a 114 K protein with 1019 amino acid residues (Fig. 1). Table 2

```
      GG CAA AAT CGA GCG AAG AAT CCT TTA TCC TCG TGA GCT ATG GAC      44
                                                     M   D        2
AGT ACC GGC CGA GCA TAT GAC GGT GCG TCT GAA TTC AAA AGT GTC CTT      92
 S   T   G   R   A   Y   D   G   A   S   E   F   K   S   V   L      18
GTT ACC GAG GGT ACT TCA CAC TAC ACA CCA GTG GAA GTA TAT AAC ATC     140
 V   T   E   G   T   S   H   Y   T   P   V   E   V   Y   N   I      34
CTC GAT GAA TTG AAA ACC ATT AAA ATA ACG TCT ACC ATA GCT GAG CAA     188
 L   D   E   L   K   T   I   K   I   T   S   T   I   A   E   Q      50
TCA GTG GTG TCA CGT ACT CCA ATT CCT CTT TCG AAA ATT GGT TTA TCA     236
 S   V   V   S   R   T   P   I   P   L   S   K   I   G   L   S      66
GAC GTT AAG AAA TTG TTT GAC ATT AAT GTC ATA AAA TGT GGC TCA TCT     284
 D   V   K   K   L   F   D   I   N   V   I   K   C   G   S   S      82
TTG CGG ATA GTG GAC GAA CCT CAA GTT ACG TTC ATC GTA TCG TAT GCA     332
 L   R   I   V   D   E   P   Q   V   T   F   I   V   S   Y   A      98
AAG GAT ATC TAT GAT AAG TTC ATG TGC ATC GAA CAT GAC TCT GCC TAT     380
 K   D   I   Y   D   K   F   M   C   I   E   H   D   S   A   Y     114
GAA CCG AGT CTG ACG ATG CAT AGA GTG CGG GTT ATT TAC TCG ATG CTC     428
 E   P   S   L   T   M   H   R   V   R   V   I   Y   S   M   L     130
AAT GAC TAT TGT GCA AAA ATG ATT TCG GAG GTA CCC TAC GAA AGT AGC     476
 N   D   Y   C   A   K   M   I   S   E   V   P   Y   E   S   S     146
TTT GTA GGG GAG CTC CCT GTC AAA TCA GTA ACT CTA AAT AAA CTG GGT     524
 F   V   G   E   L   P   V   K   S   V   T   L   N   K   L   G     162
GAC AGA AAT ATG GAT GCT TTA GCT GAG CAT CTC TTA TTT GAG CAG GAC     572
 D   R   N   M   D   A   L   A   E   H   L   L   F   E   Q   D     178
GTT GTG AAT GCT CAG CGT GAG AAC CGG ATT TTC TAT CAA AGA AAG TCC     620
 V   V   N   A   Q   R   E   N   R   I   F   Y   Q   R   K   S     194
GCG CCT GCT GTG CCA GTC ACA TTC GGA GAT GAC TTG GAG CCG GCT GTC     668
 A   P   A   V   P   V   T   F   G   D   D   L   E   P   A   V     210
CGT GAA AGA GCT AAC CTC TAC CAT CGA TAT TCG GTT CCG TAT CAC CAA     716
 R   E   R   A   N   L   Y   H   R   Y   S   V   P   Y   H   Q     226
ATC GAG CTG GCA TTG CAT GCA TTA GCA AAT GAT CTA CTT TCA ATA CAG     764
 I   E   L   A   L   H   A   L   A   N   D   L   L   S   I   Q     242
TAC TGT CAT CCT ACA GTA GTG TAT AAT TAT TTG AGT TCC AGA GCT CCA     812
 Y   C   H   P   T   V   V   Y   N   Y   L   S   S   R   A   P     258
AAC TTT CTA AGG TTG GAT GAT CAA GTA TCG CTG AAA CTT ACA TCC GCT     860
