Mol Gen Genet (1990) 221:1–7

MGG

Cloning and characterization of a cDNA encoding a rice 13 kDa prolamin

Takehiro Masumura, Takashi Hibino, Kunitomo Kidzu, Norihiro Mitsukawa, Kunisuke Tanaka, and Shoji Fujii

Department of Biochemistry, College of Agricultue, Kyoto Prefectural University, Shimogamo, Kyoto 606, Japan

Summary. A cDNA library constructed from mRNAs obtained from developing rice endosperm was screened with a cDNA clone (λRM7) of highest frequency of occurrence (1.8%). The translation product directed by the mRNA which was hybrid-released from λRM7 cDNA in a wheat germ cell-free system showed a molecular size of 13 kDa when coexisting with the protein body fraction of developing maize endosperm. A polypeptide sequence composed of 156 amino acids was deduced from the nucleotide sequence. By comparison with the 19 N-terminal amino acids obtained from Edman degradation of the isolated rice 13 kDa prolamin fraction, the signal sequence was determined as consisting of 19 amino acids. The deduced polypeptide is rich in hydrophobic amino acids such as Leu and Val, and also in Gln, but lacks Lys. Hence, the amino acid composition is consistent with that of rice 13 kDa prolamin. By homology with previously reported cereal prolamins, only a single octapeptide sequence, Gln-Gln-Gln-Cys-Cys-Gln-Gln-Leu, which was observed in 15 kDa and 27 kDa zein, B- and γ-hordein, α/β- and γ-gliadin, and γ-secalin was conserved in the rice 10 kDa and 13 kDa prolamin. No repetitive sequences and/or sequences homologous to other cereal prolamins, except the above octapeptide, were observed for the mature 13 kDa prolamin polypeptide. The signal sequence region of the 13 kDa prolamin, however, shows homology of more than 65% in both the nucleotide sequence and the amino acid sequence with rice 10 kDa prolamin and maize zein.

Key words: Prolamin – *Oryza sativa* – Storage protein – cDNA – N-terminal sequence

Introduction

Plant seed storage proteins are unique in that they are expressed at a very specific stage after pollination, and then only by a specific cell type. Many legumes generally express their storage proteins in the cotyledons. In contrast, cereal storage proteins are expressed strongly in scutellar and/or endospermal cells. The storage proteins are compartmentalized within subcellular organelles, known as protein bodies, which serve to stabilize the protein reserves for the next generation (Pernollet 1978). Such storage proteins are also important as food for animals, including humans and as such are agronomically important.

Offprint requests to: K. Tanaka

Enormous amounts of biological and nutritional information have been accumulated about these proteins, especially information concerning the types of proteins and their amino acid compositions for agronomically important seeds. Storage proteins typical of cereals and legumes have been investigated extensively and their amino acid sequences have been determined. Generally, however, it is difficult to elucidate full-length amino acid sequences by conventional methods, such as Edman degradation (Edman and Begg 1967), even when using the recently developed automated amino acid sequencers. The major difficulty in the sequencing of seed storage proteins stems from their microheterogeneity which hinders the isolation and purification of individual storage polypeptides. However, using DNA cloning techniques, information on the amino acid sequence for a single storage polypeptide can be obtained even though microheterogeneity exists in the original mRNA fraction. Thus, the amino acid sequences for many of the globulins of legumes and the prolamins of cereals have been deduced by cDNA analysis. Such accurate information on individual storage polypeptides is essential to understand fully the detailed mechanisms responsible for the deposition of polypeptides to form a storage protein body.

In the case of rice, prolamin accumulates within 1- to 2-μm-diameter spherical protein particles with a concentric ring structure as observed by thin section electron microscopy (Yamagata and Tanaka 1986). This type of protein body originates directly from the endoplasmic reticulum membrane and is called a type I protein body, PB-I (Yamagata et al. 1982). On the other hand, glutelin accumulates within 3- to 4-μm-long ovary-shaped protein bodies which are stained more intensely than PB-I by osmium tetroxide and lead acetate staining, and they lack the concentric ring structure (Yamagata and Tanaka 1986). This type of protein body appears to originate from a vacuolar membrane and is called a type II protein body (PB-II) (Bechtel and Juliano 1980; Yamagata et al. 1982). Existence of the two types of protein bodies within a single cell is quite unique. Rice endosperm is the only case of such protein body polymorphism reported to date.

