Original Paper Plant Biotechnology, **16**(2), 103-113(1999)

Amino Acid Sequencing and cDNA Cloning of Rice Seed Storage Proteins, the 13kDa Prolamins, Extracted from Type I Protein Bodies

Norihiro MITSUKAWA*[1], Ryoichi KONISHI*, Kunitomo KIDZU*, Kozo OHTSUKI**, Takehiro MASUMURA* and Kunisuke TANAKA*[2]

*Laboratory of Genetic Engineering, Faculty of Agriculture
**Department of Food Science and Nutrition,
Kyoto Prefectural University,
Shimogamo, Kyoto 606-8522, Japan

Received 20 August 1998; accepted 23 October 1998

Abstract

The 13 kDa prolamins, one of the rice storage proteins, consist of complex mixtures of polypeptides encoded by multigene family that show heterogeneity both in size and solubility. Although many researchers have isolated prolamin cDNA clones, it has not been possible to correlate most of these cDNA clones with individual 13 kDa prolamin mature polypeptides. We isolated three new prolamin cDNA clones, λRM1, λRM4 and λRM9. Further more, we purified six 13 kDa prolamin polypeptides from rice type I protein bodies, and determined these amino acid sequences. Here we demonstrate a classification for the 13 kDa prolamin polypeptides which can be divided four classes, 13-I, 13-IIa, 13-IIb and 13-III. Cysteine labeling of the prolamin polypeptides indicated that 13-I contains cysteine residues, but 13-IIa or 13-IIb have no cysteine residues. The 13-I polypeptide was soluble in nonreducing solution when their cysteine residues form intramolecular disulfide bonds, but not soluble at intermolecular bonding.

Abbreviations

ABD-F, 4-(aminosulfonyl)-7-fluoro-2,1,3,-benzoxadiazole; CBB R-250, Coomassie Brilliant Blue R-250; EST, expressed sequence tag; FPLC, Fast Protein Liquid Chromatography; IEF, isoelectrofocusing electrophoresis; PB-I's, type I protein bodies; PB-II's, type II protein bodies; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TBP, tri-*n*-buthyl phosphine; TFA, trifluoroacetic acid

1. Introduction

Rice seed endosperm cells contain two types of storage proteins, prolamin and glutelin that are defined as alcohol-soluble and alkaline (and/or acid) soluble proteins, respectively. It has been shown that prolamin accumulates in the type I protein bodies (PB-I's), and glutelin in the type II protein bodies (PB-II's) [Tanaka *et al.* 1980]. The population of PB-I proteins are 18-35% of all endosperm protein [Ogawa *et al.*, 1987, Li and Okita 1993, Krishnan and White 1995], but almost these proteins are wasted from the viewpoint of human nutrition; cooked PB-I's pass through the human digestive tract as discrete particles [Tanaka *et al.*, 1975, Tanaka *et al.*, 1978, Resurreccion and Juriano 1982]. Although rice is identified as one of the carbohydrate sources for World's population, many Asian people depend on rice as a major protein source. In order to improve the availability of rice PB-I proteins, it is necessary to elucidate the correspondences of prolamin genes in PB-I's with their proteins.

Prolamins are encoded by the multigene family, and separate into three major polypeptide bands with apparent molecular sizes of 10 kDa, 13 kDa and 16 kDa by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [Ogawa *et al.*, 1987]. Micro heterogeneity exists within the 13 kDa and the 16 kDa prolamin classes. Ogawa *et al.*, [1989] reported five distinct 13 kDa prolamin polypeptides separated by two-dimentional electrophoresis. They also found that some 13 kDa prolamins, designated "13a", could be extracted in the complete absence of reducing agents, while the remaining 13 kDa prolamins "13b" required reducing conditions, as do the 10 kDa and 16 kDa prolamins. Hibino *et al.*,

[1] *Present address: Mitsui Plant Biotechnology Research Institute, TCI-D21, Sengen 2-1-6, Tsukuba, Ibaraki 305-0047, Japan*
[2] To whom correspondence should be addressed.
[3] The nucleotide sequences were submitted to DDBJ, EMBL and GenBank nucleotide sequence databases. The accession numbers of λRM1, λRM4 and λRM9 are AB016503, AB016504 and AB016505, respectively.

[1989] extracted prolamins with 60% 1-propanol and showed that isoelectrofocusing electrophoresis (IEF) separated the 13 kDa prolamins into seven polypeptides. Horikoshi *et al.*, [1991] reported resolution of five 13 kDa prolamins by IEF.

Amino acid sequencing and cDNA cloning of rice prolamins have been reported. A cDNA encoding the 10 kDa prolamin has been isolated by Masumura *et al.* [1989b] . The predicted NH_2-terminal sequence is identical to that of purified 10 kDa prolamin polypeptide. Horikoshi *et al.*, [1991] purified a "major prolamin" and reported its complete amino acid sequence, the molecular size of the "major prolamin" was estimated at 15,000 by SDS-PAGE [Horikoshi *et al.*, 1991] . It's sequence is highly homologous (85-99% sequence identity) to deduced amino acid sequences from isolated cDNA clones, pProl 7 [Kim and Okita 1988a] , pProl 14 [Kim and Okita 1988b] , pX24, pS23 [Shyur and Chen 1990] and λRM2 [Yamagata *et al.* 1992] . In addition, three cDNA clones, λRM7 [Masumura *et al.*, 1990] , pS18 [Shyur *et al.* 1992] and pProl 17 [Kim and Okita 1988b] have also been reported, but these amino acid sequences share lower homology (58-70% of amino acid identities) with that of the "major prolamin" purified by Horikoshi *et al.*, [1991] . Most of the prolamin cDNA clones have not been identified by corresponding sequences of purified prolamin polypeptides.

In order to resolve the characteristics of the prolamin polypeptides , it is necessary to identify the amino acid sequences of prolamin polypeptides corresponding to cDNA clones. We report here purification of six 13 kDa prolamins by reverse-phase column chromatography and determination of partial amino acid sequences for their tryptic peptides. We have also isolated three new prolamin cDNAs distinct from previously reported prolamin genes. By these approaches it should be possible to identify proteins corresponding to specific prolamin clones, and to classify the cDNA clones, together with the prolamin clones isolated previously, into four classes.