 N   F   L   R   L   D   D   Q   V   S   L   K   L   T   S   A     274
GGG ATA GGT ACT CTT ATG CCT AGA CCT GTG GTA CAG CTG CTT GAT TAT     908
 G   I   G   T   L   M   P   R   P   V   V   Q   L   L   D   Y     290
GAT CTA GTT TAT ATG TCA CCG CTA GCC CTG AAC AAT TTA GCC TCC CGA     956
 D   L   V   Y   M   S   P   L   A   L   N   N   L   A   S   R     306
TTG TTG AGA AAG ATC TCT CTA CAC TTA GTA ATG CAA ATG GTT ACC GCA    1004
 L   L   R   K   I   S   L   H   L   V   M   Q   M   V   T   A     322
GTG CAG CAA GAT CTG GGT GAA GTG GTG AGT GTG TCC TCC AAT GTT ACT    1052
 V   Q   Q   D   L   G   E   V   V   S   V   S   S   N   V   T     338
AAT CCT GCT AGC GCT TGC TTG GTA AGG ATG AAC GTT CAA GGG GTC CAA    1100
 N   P   A   S   A   C   L   V   R   M   N   V   Q   G   V   Q     354
ACC TTG GCT GTA TTC ATA GCC CAG TCT ATG TTA AAT CCA AAC ATA TCA    1148
 T   L   A   V   F   I   A   Q   S   M   L   N   P   N   I   S     370
TAT GGT ATG CTA TCT GGA TTG ACA CTT GAC TGT TTT TCG AAC TTT ATT    1196
 Y   G   M   L   S   G   L   T   L   D   C   F   S   N   F   I     386
TAC GGC GCA TGT CTG ATG TTG TTT CAA GCC CTC ATT CCC CCC AGC GCT    1244
 Y   G   A   C   L   M   L   F   Q   A   L   I   P   P   S   A     402
TTG ACG GCT AGA CAG AGA TTA GAC ATA AAT AAC CGT TTT GCA TAT TTT    1292
 L   T   A   R   Q   R   L   D   I   N   N   R   F   A   Y   F     418
TTA ATT AAA TGT CAT GCT ACA CAA GCA ACG ACA GCT AGC GTT GTG CCT    1340
 L   I   K   C   H   A   T   Q   A   T   T   A   S   V   V   P     434
AAT CAA GTT ATT TAC CCA GTG GAT GCA ATT GAT CAG TGG CAG TCT AAT    1388
 N   Q   V   I   Y   P   V   D   A   I   D   Q   W   Q   S   N     450
CGA AGA GAC GTT TTG GTT GCA ATT TAT AAT AAT TTG CTG CCA GGT GAG    1436
 R   R   D   V   L   V   A   I   Y   N   N   L   L   P   G   E     466
TTG GTA TTA AGT AAT TTG ATA CAG ACT TAC TTT AGA GGT AAT ACT GCA    1484
 L   V   L   S   N   L   I   Q   T   Y   F   R   G   N   T   A     482
CAG CAA GCG GCT GAA ATA TTG ATT CCT GCG GAC CAG ACT TCT TAT GGC    1532
 Q   Q   A   A   E   I   L   I   P   A   D   Q   T   S   Y   G     498
GCC AAT GAG ACT CGC GCT CTA TCG GCA CCA TAT TTG TTT GGA GCT CCA    1580
 A   N   E   T   R   A   L   S   A   P   Y   L   F   G   A   P     514
ATC AAT ATG CTC GCC CCA GAT GCT AGA TTG TCA ACT TAT AAG CGT GAT    1628
 I   N   M   L   A   P   D   A   R   L   S   T   Y   K   R   D     530
CTC GCT TTG CCT GAT CGC TCC CCG ATA CTA ATC ACG ACT GTT GAG GGG    1676
 L   A   L   P   D   R   S   P   I   L   I   T   T   V   E   G     546
```