Information on the nucleotide sequence of rice storage polypeptides is rare. Takaiwa et al. (1986, 1987a, b) reported on the amino acid sequence of rice glutelin, and the major rice polypeptides were studied by Kim and Okita (1988a, b). We recently reported the amino acid sequences of a 10 kDa prolamin polypeptide (Masumura et al. 1989a) and a major glutelin polypeptide (Masumura et al. 1989b)

as deduced from the cDNA sequences. Accurate amino acid sequences deduced from nucleotide analysis reinforce our understanding of the mechanisms of storage protein biosynthesis, transport, processing, and deposition. In the present work we describe the nucleotide sequences corresponding to the polypeptide of the major rice prolamin (13 kDa) together with the deduced amino acid sequence. The biological significance of the information derived through analysis of the amino acid sequence is indicated.

Materials and methods

Plant materials. Oryza sativa L. *japonica* cv Nipponbare was grown in the experimental field of Kyoto Prefectural University from the end of May to September. The rice seeds harvested at 8, 10, 16, 22, and 29 days after flowering (DAF) were stored under liquid nitrogen until extraction of the RNA.

Construction of cDNA library. Poly(A)$^+$ RNA was prepared from 22 DAF seeds following the method reported previously using sodium dodecylsulfate (SDS)-phenol extraction and oligo(dT) column chromatography (Yamagata et al. 1986). The synthesis of cDNA was carried out using commercially available cDNA synthesis kits (Pharmacia) and following the instructions in the supplier's manual. A cDNA library consisting of 1.4×10^6 independent clones was prepared using λgt11 and transfecting host cells of *Escherichia coli* Y1088 (Huynh et al. 1985) with double-stranded cDNAs ligated with *Eco*RI adaptor sequences.

Screening of the library by use of ^{32}P-labeled single-stranded cDNA. The following screening procedure was used to isolate a cDNA clone for the rice 13 kDa prolamin which was preferentially expressed at the later stage of rice seed development. Single-stranded cDNA probes (8 μCi/μg DNA) were prepared by reverse transcription with AMV-reverse transcriptase (Life Science) and 3000 Ci/mmol [α-^{32}P]dCTP (ICN) (Grunstein and Hogness 1975). Duplicate nitrocellulose filters (NCFs) for screenings using both the 10 DAF probe and the 22 DAF probe were prepared by transferring 2.0×10^5 independent plaques from agar plates with a lawn of *E. coli* Y1088 host cells. Positive clones were selected which produced much stronger signals with the 22 DAF probe than with the 10 DAF probe. Finally, 5 clones possessing an insert of >500 bp and which did not hybridize to glutelin or to 10 kDa prolamin cDNAs were selected for further analysis.

Subcloning and the population of λRM7 sequence in the library. cDNA inserts were subcloned into Bluescript M13$^+$ KS (Stratagene) after the inserts were recovered by *Eco*RI digestion. The subclone of interest was named pλRM7. The population of the cloned DNA in the library was determined using insert DNA ^{32}P-labeled with a random primed DNA labeling kit (Boehringer Mannheim) (Feinberg and Vogelstein 1983, 1984). The λRM7 clone exhibited the highest frequency of appearance of 1.8%, and was used for further investigation.

Hybridization-released translation. This procedure was carried out essentially according to the method of Maniatis et al. (1982). The denatured insert cDNA (about 620 bp, 5 μg) was fixed onto a NCF which was treated with the poly(A)$^+$ RNA fraction obtained from 22 DAF rice seeds. The poly(A)$^+$ RNA specifically hybridized to the fixed cDNA insert was eluted with water and then translated in an in vitro translation system using wheat germ extract (Roberts and Paterson 1973). The translation product was fractionated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli 1970), and then fluorographed (Booner and Laskey 1974). To remove the signal sequence from the nascent polypeptide, the protein body fraction obtained from developing maize seeds (28 DAF) was used (Masumura et al. 1988).

DNA sequencing. The supercoil sequencing method of Chen and Seeburg (1985) was used for nucleotide sequence determination after the cDNA insert was digested with *Pst*I endonuclease and the fragments subcloned.