2. Materials and Methods

2. 1 Preparation of the rice PB-I fraction

The PB-I fraction was prepared by the method of Ogawa *et al.*, [1987] with some modifications. Matured rice seeds (100 g) (*Oryza sativa*, L. Japonica cv. Nipponbare) were homogenized in homogenization buffer (50 mM Tris-HCl, pH7.5, 50 mM potassium acetate, 5 mM magnesium acetate including 0.5 M sucrose) at 0°C using a polytron homogenizer (Kinematic). The homogenate was centrifuged at 5,000×g for 10 min. The precipitate was resuspended in 50 m*l* of the homogenization buffer supplemented with 0.2% (w/v) Cellulase "Onozuka RS", 0.2% (w/v) Maserozyme and 0.02% (w/v) Pectolyase, and incubated at 37°C for 18 h. The incubated mixture was centrifuged at 200×g for 10 min. The supernatant was recentrifuged at 5,000×g, and the 5,000×g precipitate was washed with the homogenization buffer. The fraction named "total PB fraction" was resuspended in 10 m*l* of the homogenization buffer and stored at 4°C before using. To prepare the PB-I fraction, the total PB fraction was resuspended in 0.2 M sodium acetate-HCl, pH1.7 containing 10 μg ml^{-1} pepsin (2,660 units mg^{-1}, Sigma). The mixture was incubated at 37°C for 2 h with gentle shaking. After incubation, the remaining particles were pelleted and washed with the homogenization buffer, then stored at 4°C until use.

2. 2 Fractionation of polypeptides from the rice PB-I fraction

Polypeptides were extracted from the PB-I fraction by shaking for 1 h at room temperature in N buffer (62.5 mM Tris-HCl, pH7.0 with 4 M urea and 2% (w/v) SDS). After shaking, suspensions were centrifuged at 15,000×g for 10 min. The supernatant and the pelleted materials were subjected to SDS-PAGE using a 16% (w/v) acrylamide gel, after resuspending in R buffer (62.5 mM Tris-HCl, pH7.0 with 4 M urea, 10 mM tri-*n*-buthyl phosphine (TBP) and 2% (w/v) SDS). Separated polypeptide bands were stained by Coomassie Brilliant Blue R-250 (CBB R-250).

2. 3 Reduction and fluorescent labeling of polypeptides

Fluorescent labeling of protein fractions by reaction with 4-(aminosulfonyl)-7-fluoro-2,1,3,-benzoxadiazole (ABD-F, Dojin) was performed according to the method previously described by Kirley [1989] . Reactions were performed in 50 mM Tris-HCl, pH7.0, 4 M urea, 2% (w/v) SDS, 15% (v/v) DMSO and 8 mM ABD-F at 56°C for 1 h at protein concentrations ranging from 0.2 to 2 mg ml^{-1}. After labeling, the reaction mixture was separated by SDS-PAGE. Gels containing fluorescently labeled polypeptides were placed under UV illumination (365 nm) and photographed using Polaroid film (No. 667). Afterwards the gels were stained with CBB R-250.

2. 4 Purification and amino acid sequencing of the 13 kDa prolamin polypeptides

The 13 kDa prolamin polypeptides were extracted from the rice PB-I fraction by shaking in N buffer at 55°C for 2 h. An equal volume of ethanol was added to the extract and the mixture was centrifuged at

15,000×g for 10 min. The 50% ethanol soluble prolamin fraction was evaporated *in vacuo* at 25°C until it's volume was reduced to half. The prolamin fraction (N fraction) was concentrated by ultrafiltration using Centriprep-10 (Amicon). After extraction of N fraction, another prolamin fraction (R fraction) was extracted from residual PB-I fraction by shaking in R buffer.

The N fraction was further fractionated on a Fast Protein Liquid Chromatography (FPLC) system (Pharmacia) using a Cosmosyl $_{10}C_{18}$-500 column (Nakalai Tesque) equilibrated with 0.1% trifluoroacetic acid (TFA). The fraction (500 μg of proteins) was applied to the column, then eluted with a 30% to 60% linear acetonitrile gradient in 0.1% TFA at a flow rate of 0.2 ml min^{-1} at room temperature. Elution was monitored by UV absorbance at 280 nm. Five eluted peaks were rechromatographed to enhance their purity. After labeling with ABD-F, the reaction mixtures were desalted by gel filtration using Bio-Gel P-6 (BIO-RAD) in 5 mm×100 mm disposable columns. The R fraction was also fractionated using nearly the same procedure after S-pyridylethylation by 4-vinylpyridine [Geisow and Aitkin 1989].

Each purified polypeptide (100 μg) was digested with 4 μg of trypsin (Promega) in 100 μl of 50 mM Tris-HCl and 4 M urea, pH8.2 under N_2 gas for 6 h at 37°C as described [Findlay and Geisow 1989]. Tryptic peptides were isolated by the Tricine SDS-PAGE system [Schägger and von Jagow 1987] using 10% polyacrylamide gels. The separated peptides were blotted to PVDF membranes (BIO-RAD). The peptides on the membranes were sequenced using a model 477A protein sequencer (Applied Biosystems).

Proteins quantities were estimated by the Bradford method [Bradford 1976] using bovine γ-globulin as the standard.

2.5 *Isolation of novel prolamin cDNAs from a developing rice seed cDNA library*

To obtain candidate clones for seed storage proteins, a λgt11 cDNA library prepared from developing rice seeds (22 days after flowering; DAF) [Masumura *et al.*, 1990] was screened with a single stranded cDNA probe. The plaques were transferred onto Hybond-N nylon membranes (Amersham). Total RNAs were prepared from developing seeds (14 DAF) by the SDS-phenol method [Yamagata *et al.* 1986]. Poly(A)$^+$ RNA was purified from the total RNA using a oligo(dT)-cellulose column [Yamagata and Tanaka 1986]. The poly(A)$^+$ RNA was reverse-transcribed in the presence of [α-^{32}P] dATP by RAV-2 reverse transcriptase (TAKARA) to generate a single stranded cDNA probe. The membranes were hybridized with the single stranded cDNA probe for 16 h at 37°C in 6×SSC solution containing 50% formamide, 0.1% SDS, 5×Denhardt's solution, and 200 μg ml^{-1} denatured salmon sperm DNA. The membranes were washed twice at 25°C for 45 min with 6×SSC solution containing 0.1% SDS.

To select novel prolamin cDNA clones, candidate clone's plaques were transferred to membranes and then hybridized with a mixture of previously isolated rice storage protein cDNA clones; λRP10 (10 kDa prolamin) [Masumura *et al.*, 1989b], λRG21 (glutelin) [Masumura *et al.*, 1989a] and λRG1 [Masumura *et al.*, unpublished] (identical to pREE61 [Takaiwa *et al.*, 1986]). The membranes were washed twice at 25°C for 45 min with 1×SSC solution containing 0.1% SDS. Insert fragments of nonhybridizing clones were subcloned into the *Eco*R I site of pBluescript KS$^+$ (Stratagene), and their end sequences were determined.