Fig. 1

```
CAA AAT TCG ATC TCC ATC GAA ATC TTG AGG CAT AAG ACG GGC TTG ATA    1724
 Q   N   S   I   S   I   E   I   L   R   H   K   T   G   L   I     562
CGT GCT ATG TAT CTG AAC GGC TTC GTC ACG CAA CCT CCA GCG TGG ATT    1772
 R   A   M   Y   L   N   G   F   V   T   Q   P   P   A   W   I     578
CGT AAC GCG AAT TCG AAT ACT GCG CTG CTA TCA CGA TTT CTT GAC GTT    1820
 R   N   A   N   S   N   T   A   L   L   S   R   F   L   D   V     594
ACG CCG AAT TTA TTA GGT ATT TAC GAA GCT ATT TTA GCT AAC ACG TAT    1868
 T   P   N   L   L   G   I   Y   E   A   I   L   A   N   T   Y     610
GCA AAT GCG GTA AAC GTT TAT TGC GAT TCC GTC TAC CGG GCA GAT ATA    1916
 A   N   A   V   N   V   Y   C   D   S   V   Y   R   A   D   I     626
CCT ACT GAA TGG AAA TTG CAC CAA TCA GTG GAT CCT CAG GAT TTA TTG    1964
 P   T   E   W   K   L   H   Q   S   V   D   P   Q   D   L   L     642
TTT GGT GTG TTT GGT ATT GTT CCG CAA TAT CAA ATT TTG AAT GAA GCG    2012
 F   G   V   F   G   I   V   P   Q   Y   Q   I   L   N   E   A     658
GTT CCG GAT TTC TTC GCT GGG GGT GAA GAC ATC CTA ATA CTA CAG CTT    2060
 V   P   D   F   F   A   G   G   E   D   I   L   I   L   Q   L     674
ATT CGG GCT GTG TAT GAC ACG TTG TCA AAT AAG CTT GGG AGA AAT CCC    2108
 I   R   A   V   Y   D   T   L   S   N   K   L   G   R   N   P     690
GCT GAC ATA TTT CAT CTT GAC GAG GTC TTC AAA GTT ATA GAA GAG ATA    2156
 A   D   I   F   H   L   D   E   V   F   K   V   I   E   E   I     706
GTG TCG GTT TTA GTT CAA CAA AAG GTT GAC GCT AGA AAG TAC TTC ACT    2204
 V   S   V   L   V   Q   Q   K   V   D   A   R   K   Y   F   T     722
GAA AGT ATG AGA AGT GGC TCA TTC TCA AAA CCT AGA TGG GAT AAC TTT    2252
 E   S   M   R   S   G   S   F   S   K   P   R   W   D   N   F     738
CTA AGA CGT CCA GTC GCT CAA CGA CTA CCG AAC TTA GAC AGC GTC ATC    2300
 L   R   R   P   V   A   Q   R   L   P   N   L   D   S   V   I     754
ATG ACG CAG GCA GAT CAT GTG TAC AAC TAC ATG ACT CAG CTT ACC CAT    2348
 M   T   Q   A   D   H   V   Y   N   Y   M   T   Q   L   T   H     770
ATA ATA CCA ATT ACT GAT TGC TTT TAC ATA GTA AAG AAT TCG GGA TTC    2396
 I   I   P   I   T   D   C   F   Y   I   V   K   N   S   G   F     786
GTC GAT CGT GGC TCG ACT GGT CCT GTG ATG GCA TCT TCA TCA GTC TAT    2444
 V   D   R   G   S   T   G   P   V   M   A   S   S   S   V   Y     802
GAA AAC GTG CTT AAG GTC GTT CAT ACC ATA GCT GAT TTT GAA GCG GCT    2492
 E   N   V   L   K   V   V   H   T   I   A   D   F   E   A   A     818
AAT GCT TTA CGC CTA CAA AGG AGA AGT GTA GAC AAT ACG TCT TAC ACT    2540
 N   A   L   R   L   Q   R   R   S   V   D   N   T   S   Y   T     834
GAC TCT CTT TCT GAC ATG TTC AAT GGG TTA CGG TCT ATC AGT TCT AGC    2588
 D   S   L   S   D   M   F   N   G   L   R   S   I   S   S   S     850
GAA TTT GTT AGA TCC GTC AAT GGT CGC TCA GTG TTT ACT GAA GGA CGC    2636
 E   F   V   R   S   V   N   G   R   S   V   F   T   E   G   R     866
ATT GAT GCA ATC AAG GTT AAT ATG CGA GTA AAA TTC GAT TTG CAA TTC    2684
 I   D   A   I   K   V   N   M   R   V   K   F   D   L   Q   F     882
ATC ACT GAG GAG GGC GGC TAC TCA AAA CCT CCA AAT GTG AAA AAG CTT    2732
 I   T   E   E   G   G   Y   S   K   P   P   N   V   K   K   L     898
ATG TTT TCG GAC TTT CTG AGC TTC TTA GAT AGT CAT AAG AGC GAT TAC    2780
 M   F   S   D   F   L   S   F   L   D   S   H   K   S   D   Y     914
AGG CCA CCA TTG CTT ACT GTG CCG ATC ACC ATA GGA TTA AAT AAC CTT    2828
 R   P   P   L   L   T   V   P   I   T   I   G   L   N   N   L     930
GGT GAA ACT AAT TCG AAC ACA CTG CGT ATG CGT TCG GAG GCA ATT GAT    2876
 G   E   T   N   S   N   T   L   R   M   R   S   E   A   I   D     946
GAA TAC TTT TCA AGT TAC GTC GGT GCA CAA ATT TTG GTA CCG ATT AAC    2924
 E   Y   F   S   S   Y   V   G   A   Q   I   L   V   P   I   N     962
GTC GTA GAC ACT CGA GTC TAT ACT GAA TTC AGT GAG CTA CGA AAT TTC    2972
 V   V   D   T   R   V   Y   T   E   F   S   E   L   R   N   F     978
TTT ACT GGT GAT GTG GTC ATT ACA GAT GAC CCA TTT GAC GTC TGG GAT    3020
 F   T   G   D   V   V   I   T   D   D   P   F   D   V   W   D     994
GGC GTC AAG GCA ACC TAC ATC CCG ATT GGT GTG CAT GGA GTT CGC TTG    3068
 G   V   K   A   T   Y   I   P   I   G   V   H   G   V   R   L    1010
GAT CCT AAT GGA GAT CAG CCG CCT CTG TGA CGC CCC GGA TGG CAT AGC    3116
 D   P   N   G   D   Q   P   P   L   *                           1019

TCT GCG ACG TGC GAT GCA GCG CAA ACG AGG ACT AGG CTA CCG AGG GGG    3164
CTG AAG TGG ATT GTT TTG CTT GGT TCC TGA T                        3195
```