N-terminal sequencing. The rice prolamin fraction was extracted with 60% (v/v *n*-propanol from protein-rich rice bran which had both the albumin and globulin fractions removed by conventional solvent extraction (Hibino et al. 1989). The prolamin fraction was further fractionated by high performance liquid chromatography (HPLC) (Japan Optic, Model Tri Roter) using a reverse phase column (Chemco SORB 300-7C_4, 4.6×150 mm) and eluting with a 10% (v/v) to 95% (v/v) *n*-propanol gradient at 8 ml/min. The major peak was recovered and subjected to N-terminal amino acid sequence determination (19 cycles) on an automated protein sequencer (Applied Biosystems Model 470A) and PTH amino acid Analyzer (120A) (Hunkapiller and Hood 1983).

Northern blot analysis. Total RNA fractions from 8, 16, and 29 DAF developing seeds were prepared using SDS-phenol (Brawerman 1974). Samples of 10 μg of each total RNA were fractionated on a 1.5% (w/v) agarose gel, then blotted onto a nitrocellulose filter according to Maniatis et al. (1982). The ^{32}P-labeled cDNA probe for the cDNA insert from pλRM7 was hybridized to the filter. The hybridized filter was then washed twice at 42° C for 45 min in $3 \times$ SSC ($1 \times$ SSC is 0·15 M NaCl, 15 mM sodium citrate), 0.1% (w/v) SDS, and subjected to autoradiography.

Southern blot analysis. Genomic DNA was isolated from rice germ (Ausubel et al. 1987) and aliquots digested with each of three restriction endonucleases (*Bam*HI, *Eco*RI, and *Hin*dIII). Each 10 μg DNA sample was fractionated on 0.8% (w/v) agarose gel, then blotted onto a NCF. Samples of *Bam*HI-digested pλRM7 DNA, and the equivalent amount of 0.5, 1, 2, 4, and 8 copies/haploid genome (Domoney and Casey 1985), were loaded on the same gel. The ^{32}P-labeled probe for the cDNA insert from pλRM7 was then hybridized to the filter. The hybridized filter was washed twice at 65° C for 45 min in $0.2 \times$ SSC, 0.1% (w/v) SDS and then autoradiographed.

Results

Identification of the polypeptide encoded by λRM7

Poly(A)$^+$ RNA that specifically hybridized to the λRM7 cDNA sequence directed the synthesis of a polypeptide of approximately 15 kDa in a wheat germ cell-free protein biosynthesizing system (Fig. 1, lane 3). The in vitro synthesized

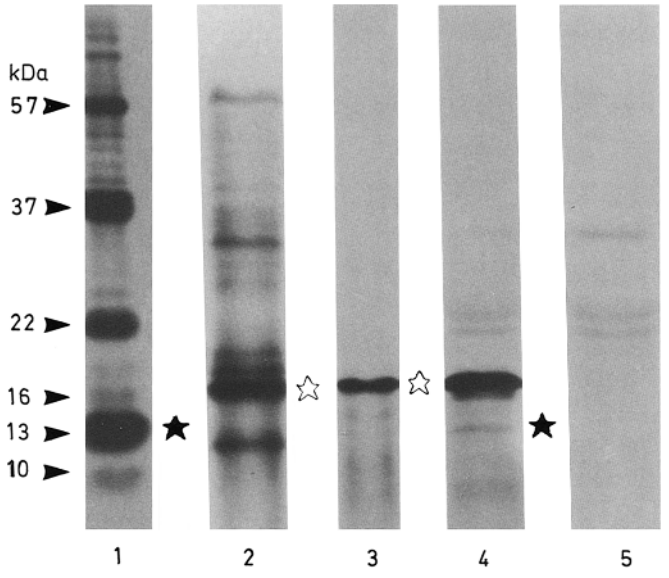

Fig. 1. Identification of cloned cDNA using hybridization-released translation. Lane 1, in vivo-labeled storage proteins of developing rice seed at 15 days after flowering (DAF); lane 2, total translation products; lane 3, translation products of mRNA released from λRM7 cDNA; lane 4, in vitro processing products of lane 3 using maize protein body (PB) membrane; lane 5, in vitro translation products using maize PB fraction without mRNA (no exogenous polypeptides). The mRNA was translated in a wheat germ system and the ^{3}H-labeled products were separated on 13.5% SDS-polyacrylamide gel and fluorographed. The precursor of the 13 kDa prolamin is marked by *open stars*. The mature sized polypeptide of the 13 kDa prolamin is marked by *solid stars*. The sizes given at the left refer to polypeptides appearing in the developing rice endosperm