2.6 *DNA sequencing and analyses of novel prolamin cDNAs.*

Nucleotide sequencing were performed using the Taq Dye Primer Cycle Sequencing kit in combination with a model 373A DNA sequencer (Applied Biosystems). Nucleotide and amino acid sequence data were analyzed using the GENETYX (SDC), as well as a multiple sequence alignment program, Clustal W [Thompson *et al.*, 1994], and a phylogenic tree drawing program, TREEVIEW [Page 1996], and signal peptide prediction program, Signalp Ver.1.0 [Nielsen *et al.*, 1997]. A phylogenic analysis was carried out by the neighbor-joining method [Saito and Nei 1987]. Database searches were performed with the NCBI BLAST algolithum search program [Altschul *et al.*, 1997].

2.7 *Northern and Southern blot analyses*

Total RNAs were prepared from 8, 14, 22 and 28 DAF developing seeds by the SDS-phenol method [Yamagata *et al.*, 1986]. Ten μg of total RNAs from each tissue were fractionated on 1.2% (w/v) agarose gels containing formaldehyde, then blotted onto Hybond-N membranes. Filters were hybridized with ^{32}P-labeled cDNA inserts of prolamin clones, λRM1, λRM4 and λRM9, in 50% formamide, 6×SSC, 5×Denhardt's solution, 0.5% SDS and 100 μg ml^{-1} salmon sperm DNA at 42°C overnight, and washed with 0.2×SSC and 0.1% SDS at 65°C.

For Southern blotting, genomic DNA was prepared from rice germ as described [Ausubel *et al.* 1987], then digested with restriction enzymes (*Eco*R I and *Hind* III). These DNA fragments were separated on 0.8% agarose gels and transferred to Hybond-N

membranes. Filters were hybridized with ^{32}P-labeled cDNA inserts of prolamin clones, λRM1, λRM4 and λRM9, in 6×SSC, 5×Denhardt's solution, 0.5% SDS and 100 μg ml^{-1} salmon sperm DNA at 55°C overnight, and washed with 0.2×SSC and 0.1% SDS at 65°C.

3. Results

3. 1 Extraction of 13 kDa prolamin polypeptides

We extracted prolamin polypeptides from the PB-I fraction with solutions containing 4 M urea and 2% SDS in the presence or absence of a reducing agent, TBP (**Fig. 1-A**). The PB-I fraction was enriched from a homogenate fraction of developing seeds by centrifugation, and then purified by pepsin digestion as described [Ogawa *et al.*, 1987] . Our preliminary experiments indicated that neither 4 M urea nor 2% SDS alone permitted extraction of prolamin polypeptides, but buffer containing both these reagents led to a good extraction of the 13 kDa polypeptides (**Fig. 1-B**). Under nonreducing conditions, mainly the 13 kDa polypeptides were extractable in N fraction from the PB-I fraction. In the presence of TBP, more the 13 kDa polypeptides were extracted in R fraction from the residue together with the 10 kDa and 16 kDa prolamins. These results indicated that there were two classes of the 13 kDa prolamin polypeptides which differed in their solubility under reducing conditions. Ogawa *et al.*, [1987] defined two types of the 13 kDa prolamins, 13a and 13b, based on a slight difference of their mobility in an SDS-PAGE gel in the presence of the reducing agent. As "13a" was extracted both in the presence and in the absence of the reducing agent, we defined "13a-1" as the 13 kDa prolamin polypeptides that were extracted under nonreducing conditions, and "13a-2" as the 13 kDa prolamin polypeptides whose extraction required reducing conditions. All 13b polypeptides were extractable under nonreducing conditions (**Fig. 1-B**).

3. 2 Cysteine contents of the 13 kDa prolamin polypeptides

The presence or absence of cysteine residues in the 13 kDa prolamins polypeptides was determined using a cysteine labeling reagent, ABD-F (**Fig. 2**). The 13b polypeptides showed no labeling with the reagent even after reduction with TBP, suggesting no cysteine residue in the polypeptides. Indeed, the prolamin polypeptide sequenced by Horikoshi *et al.*, [1991] contains no cysteine residue. It is likely that this prolamin belongs to the 13b prolamin. The 13a-1 polypeptides were not labeled with ABD-F before reduction. However, the reduced polypeptide was labeled with the reagent. The molecular sizes of 13a-1 polypeptides that were determined using an SDS-PAGE without any reducing agent was smaller than that of the reduced polypeptide (**Fig. 1-B**). These results indicate that the 13a-1 polypeptides have neither free SH groups nor intermolecular disulfide bonds.

The 13a-2 polypeptides extracted under reducing conditions were also treated with the labeling reagent, and were indeed labeled. The amounts of the 13a-2 polypeptides extracted positively correlated with the concentration of the reducing agent used for extraction as well as that of 10 kDa and 16 kDa polypeptides (**Fig. 3**). The increases of soluble polypeptides also

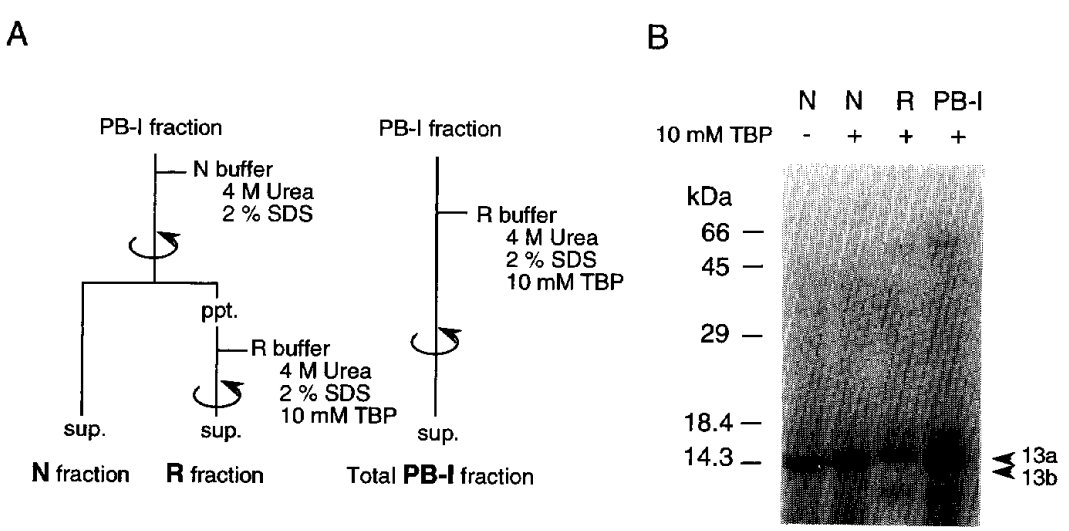

Fig. 1 Protein extraction from the rice PB-I fraction.
(A) Schematic representation of the fractionation of rice PB-I proteins. (B) SDS-PAGE profile of fractions from rice PB-I. N, N fraction; R, R fraction; PB-I, total PB-I fraction. Proteins in the N fraction were reduced or not with 10 mM TBP before SDS-PAGE. After separation by SDS-PAGE using a 16% (w/v) gel, gel was stained by CBB R-250.