Fig. 1

Complete nucleotide and deduced amino acid sequences of RDV Chinese isolate gene segment S3

shows a comparison of the nucleotide and deduced amino acid sequences of 3 Japanese RDV isolates with those of the Chinese RDV isolate.

Comparison of amino acid sequence of RDV P3 with those of other proteins

It has been found that RDV P3 contains RDRP motif-like elements (Suzuki *et al.*, 1990) such as DRNMD (motif I), SGLTLCFSN (motif II), GDD (motif III) and ENRIFY (highly conserved motif IV), present also in the VP3s of rotavirus and bluetongue virus (BTV) localized in the virus core (Mitchell, 1990). The VP3 motifs I and III are identical with the RDRP motifs I and III, and the VP3 motifs II and IV are only homologous to the RDRP motifs II and IV. It would be of great interest to characterize the role of RDV P3 in the virus genome replication (synthesis of negative strand RNA) and transcription, because it was konwn that in the case of L-A virus (dsRNA virus) a structural protein of the virion might be responsible for the synthesis of the negative-strand RNA (Fujimura *et al.*, 1988).

We searched for proteins of significant homology to RDV P3 using blast from Internet and found the rice gall dwarf virus (RGDV, genus *Phytoreovirus*) S3-coded core capsid protein (homology over 65%, identity over 43% in full length) (Fig. 2) (Takahashi *et al.*, 1994), rotavirus (genus *Rotavirus*) VP4 (homology over 46%, identity over 30% within a length over 157 amino acids) (Fig. 3A) (Taniguchi *et al.*, 1994), SPH (homology over 47%, identity over 31% within a length over 95 amino acids) (Hall *et al.*, 1991) (Fig. 3B), the clp-like ATP-dependent protease binding subunit (homology over 60%, identity over 34% within a length over 72 amino acids) (Fig. 3C) and the ATP-dependent protease ATP-binding subunit (homology over 64%,