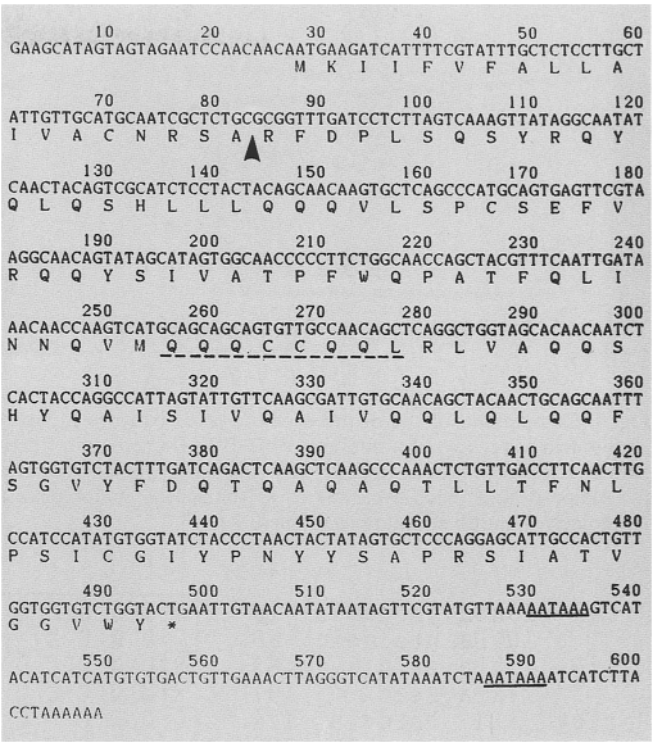

Fig. 2. Nucleotide and deduced amino acid sequences of the rice 13 kDa prolamin cDNA clone λRM7. The *arrowhead* indicates the cleavage site of the signal sequence; a unique octapeptide is shown by the *dashed underline;* the stop codon is indicated by an *asterisk;* the putative polyadenylation signals are *underlined*

peptide was 2000 dalton larger than the major rice prolamin polypeptide and coincides with the size of the previously identified precursor polypeptide for the 13 kDa prolamin. When protein synthesis took place in the presence of the protein body (PB) fraction obtained from developing maize endosperm, a new band appeared that was about 2000 dalton smaller (Fig. 1, lane 4). This new band was the same size as the 13 kDa prolamin that was prepared in vivo using ^{3}H-leucine (Fig. 1, lane 1). However, the new band was not synthesized in the presence of the PB fraction only (Fig. 1, lane 5). Thus, the in vitro process seems to reproduce completely the events that occur during 13 kDa prolamin formation within the developing rice endosperm cell. Therefore, the λRM7 insert must be coding for one of the 13 kDa prolamin polypeptides. The frequency of appearance of the λRM7 insertion sequence in the library was 1.8%. This value is fairly large considering that the values of the 10 kDa prolamin sequence and one of the glutelin sequences, λRG2, in this library were 0.9% and 1.3%, respectively.

DNA sequence of λRM7

The whole nucleotide sequence comprising 609 bp was determined and is shown in Fig. 2. The sequence has a long open reading frame starting at ATG at nucleotide number 28 and ending at TGA at position 496. The AACAATGAA sequence at the initial Met codon resembles the plant consensus initiation codon of AACAATGGC (Lutcke et al. 1987). No ATG codon is present upstream of position 28 and a stop codon, TAG, is observed at position 10. Thus, the first ATG must correspond to the initiation codon. The two AATAAA sequences at positions 530 and 586 are consensus poly(A) signals found in eukaryotic mRNA (Nevins 1983). From the deduced amino acid sequence, the precursor polypeptide of the rice 13 kDa prolamin consists of 156 amino acids and has a molecular weight of 17890 dalton. As shown in Fig. 3 the split site of the signal sequence is between Ala^{19} and Arg^{20}, the mature polypeptide consisting of 137 amino acids with a molecular weight of 15828 dalton. Within the polypeptide, Gln is present in the highest concentration at 21.7%, the Leu content is 10.1% and the Val content is 7.3%. The content of hydrophobic amino acids such as Ala, Ile, Pro, and Phe is about 5% each. No Lys residue was found. The nature of the deduced polypeptide coincides well with those of cereal prolamins already reported (Lásztity 1984). These facts are consistent with the inference that the nucleotide sequence of λRM7 is that of one of the 13 kDa prolamin polypeptides.