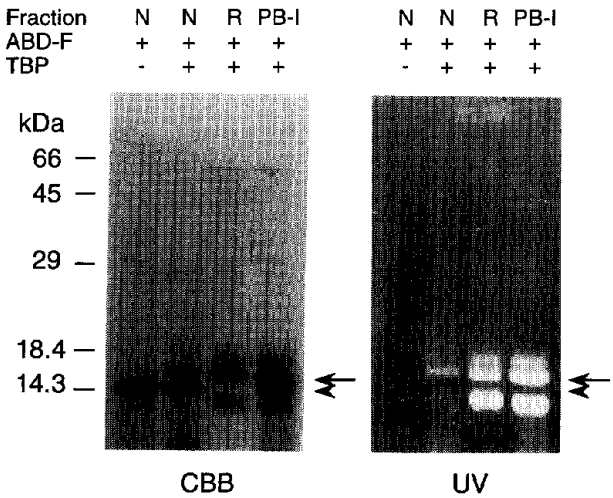

Fig. 2 Fluorescent cysteine labeling of PB-I proteins. N, N fraction extracted with N buffer (62.5 mM Tris-HCl, pH7.0 with 4 M urea and 2% (w/v) SDS) from the PB-I fraction; R, R fraction extracted with R buffer (62.5 mM Tris-HCl, pH7.0 with 4 M urea, 10 mM TBP and 2% (w/v) SDS) from the residual PB-I fraction after extraction with N buffer; PB-I, total PB-I fraction. Proteins in fractions were labeled with ABD-F in the absence or in the presence of TBP. After separation by SDS-PAGE using a 16% (w/v) gel, protein bands were visualized by staining with CBB R-250 (left) and by UV light illumination (right). Arrows and arrow head show cysteine-containing and cysteine-lacking 13 kDa prolamins, respectively.

correlated with the increased labeling of the polypeptides with ABD-F, indicating SH groups were present in the reduced polypeptides.

3.3 Purification and amino acid sequencing of 13 kDa prolamin polypeptides

The 13a-1 and 13b polypeptides extracted in N fraction without reducing agents were purified by reverse-phase column chromatography for amino acid sequencing (**Fig. 4**). The five major fractions (13n-1, 13n-2, 13n-3, 13n-4 and 13n-5) were collected for amino acid sequencing. The R fraction that was extracted from the PB-I fraction with reducing agents contained the 13a-2 polypeptides. The 13r-1 fraction were also purified from the R faction by reverse-phase column chromatography (data not shown). SDS-PAGE analysis indicated that the molecular sizes of 13r-1, 13n-1, 13n-2 and 13n-4 polypeptides were equal to the size of 13a polypeptides, the molecular sizes of the 13n-3 and 13n-5 polypeptides correspond to that of 13b polypeptides. These polypeptides were subjected to trypsin digestion. Each tryptic peptide was separated by Tricine SDS-PAGE and sequenced. NH_2-terminal sequences of the tryptic peptides are shown in **Table 1**.

After trypsin digestion, purified 13n-1 polypeptide was separated by Tricine SDS-PAGE in the presence or absence of a reducing agent, TBP. A peptide band with an apparent molecular mass of 12,000 in the absence of TBP migrated as three separate bands with apparent molecular masses of 7,000, 4,000 and 3,000 when TBP was present (data not shown).

3.4 Isolation and sequencing of novel 13 kDa prolamin cDNAs

Toward the comprehensive isolation of the 13 kDa prolamin cDNAs, we screened cDNA clones which were relatively abundant in a developing seed cDNA library. For isolating abundant cDNA clones from

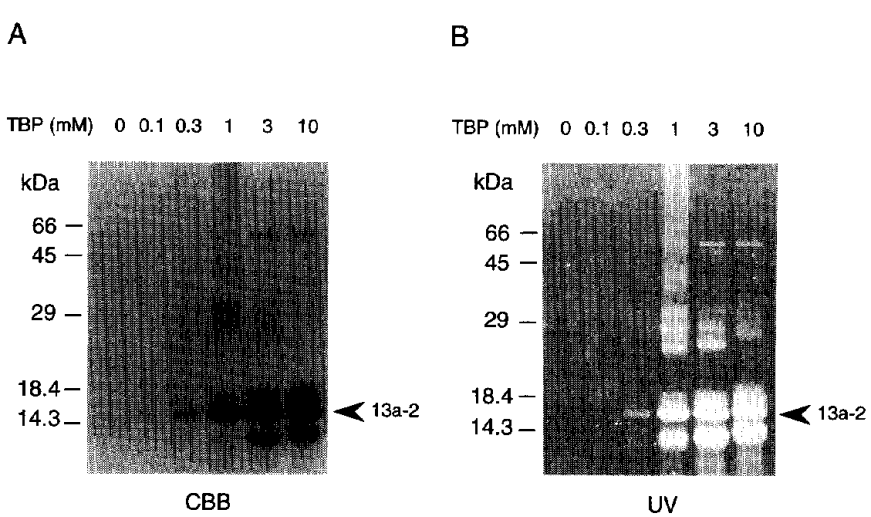

Fig. 3 Polymerization of polypeptides in rice PB-I's by intermolecular disulfide bonds. After removing of the N fraction, residual PB-I polypeptides were extracted in solutions containing various concentrations of TBP as a reducing agent. The soluble polypeptides in the solutions were labeled with ABD-F prior to SDS-PAGE. After separation by SDS-PAGE using 16% (w/v) gels, protein bands were visualized by staining with CBB R-250 (A) and by UV light illumination (B).