Fig. 2
Comparison of the deduced amino acid sequences of RDV P3 and RGDV P3
The top and bottom lines refer to RDV and RGDV, respectively. Identical (*) and similar (.) amino acids.

```
RDV P3    555 LRHKTGLIRAMYLNGFVTQPPA == 71 == GIVPQYQILNEAVPDFFAGGED 668
ROTA VP4  362 IRNKPGKVNYAYLNGGFAQVDA == 71 == GITVSYTVMKPSDPDFITGGEN 475
              .*.* * .   ****   *  *          **   * ..  . ***  ***.

RDV P3    672 LQLIRAVYDTLSNKLGRNPADIFHLDEVFKVIEEIVSVLVQQK 714
ROTA VP4  504 IQQVTSAVFTAITNLGELPGLFSNITKVFSKTKEALSKLKSRK 546
              .* . .   *    **  *    .. .**   .* .* *  .*

RDV P3    804 NVLKVVHTIADFEAANALRLQRRSVDNTSYTDSLSDMFNGLRSISSSEFVRSV 856
ROTA VP4  690 NVRKQLHINNSFDTPTYGQLVERILDDGQLLDILGKLNPNSVEELFSEFLHRI 742
              ** * .*    *.     .*  * .*.    * *  .         ***.  .

RDV P3    893 PNVKKLMFSDFLSFLDSHKSDY 914
ROTA VP4  728 PNSVEELFSEFLHRIQHQLREY 749
              **  . .**.**  .     .*

                                    (A)

RDV P3    468 VLSNLIQTYFRGNTAQQAAEILI  490
SPH         8 IFNNRFYIYKRMNTVQILVVILI   30
              . .*    * * ** *    ***

RDV P3    751 DSVIMTQADHVYNYMTQLTHIIPITDCFYIVKNSGFVDRGSTG  793
SPH        73 DLTVFDPNDNVFNVEEKWRCASTNNNIFYAVSTFGFLSTESTG  115
              *  .    *.*.*   .        . ** *   **.   ***

RDV P3    801 VYENVLKVVHTIADFEAANALRLQRRSVDNTS  832
SPH       130 LFSRIIKIVYDPCTVETSNDCRLLRLLMANTS  161
              ..  ..*.*.     * .*  ** *  . ***

                                    (B)

RDV P3     50 ELKTIKITSTIAEQSVVSRTPIPLSKIGLSDVKKL 71
BOVCLPAB  366 ESERVVATPSDVAAAVERMTGIPVSKMGASDIERL 400
              * . .  * .    .*   * **.**.* **...*

RDV P3    680 DTLSNKLGRNPADIFHLDEVFKVIEEIVSVLVQQKVDAR 718
BOVCLPAB  502 NTLTEKVRRHPYSIVLLDEIEKANPQVITLLLQVLDDGR 540
              .**. *. *.*  *  ***. *   .....*.*   * *

                                    (C)

RDV P3     50  QSVVSRTPIPLSKIGLSDVKKLFDIN 75
CLPL-LACLA444  QSVERLTGIPVSDMGANDIEHLKNLD 469
               ***   * **.* .* .*.. * ...

RDV P3    680 DTLSNKLGRNPADIFHLDEVFKVIEEIVSVLVQQKVDAR 718
CLPL-LACLA567 NTLTERVRRNPYSVILLDEIEKADPQVLTLLLQVMDDGR
              .**. .. ***  .  ***. *   .....*.*   * *
```

Fig. 3

Comparison of the deduced amino acid sequences of RDV P3 and of some other proteins

A: rotavirus VP4 (ROTA VP4); B: amsacta moorei entomopoxvirus spheroidin (SPH); C: clp-like ATP-dependent protease binding subunit; D: ATP-dependent protease ATP-binding subunit. Identical (*) and similar (.) amino acids.

identity over 33% within a length over 61 amino acids) (Fig. 3D) (Gottesman *et al.*, 1990).

Both RGDV P3 and RDV P3 have the same function, they are major core proteins (Suzuki *et al.*, 1990a; Takahashi *et al.*, 1994). The rotavirus VP4 forms the spikes on virus particles and has been identified as the protein linked to the trypsin-enhanced virus growth in tissue culture (Tangiguchi *et al.*, 1994). The trypsin activation of rotavirus growth is associated with cleavage of the viral haemagglutinin VP4 into 60K and 28K fragments, VP5* and VP8*, respectively. These two small proteins can activate rotavirus penetration into the cells (Kaarel *et al.*, 1988). Sazaki *et al.* (1985) have demonstrated that rotaviruses penetrate into the cells rapidly like trypsin-activated infectious viruses. Rotavirus VP4 has also other functions, e.g. in haemagglutination, neutralization and virulence (Venkataram *et al.*, 1990).