Homologous sequences with other prolamins

A unique octapeptide, ^{77}Gln-Gln-Gln-Cys-Cys-Gln-Gln-Leu^{84}, is contained in the rice 13 kDa prolamin amino acid sequence. This octapeptide is a unique repetitive sequence reported in rice 10 kDa prolamin (Masumura et al. 1989a), maize 15 kDa zein (Pedersen et al. 1986) and 27 kDa zein (Prat et al. 1985), barley B-hordein (Forde et al. 1985) and γ-hordein (Mills and Brandt 1988), wheat α/β-gliadin (Okita et al. 1985) and γ-gliadin (Scheets et al. 1985), and γ-secaline

```
( I )      M K I I F V F A L L A I V A C N R S A R F D P L S 25

( II )                                    (a)  Q F D V L Y  6
                                          (b)  A G Y G G G  6

( I )      Q S Y R Q Y Q L Q S H L L L Q Q Q V L S P C S E F 50

( II )  (a)  Q S V G Q Y Q L Q G P A L 19
        (b)  Q S Y R E P A V V S K V S 19
```

Fig. 3. Comparison of the deduced amino acid sequence of rice 13 kDa prolamin cDNA clone λRM7 with two sets of the N-terminal amino acid sequences of a mixture of rice 13 kDa prolamin fraction. I, Deduced amino acid sequence from λRM7 insert. II, N-terminal amino acid sequences of rice 13 kDa prolamin polypeptide fraction purified by HPLC after propanol extraction, and obtained as the major peak in each cycle (a) or as the second major peak (b). The *arrowhead* indicates the cleavage site of the signal sequence. Homologous amino acids are *underlined*

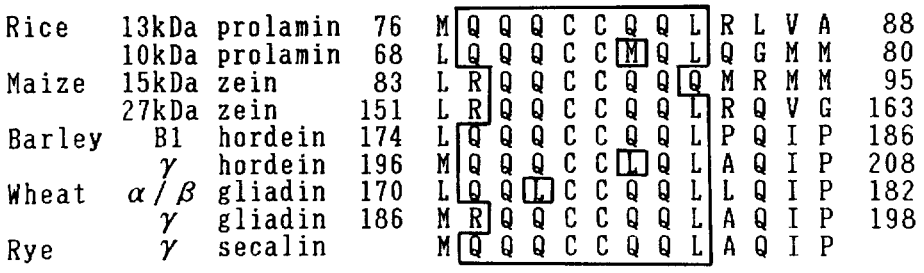

Fig. 4. Conserved amino acid sequences among cereal prolamins. Comparison of the unique amino acid sequence of 13 kDa prolamin (λRM7) with 10 kDa prolamin, 15 kDa zein, 27 kDa zein, B1-hordein, γ-hordein, α/β-gliadin, γ-gliadin, and γ-secalin deduced from the nucleotide sequences. Conserved amino acid sequences are *boxed* and numbered starting with the initiation codon Met