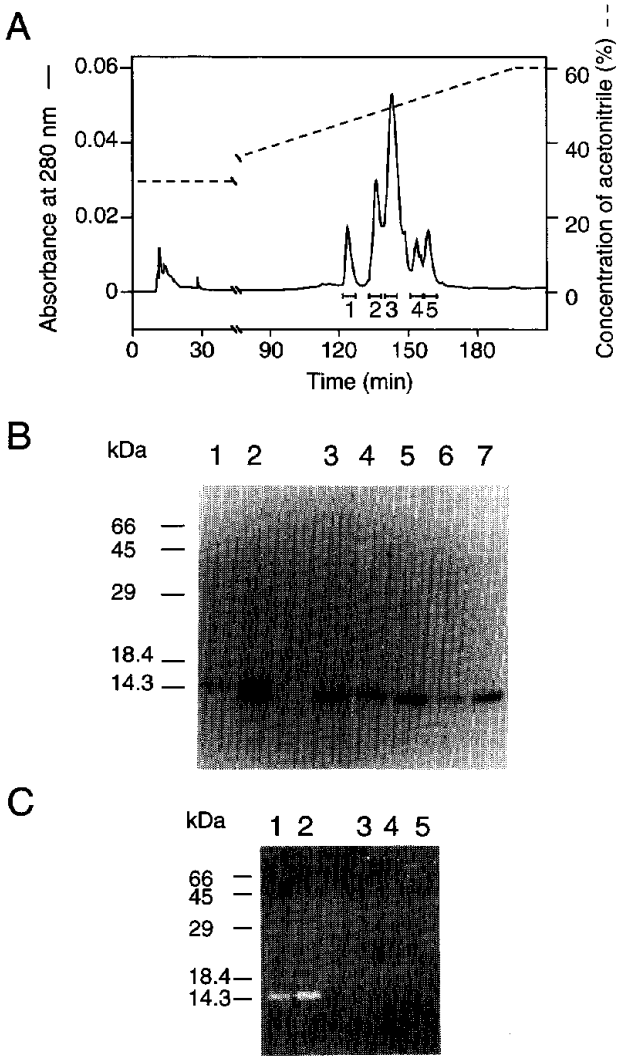

Fig. 4 Purification of the 13 kDa prolamin polypeptides.
(A) Fractionation by reverse-phase column chromatography of the N fraction from rice PB-I's. Numbers with solid bars indicate fractions (13n-1, 13n-2, 13n-3, 13n-4 and 13n-5) applied for further analyses after rechromatography. (B) and (C), The SDS-PAGE profile of fractions separated by the reverse-phase column chromatography. Lane 1-5, reduced and ABD-F treated fractions corresponding to the numbered fractions in panel A; lane 6 and 7, nonreduced fractions 13n-1 and 13n-2, respectively. After separation by SDS-PAGE using 16% (w/v) gels, protein bands were visualized by staining with CBB R-250 (B) and by UV light illumination (C).

the cDNA library, the library was screened with a single-stranded cDNA probe synthesized from poly $(A)^+$ RNA of developing seeds (14 DAF). Twenty positive clones were tested to hybridize with a mixture of previously reported storage protein cDNAs encoding 10 kDa prolamin and glutelins. Seven clones didn't hybridize with the mixed probe. These clones were subcloned into the pBluescript KS^+ and sequenced. Searches of protein database using the BLAST algorithm revealed that three clones were novel prolamin cDNAs. These clones, λRM1 (601 bp), λRM4 (596 bp) and λRM9 (617 bp) seemed to be full-length clones. The amino acid sequences deduced from these cDNAs were homologous each other, although the gene product of λRM4 carried deletions of five amino acid residues, in comparison with that of λRM1 (**Fig. 5**). A striking feature of the prolamin encoded by λRM4 was the absence of cysteine residues in their amino acid composition. The hydrophobic regions (19 residues) at the NH_2-termini of the prolamins seemed likely to be signal sequences for targeting to the ER membrane.

Based on the amino acid sequences, we classified these newly isolated prolamin genes, together with those previously isolated, into four classes, 13-I (lRM1, λRM7 [Masumura *et al.*, 1990] and pS18 [Shyur *et al.* 1992]), 13-IIa (λRM4, pProl 7 [Kim and Okita 1988a], pProl 14 [Kim and Okita 1988b] and pX24 [Shyur and Chen 1990]), 13-IIb (λRM2 [Yamagata *et al.*, 1992], and pS23 [Shyur and Chen 1990]) and 13-III (λRM9 and pProl 17 [Kim and Okita 1988b]). The polypeptides corresponding to 13-III prolamin have not been purified. However the two clones are classified as 13-III (**Fig. 6** and **Table 2**), because the nucleotide and amino acid sequences of the two cDNA clones share high homology with those of the other 13 kDa prolamin cDNA clones.

Comparison of amino acid sequences determined by protein sequencing and cDNA cloning of the 13 kDa prolamins showed that λRM1 (156 amino acid residues with the molecular mass of 15,672) encodes the 13n-1 polypeptide, while λRM4 (150 amino acid residues with the molecular mass of 14,747) encodes the 13n-3 polypeptide (**Table 1**). The amino acid sequence encoded by λRM4 perfectly matches a polypeptide purified and sequenced by Horikoshi *et al.* [1991].

The gene product predicted for λRM9 comprises 150 amino acid residues with a molecular mass of 14,888. No amino acid sequence of purified polypeptides which correspond to the cDNA clone (λRM9) has been determined in any study, including this manuscript. Therefore, it is not clear whether the 13-III prolamin gene products delineated above belong to the 13 kDa prolamins (see Discussion).

We used the nucleotide sequences of the three clones as queries to search the GenBank expressed sequence tag (EST) data base, and identified a great number of EST entries that were nearly identical to the cDNA clones (**Table 2**). The numbers of entries for 13-I, 13-IIa, 13-IIb and 13-III prolamins are 105, 95, 1 and 162, respectively.

Table 1. Correspondences of the purified 13 kDa prolamins with the prolamin cDNA clones.

Fraction	Class*1	Peptide sequences	Corresponding clones (position)
13r-1	13a	QQYSIVATPFXQPATFQLINNQV	λRM7 (53-75)
13n-1	13a	QYXLXXXLLLXXQVLXPC	λRM1 (30-47), λRM7 (30-47)
13n-2	13a	QQYSIVATPFWQPATFQLINNQVMQ	λRM7 (53-77)
13n-3	13b	QQYGIAASPFLQSAAFQLXNNQVWQ	λRM4 (53-77), pProl 14 (53-77) pX24 (52-76), pProl7 (53-77)
13n-4	13a	QQYSIAASTFLQSAAFQLXNNQVLQ	λRM2 (53-77)
13n-5	13b	QYQVQSPLLLQQQVLSLY	pS23 (29-46)

*1 These class were defined by Ogawa *et al.*, [1989] based on molecular sizes of the 13kDa prolamins on SDS-PAGE.