Whether phytoreovirus S3-encoded major core protein has amino acid residues exposed on the surface of the virus particles and involved in replication are open questions. In addition, a phytoreovirus not only infects plants, but also multiplies in its insect vectors. In the process of infecting insect cells, there must be a recognition mechanism between the virus and the insect cell. It has been found that RDV P2 is essential for the virus attachment and/or penetration of the virus into the vector cell (Omura *et al.*, 1994). But it is not known whether RDV P3 is involved in the processes of recognition and penetration. The amino acid sequences of the structural proteins P3, P5, P7 and P8 are very conserved in three phytoreoviruses (RDV, RGDV and wound tumour virus (WTV)) as compared with the non-structural proteins. RDV P5, P7 and P8 have a significant homology to WTV P5, P7 and P8 (over 46.0, 35.0% and 53.6% amino acid identity, respectively). RDV P3 and P8 have a significant homology also to RGDV P3 and P8 (over 35 and 50% amino acid identity, respectively) (Li, Y., unpublished results). The structural proteins P2, P3 and P5 of RDV

Table 2. Comparison of the nucelotide and deduced amino acid sequences of the gene segment 3 of three RDV Japanese isolates with those of the RDV Chinese isolate

Chinese RDV isolate	Nucleotide sequence homology	Amino acid sequence homology	Reference
RDVSEG3A	94.3%	97.0%	Suzuki *et al.* (1990)
RDVS3CP	93.8%	97.0%	Kano *et al.* (1990)
RDVSEG3	94.0%	97.0%	Yamada *et al.* (1990)

also have a significant regional homology to rotavirus VP2 (102 K), VP4 (87 K) and VP3 (98 K), respectively (Li, Y., unpublished results). The same phenomenon has also been found by Suzuki *et al.* (1992b) when they compared the RDRP sequences of members of the family *Reoviridae*. It seems that phytoreoviruses are evolutionarily related more to rotaviruses.

In addition, we found that RDV P3 has a significant homology to SPH, the clp-like ATP-dependent protease binding subunit and the ATP-dependent protease ATP-binding subunit (Hall and Moyer, 1991; Gottesmann *et al.*, 1990). SPH is the major protein of the occlusion body of amsacta moorei entomopoxvirus. It protects the virions during transmission from one insect to another (Hall and Moyer, 1991). The function of SPH is similar to that of the baculovirus polyhedrin protein and the SPH gene can be used as an insertion site of an invertebrate expression vector (Arif, 1995). The significance of the homology of RDV P3 with SPH, the clp-like ATP-dependent protease binding subunit and the ATP-dependent protease ATP-binding subunit is unclear.

Expression of RDV S3 in E. coli and immunodetection of P3 in E. coli and RDV-infected rice plants

The full length coding region of RDV S3 failed to be expressed in *E. coli* while its part covering nt 1004-3195 (2192 bp long) was successfully expressed. By Western blot analysis, a specific band of 70 K was detected in *E. coli* transformed with pTrcBRS3P, and a specific band of 110 K was detected in RDV-infected rice and purified RDV preparations (Fig. 4). The expression of the RDV segment 2 full length sequence in the pTrcHisA expression vector gave also a negative result (Lu, RF, personal communication) It seems that the complete RDV gene segments S2 and S3 are not suitable for expression in the *E. coli* system.

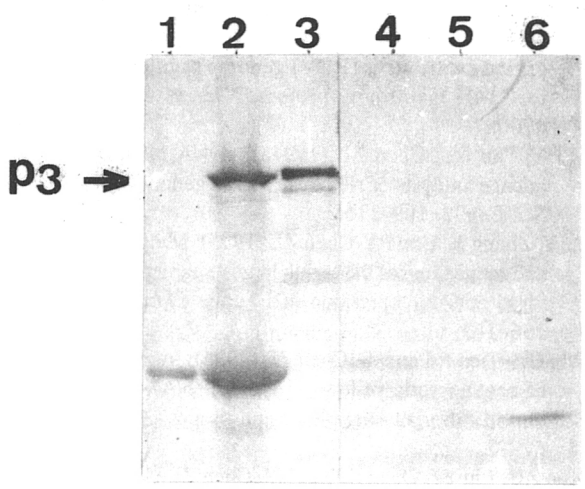

Fig. 4
Western blot analysis of RDV P3
Total protein extract from healthy rice leaves (1); total protein extract from RDV-infected rice leaves (2); partially purified RDV (3); lysate of bacterial cells transformed with pTrcHisB (control) (4); lysate of bacterial cells transformed with pBSRS3P but not induced (5); the same as lane 5, but cells induced for 4 hrs (6).