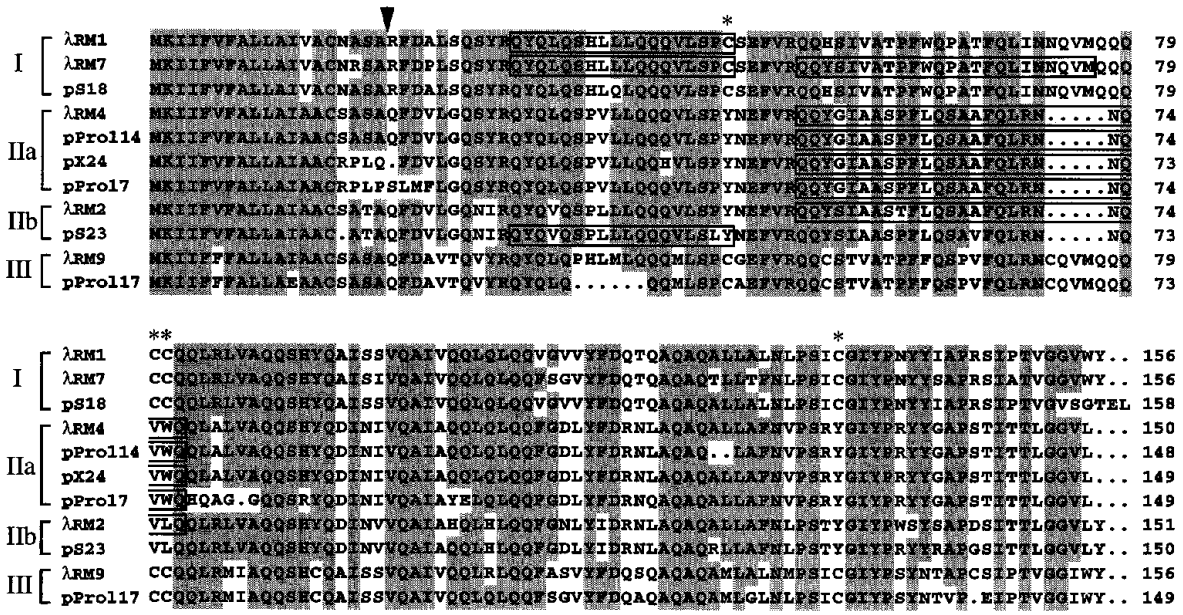

Fig. 5 Comparison of amino acid sequences deduced from rice prolamin cDNAs.
An alignment was calculated using the multiple alignment program Clustal W [Thompsom *et al.* 1994] with following rice prolamin sequences (GenBank accession numbers of the sequences are indicated in parentheses): λRM1, λRM4 and λRM9 (this work), pProl 7 (M23743) [Kim and Okita 1988a] , pProl 14 (M23744) [Kim and Okita 1988b] , pProl 17 (M23745) [Kim and Okita 1988b] , pS23 (L36819) [Shyur and Chen 1990] , pX24 (M28159) [Shyur and Chen 1990] , pS18 (X60979) [Shyur *et al.*, 1992] , λRM7 (X14392) [Masumura *et al.*, 1990] , λRM2 (D11385) [Yamagata *et al.*, 1992] . The conserved amino acid residues among rice 13 kDa prolamins are shaded. Dashes indicate gaps in the amino acid sequences used to optimally align the sequences. The putative cleavage site of the NH_2-terminal signal peptide is marked by an arrowhead. The positions of four conserved cysteine residues among 13-I and 13-III prolamins are marked by asterisks. The amino acids identified by peptide sequencing are boxed.

3. 5 Northern and genomic Southern blot analyses

Expression of the prolamin genes was examined in developing seeds by Northern blot analysis (**Fig. 7**). Full-length inserts of the cDNA clones were used as probes. Transcripts of the prolamin genes were weakly detected at 8 DAF, and increased after 14 DAF and remained high until 28 DAF.

Genomic Southern blot analysis was performed using the same set of the probes as Northern blot analysis. Multiple signals were detected with all probes (**Fig. 8**). The hybridization patterns of λ RM1 and λRM4 were similar to those of λRM7 [Masumura *et al.*, 1990] and pProl 14 [Kim and Okita 1988b] , respectively. The numbers of genes hybridized with λRM7 and pProl 14 have been estimated 7-10 [Masumura *et al.*, 1990] and 80-100 [Kim and Okita 1988b] copies per haproid genome, respectively. Our results support that all 13 kDa prolamin polypeptides (13-I, IIa and III) are encoded by multigene families.

4. Discussion

In this study, we purified and sequenced six 13 kDa prolamin polypeptides from the rice PB-I fraction. And we cloned three cDNAs that encode 13 kDa prolamins. Based on the homology of amino acid sequences and the numbers of cysteine residues, we

Table 2 Classification of rice prolamins and their corresponding cDNA clones.

Classes	Purified proteins	Cysteine residues	EST clones*1	Corresponding cDNA clones/ Reference
13kDa prolamin				
13-I	13r-1, 13n-1, 13n-2	4	105	λRM1[this work] λRM7[Masumura *et al.* 1990] pS18[Shyur *et al.* 1992]
13-IIa	13n-3, major prolamin [Horikoshi *et al.* 1991]	0	95	λRM4[this work] pProl 14[Kim and Okita 1988b] pX24[Shyur and Chen 1990] pProl 7 [Kim and Okita 1988a]
13-IIb	13n-4, 13n-5	0	1	λRM2[Yamagata *et al.* 1992] pS23 [Shyur and Chen 1990]
13-III*2	not purified	7-8	47	λRM9[this work] pProl 17[Kim and Okita 1988b]
10kDa prolamin	10kDa prolamin	10	162	λRP10[Masumura *et al.* 1989b]
16kDa prolamin	16kDa prolamin	12	38	λRP16 [Mitsukawa *et al.* unpublished results]

*1Numbers of EST clones which show more than 90% homology with the corresponding cDNA clones in thier amino acid sequences.

*2Proteins in this class have been not purified and estimated their molecular sizes. However due to similarities of type III cDNA clones with the cDNAs in other 13 kDa classes, the type III clones are clasified as one of the 13 kDa prolamin classes.

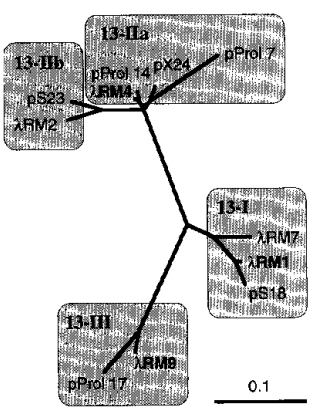

Fig. 6 Unrooted phylogenic tree inferred from amino acid sequences of following rice 13 kDa prolamins (GenBank accession numbers of the sequences are indicated in parentheses): λRM1, λRM4 and λRM9 (this work), pProl 7 (M23743) [Kim and Okita 1988a] , pProl 14 (M23744) [Kim and Okita 1988b] , pProl 17 (M23745) [Kim and Okita 1988b] , pS23 (L36819) [Shyur and Chen 1990] , pX24 (M28159) [Shyur and Chen 1990] , pS18 (X60979) [Shyur *et al.*, 1992] , λRM7 (X14392) [Masumura *et al.*, 1990] , λRM2 (D11385) [Yamagata *et al.*, 1992] . The amino acid sequences were aligned using the multiple alignment program Clustal W [Thompson *et al.*, 1994] . Based on this alignment, a phylogenic analysis was carried out by the neighbor-joining method [Saito and Nei 1987] and a phylogenic tree was drawn using the software TREEVIEW [Page 1996] . The scale bar indicates a divergence of 0.1 amino acid substitution per site.

classified the 13 kDa prolamins into four classes, 13-I, 13-IIa, 13-IIb and 13-III.

13-III prolamin polypeptides have not been purified. So there is no conclusive evidence that 13-III prolamin cDNA clones belong to the 13 kDa prolamin family. The complete amino acid sequences of the "major prolamin" determined by Horikoshi *et al.*, [1991] matched perfectly with the deduced amino acid sequence of λRM4 (13-IIa prolamin), which was estimated to have a molecular mass of about 15,000. The partial amino acid sequences determined from NH_2-termini of 13 kDa prolamin polypeptides by Masumura *et al.* [1990] also correspond with 13-IIa prolamin cDNA clones. Although Shyur *et al.*, [1992] reported molecular cloning of a "16 kDa prolamin" and NH_2-terminal sequencing of a prolamin, the gene and the protein belong to the 13-I and 13-IIa prolamin, respectively. The lengths of ORFs of pProl 17 (149 a.a.) and λRM9 (156 a.a.) are very similar to the length of ORFs of 13-I and 13-II prolamin cDNA clones (148 a.a.-156 a.a.). But we cannot exclude the possibility that the 13-III polypeptides might migrate as 16 kDa bands in SDS-PAGE due to their peculiar amino acid composition. However we classify 13-III prolamins as one of the 13 kDa prolamins defined by Tanaka *et al.*, [1980] because amino acid and nucleotide sequences of 13-III cDNA clones share higher homology with 13 kDa prolamins than a 16 kDa prolamin [Mitsukawa *et al.* unpublished results] .

Cysteine contents and disulfide binding manner of 13 kDa prolamins affect their solubility in nonreducing solutions. We demonstrated that the insolubility of

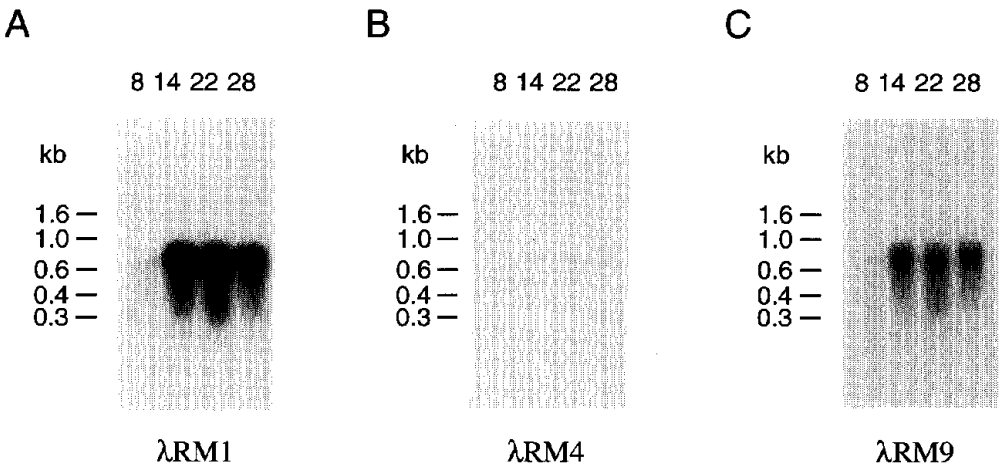

Fig. 7 Expression of λRM1, λRM4 and λRM9, during seed maturation. Total RNAs (10 μg) from developing seeds (8, 14, 22 and 28 DAF) were used Northern blot analysis. Filters transferred the RNAs were hybridized with the ^{32}P-labeled inserts from λRM1 (A), λRM4 (B) and λRM9 (C) cDNAs, and washed with 0.2×SSC and 0.1% SDS at 65°C.

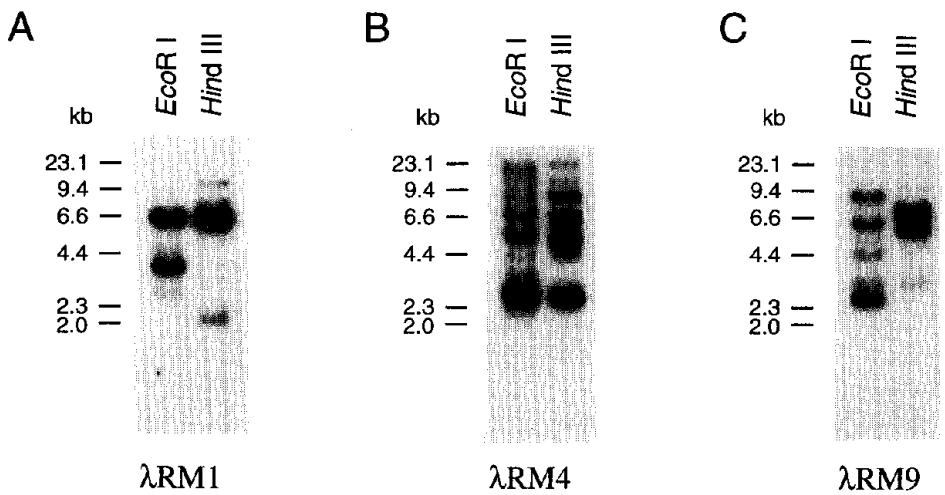

Fig. 8 Southern blot analyses of λRM1, λRM4 and λRM9. Rice genomic DNA (10 mg) was digested with *Eco*R I and *Hind* III, and separated on 0.8% agarose gels. Filters transferred genomic DNAs were hybridized with the ^{32}P-labeled inserts from λRM1 (A), λRM4 (B) and λRM9 (C) cDNAs, and washed with 0.2×SSC and 0.1% SDS at 65°C.

13-I prolamin polypeptides in nonreducing solutions is attributed to intermolecular disulfide bonds which consequently polymerize the individual polypeptides into much larger aggregates (**Fig. 3**). The 13-I prolamin polypeptides contain four cysteine residues. These polypeptides were soluble in nonreducing solution when their cysteine residues form intramolecular disulfide bonds, but were not soluble at intermolecular bonding. The position of cysteine residues in amino acid sequences of 13-I and 13-III prolamins are conserved as in an oat avenin sequence [Egorov *et al.*, 1994] . All eight cysteine residues in the avenin form intramolecular disulfide bonds [Egorov *et al.* 1994] , so this avenin polypeptide is soluble in nonreducing solutions. 13-IIa and 13-IIb prolamin polypeptides have no cysteine residues, they are easily extracted from PB-I's using nonreducing solutions.

It seems that both 13-I and 13-IIa prolamins are major polypeptides among all 13 kDa prolamin classes. The total amount of purified 13-I prolamin polypeptides (13r-1, 13n-1, 13n-2) is equal to the amount of 13-IIa prolamins. The 13n-3 fraction was separated as a major peak with the reverse-phase column chromatography. Horikoshi *et al.*, [1991] reported purification and sequencing of a "major prolamin" encoded by λRM4. The elution profile of our reverse-phase column chromatography is similar to the profile reported by Horikoshi *et al.*, [1991] . The major prolamin seems to be same as 13n-3 fraction. Search of EST database also reveals that 13-I and 13-IIa homologues are abundant among rice prolamin homologues in the database entries (**Table 2**).