Acknowledgements. We would like to thank Dr. N. Suzuki for his generous gift of the antiserum against RDV P3, Dr. Y.X. Zhu for critical reviewing of the manuscript, Dr. S.Y. Ding for computer analysis and Ms X.H. Li for typing the manuscript. The work was supported by grant No. 863-101 of the National Science and Technology Committee, European Commission (contract No. ERB CI1*-CT94-0088), by Natural Sciences Foundation of China (contract No. 3940026) and by ICGEB (contract No. 93/095).

References

Arif BM (1995): Recent advances in the molecular biology of entomopoxviruses. *J. Gen. Virol.* **76**, 1–13.

Boccardo G, Milne RG (1984): Plant reovirus group. *CM/AAB Descriptions of Plant Viruses*, No. 294.

Chu RY, Zhang X, Pan NS, Chen ZL (1993): The cDNA cloning and nucleotide sequence of the gene encoding non-structural protein of rice dwarf virus genome segment 10. *Acta Bot. Sin.* **35**, 115–120.

Fujimura T, Wickner RB (1988): Replicase of L-A virus-like particles of saccharomyces cerevisine. *J. Biol. Chem.* **263**, 454–460.

Gottesman S, Squires C, Pichersky E, Carrington M, Hobbs M, Mattick JS, Dalrymple B, Kuramitsu H, Shirozu T, Fos-

terT, Clark WP, Ross B, Squires CL, Maurizi MR (1990): Conservation of the regulatory subunit for the clp ATP-dependent protease in prokaryotes and eukaryotes. *Proc. Natl. Acad. Sci. USA* **87**, 3513–3517.

Hall RL, Moyer RW (1991): Identification, cloning and sequencing of a fragment of Amsacta moorei entomopoxvirus DNA containing the spheroidin gene and three vaccinia virus-related open reading frames. *J. Virol.* **65**, 6516–6527.

Kaarel T, Kaljo T, Robert DS, Donald HR, Harry BG (1988): Infectious rotavirus enters cells by direct cell membrane penetration, not by endocytosis. *J. Virol.* **62**, 1136–1144.

Kano H, Koizumi M, Nodu H, Mizuno H, Tsukiharu T, Ishikawa K, Hibino H, Omura T (1990): Nucleotide sequence of rice dwarf virus (RDV) genome segment S3 coding for 114K major core protein. *Nucleic Acids Res.* **18**, 6700.

Li W, Li Y, Pan NS, Chen ZL (1994a): cDNA cloning and sequence analysis of rice dwarf virus genome S5. *J. Basic Sci. Eng.* **2**, 109–116.

Li W, Li Y, Zhang X, Chu RY, Chen ZL (1995): Molecular cloning and sequencing of the eighth largest segment of rice dwarf virus and its expression in E. coli. *Chin. J. Virol.* **11**, 156–162.

Li Y, Wei CH, Tien P, Pan NS, Chen ZL (1996): Immunodetection of beet necrotic yellow vein virus RNA3-encoded protein in different host plants and tissues. *Acta Virol.* **40**, 67–72.

Li Y, Xue ZH, Liu YF, Quan S, Pan NS, Chen ZL (1994b): cDNA synthesis, molecular cloning and sequence analysis of rice dwarf virus segment S12. *Chin. J. Virol.* **10**, 339–345.

Liu YF, Li Y, Pan NS, Chen ZL (1994): Molecular cloning and sequencing of the rice dwarf virus genome segment S6. *Chin. J. Virol.* **10**, 246–250.

Mitchell DB, Both GW (1990): Completion of the genomic sequence of the simian rotavirus SA11: Nucleotide sequence of segment S1, 2 and 3. *Virology* **177**, 324–331.