Further investigation will be needed to determine

the characteristics of 13-III prolamin polypeptide. The 13-III prolamin sequences (λRM9 and pProl 17) have a high number of cysteine residues, so the 13-III prolamin polypeptides will be extractable only in reducing solutions. Elucidation of correspondences of prolamin polypeptides with their cDNA clones will be possible to make DNA markers for individual proteins. The DNA markers for individual prolamin polypeptides should be useful for the breeding of more nutritional rice.

Acknowledgements

We are grateful to Dr. Shibata Daisuke (Mitsui Plant Biotech. Res. Inst.) for stimulating discussion and suggestions. We are indebted to Dr. Robert F. Whittier (Amersham Pharmacia Biotech) for critical reading the manuscript. This research was supported by grants from the Ministry of Education, Science and Culture (Japan), and "Research for the Future" Program of the Japan Society for the Promotion of Science.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res., **25**: 3389-3402.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K., 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience, New York.

Bradford, M. M., 1976. Rapid and quantitative method for microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem., **72**: 248-252.

Egorov, T. A., Musolyamov, A. K., Andersen, J. S., Roepstorff, P., 1994. The complete amino acid sequence and disulphide bond arrangement of oat alcohol-soluble avenin-3. Eur. J. Biochem., **224**: 631-638.

Findlay, J. B. C., Geisow, M. J., 1989. Peptide preparation and characterization. In : Findlay, J. B. C., Geisow, M. J. (Eds.) : Protein sequencing, p. 43-68, IRL press, Oxford.

Geisow, M. J., Aitkin, A., 1989. Gas- or pulsed liquid-phase sequence analysis. In : Findlay, J. B. C., Geisow, M. J. (Eds.) : Protein sequencing, p. 85-98, IRL press, Oxford.

Hibino, T., Kidzu, K., Masumura, T., Ohtsuki, K., Tanaka, K., Kawabata, M., Fujii, S., 1989. Amino acid composition of rice prolamin polypeptides. Agric. Biol. Chem., **53**: 513-518.

Horikoshi, M., Kobayashi, H., Yamazoe, Y., Mikami, B., Morita, Y., 1991. Purification and complete amino acid sequence of a major prolamin of rice endosperm. J. Cereal Sci., **14**: 1-14.

Kim, W. T., Okita, T. W., 1988a. Nucleotide and primary sequence of a major rice prolamine. FEBS Lett., **231**: 308-310.

Kim, W. T., Okita, T. W., 1988b. Structure, expression and heterogeneity of the rice seed prolamines. Plant Physiol., 88: 649-655.

Kirley, T. L., 1989. Reduction and fluorescent labeling of cyst(e)ine-containing proteins for subsequent structural analyses. Anal. Biochem., **180**: 231-236.

Krishnan, H.B., White, J.A., 1995. Morphometric analysis of rice seed protein bodies. Plant Physiol., **109**: 1491-1495.

Li, X., Okita, T.W., 1993. Accumulation of prolamines and glutelins during rice seed development: a quantitative evaluation. Plant Cell Physiol., **34**: 385-390.

Masumura, T., Hibino, T., Kidzu, K., Mitsukawa, N., Tanaka, K., Fujii, S., 1990. Cloning and characterization of a cDNA encoding a rice 13kDa prolamin. Mol. Gen. Genet., **221**: 1-7.

Masumura, T., Kidzu, K., Sugiyama, Y., Mitsukawa, N., Hibino, T., Tanaka, K., Fujii, S., 1989a. Nucleotide sequence of a cDNA encoding a major rice glutelin. Plant Mol. Biol., **12**: 723-725.

Masumura, T., Shibata, D., Hibino, T., Kato, T., Kawabe, K., Takeba, G., Tanaka, K., Fujii, S., 1989b. cDNA cloning of an mRNA encoding a sulfur-rich 10kDa prolamin polypeptide in rice seeds. Plant Mol. Biol., **12**: 123-130.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng., **10**: 1-6.

Ogawa, M., Kumamaru, T., Satoh, H., Iwata, N., Omura, T., Tanaka, K., 1987. Purification of protein body-I of rice seed and its polypeptide composition. Plant Cell Physiol., **28**: 1517-1527.

Ogawa, M., Kumamaru, T., Satoh, H., Omura, T., Park, T., Shintaku, K., Baba, K., 1989. Mutants for rice storage proteins: 2. Isolation and characterization of protein bodies from rice mutants. Theor. Appl. Genet., **78**: 305-310.

Page, R. D. M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci., **12**: 357-358.

Resurreccion, A. P., Juriano, B. O., 1982. Properties of poorly digestible fraction of protein bodies of cooked milled rice. Qual. Plant Plant Foods Hum. Nutl., **31**: 119-128.

Saitou, N., Nei, M., 1987. The neighbor-joining

method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., **4**: 406-425.

Schägger. H., von Jagow, G., 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem., **166**: 368-379.

Shyur, L. F., Chen, C. S., 1990. Nucleotide sequence of two rice prolamin cDNAs. Nucl. Acids Res., **18** : 6683.

Shyur, L. F., Wen, T. N., Chen, C. S., 1992. cDNA cloning and gene expression of the major prolamins of rice. Plant Mol. Biol., **20**: 323-326.

Takaiwa, F., Kikuchi, S., Oono, K., 1986. The structure of rice storage protein glutelin precursor deduced from cDNA. FEBS Lett., **206**: 33-35.

Tanaka, K., Sugimoto, T., Ogawa, M., Kasai, Z., 1980. Isolation and characterization of two types of protein bodies in the rice endosperm. Agric. Biol. Chem., **44**: 1633-1639.

Tanaka, Y., Hayashida, S., Hongo, M., 1975. The relationship of the faces protein particles to rice protein bodies. Agric. Biol. Chem., **39**: 515 -518.

Tanaka, Y., Resurreccion, A. P., Juriano, B. O., Bechtel, D. B., 1978. Properties of whole and undigested fraction of protein bodies of milled rice. Agric. Biol. Chem., **42**: 2015-2023.

Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. Clustal W: Improving the sensitivity of progressive multiple sequence aligment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl. Acids Res., **22**: 4678-4680.

Yamagata, H., Nomura, T., Arai, S., Tanaka, K., Iwasaki, T., 1992. Nucleotide sequence of a cDNA that encodes a rice prolamin. Biosci. Biotech. Biochem., **56**: 537.

Yamagata, H., Tamura, K., Tanaka, K., Kasai, Z., 1986. Cell-free sysnthesis of rice prolamin. Plant Cell Physiol., **27**: 1419-1422.

Yamagata, H., Tanaka, K., 1986. The site of synthesis and accumulation of rice storage proteins. Plant Cell Physiol., **27**: 135-145.