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 S8. *J. Gen. Virol.* **70**, 2759–2764.

Omura T, Takahashi A, Hibino H, Kimura I (1994): Rice dwarf virus 130K protein is essential in the adsorption or penetration processes of the virus on vector cells. *Ann. Phytopathol. Soc. Jpn.* **60**, 384.

Qu L, Li Y, Quan S, Ding SY, Suzuki N, Chen ZL (1996): Sequence analysis and expression in E. coli of segment 7 of rice dwarf virus genome. *Acta Microbiol. Sin.* **3615**, 335–343.

Qu L, Li Y, Zhu YX, Chen ZL (1995): cDNA cloning and sequence analysis of rice dwarf virus genome segment S9. *Chin. J. Virol.* **11**, 271–275.

Sambrook J, Fritsch EF, Maniatis T (1989): *Molecular Cloning: A Laboratory Manual.* 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Sazaki H, Kitaoku S, Konno T, Suto T, Ishida N (1985): Two modes of human rotavirus entry into MA104 cells. *Arch. Virol.* **85**, 25–34.

Suzuki N, Sugawara M, Kusano T, Mori H, Matsuura Y (1994): Immunodetection of rice dwarf phytoreovirus proteins in both insect and plant hosts. *Virology* **202**, 41–48.

Suzuki N, Watanabe Y, Kusano T, Kitagawa Y (1990): Sequence analysis of the rice dwarf phytoreovirus segment S3 transcript encoding for the major structural core protein of 114kDa. *Virology* **179**, 455–459.

Suzuki N, Sugawara M, Kusano T (1992): Rice dwarf phytoreovirus segment S12 transcript is tricistronic in vitro, *Virology* **191**, 992–995.

Suzuki N, Tanimura M, Watanabe Y, Kusano T, Kitagawa Y, Suda N, Kudo H, Uyeda I, Shikata E (1992b): Molecular analysis of rice dwarf phytoreovirus segment S1: Interviral homology of the putative RNA-dependent RNA polymerase between plant- and animal-infecting Reoviruses. *Virology* **190**, 240–247.

Takahashi Y, Tomiyama M, Hibino H, Omura T (1994): Conserved primary structures in core capsid proteins and reassembly of core particles and outer capsids between rice gall dwarf and rice dwarf phytoreoviruses. *J. Gen. Virol.* **75**, 269–275.

Taniguchi K, Urasawa T, Urasawa S (1994): Species specificity and interspecies relatedness in VP4 genotypes demonstrated by VP4 sequence analysis of Equine, Feline and Canine rotavirus strains. *Virology* **200**, 390–400.

Uyeda I, Suda N, Yamada N, Kudo H, Murao K, Suga H, Kimura I, Shikata E, Kitagawa Y, Kusno T, Sugawara M, Suzuki N (1994): Nucleotide sequence of rice dwarf phytoreovirus genome segment 2: completion of sequence analysis of rice dwarf virus. *Intervirology* **37**, 6–11.

Venkataram Prasad BV, Burns JW, Mariefta E, Estes MK, Chiu W (1990): Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy. *Nature* **343**, 476–479.

Xiao J, Li Y, Quan S, Gu HY, Chen ZL (1996): Nucleotide sequence of rice dwarf virus genome segment S11 and function analysis of S11 encoded protein. *Chin. J. Biotechnol.* **12**, 361–366.

Xiao J, Li Y, Zhang J, Liu JF, Chen ZL (1997): Nucleotide and protein sequence analysis of rice dwarf virus replicase (segment S1) *Acta Microbiol. Sin.* (in press).

Yamada N, Ugeda I, Kado H, Shikata E (1990): Nucleotide sequence of rice dwarf virus genome 3. *Nucleic Acids Res.* **18**, 6419.

Zhao XL, Li Y, Liang XW, Chen ZL (1996): Sequencing and function analysis of the deduced protein of rice dwarf virus genome segment S4. *Acta Microbiol. Sin.* **36**, 93–102.

Zheng HH, Li Y, Yu ZH, Li W, Chen MY, Chen ZL (1997): Recovery of transgenic rice plants expressing the rice dwarf virus outer coat protein (S8). *Theor. Appl. Genet.* **94**, 522–527.