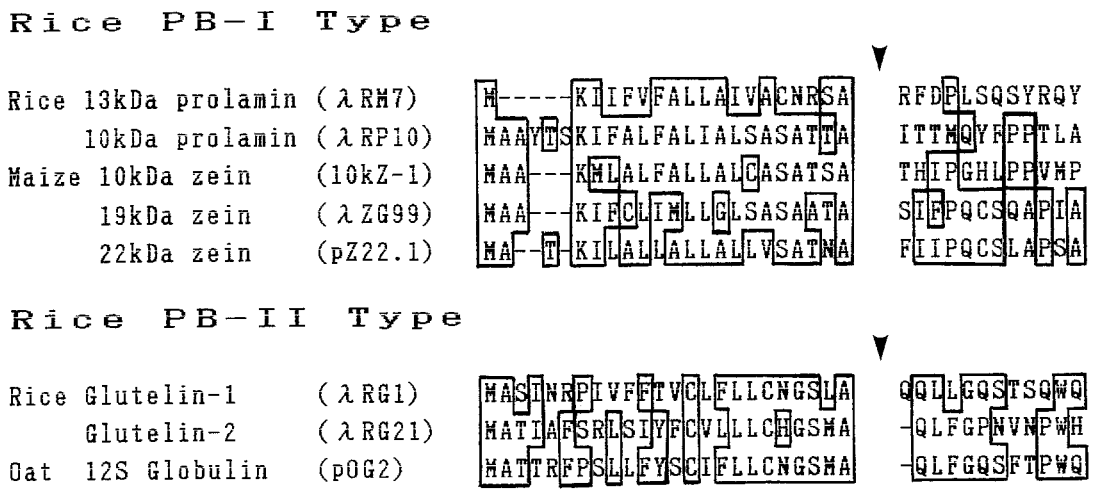

Fig. 5. Comparison of homology in the N-terminal sequences of rice prolamins, maize zeins, rice glutelins, and oat globulin. The protein bodies are classified into two types (rice PB-I type and rice PB-II type). Signal peptide sequences and partial N-terminal amino acid sequences of rice 13 and 10 kDa prolamins, maize 10, 19, and 22 kDa zeins, rice glutelins (Masumura et al. 1989b), and oat globulin (Shotwell et al. 1988) are compared. The homologous amino acids are *boxed*. An *arrowhead* indicates the cleavage site of the signal peptides

(Kreis et al. 1985) (Fig. 4). A strong homology between λRM7 and the major rice prolamin cDNAs, pProl. 14 and 17 (Kim and Okita 1988b), of more than 70% was observed. Lack of homology noted upon comparison of the nucleotide sequences of the three cDNAs corresponds to amino acid deletions rather than substitutions.

It is clear that rice prolamin polypeptides show microheterogeneity. The homology previously noted between the mature polypeptide and the corresponding nucleotide sequence of λRM7 and other cereal prolamins was not observed. However, homology of more than 60% of the nucleotide sequence, was observed in the region corresponding to the signal sequences of λRM7, rice 10 kDa prolamin, and maize zein. The amino acid sequences of the signal sequence regions for λRM7, rice 10 kDa prolamin, and maize zeins of 10 kDa (Kirihara et al. 1988), 19 kDa (Pedersen et al. 1982), and 22 kDa (Marks and Larkins 1982) are compared in Fig. 5.

Determination of N-terminal amino acid encoded by λRM7

The signal sequence comprising 24 amino acids for rice 10 kDa prolamin was previously determined by comparing the N-terminal amino acid sequence deduced from the nucleotide sequence of 10 kDa prolamin cDNA with 4 amino acid sequences obtained by Edman degradation from an isolated rice 10 kDa prolamin polypeptide (Masumura et al. 1989a). The signal sequences for 10, 19, and 22 kDa zein have also been reported (Kirihara et al. 1988; Pedersen et al. 1982; Marks and Larkins 1982; Esen et al. 1987). Comparing the homology observed in the signal sequence regions of the deduced polypeptide (λRM7), 10 kDa prolamin, and 10, 19, and 22 kDa zein, we predicted the cleavage site of the signal sequence from the proprolamin encoded by λRM7 as shown in Fig. 5. Thus, proprolamin has a signal sequence comprised of 19 amino acids.

We determined the sequence of the 19 N-terminal amino acids for rice 13 kDa prolamin using an automated amino acid sequencer. The 13 kDa prolamin extract prepared as described in the materials and methods was fractionated by HPLC and the major fraction recovered. Due to the nature of cereal prolamins, the fractionated prolamin was expected to be a mixture of several polypeptides. Actually, each cycle of Edman degradation resulted in two major peaks, except in the 7th and 8th cycles (Fig. 3, II). The cleavage site of the signal sequence in the deduced amino acid sequence of the 13 kDa prolamin precursor was determined. The N-terminal amino acid sequences obtained from the isolated 13 kDa mature prolamin fractions were aligned with the deduced amino acid sequence obtained from the λRM7 insert, as shown in Fig. 3: 14 amino acids of the 13 kDa deduced amino acid sequence were found in the 19 N-terminal amino acids of the 13 kDa mature prolamin. The signal-sequence cleavage site of the deduced 13 kDa precursor polypeptide was determined as Ala^{19}-Arg^{20}.

Northern blot analysis

Northern blot analysis was carried out using RNA obtained from three different DAF seeds to elucidate exactly when the mRNAs corresponding to the λRM7 sequence are expressed. As shown in Fig. 6, the ^{32}P-labeled probe for the pλRM7 cDNA insert hybridized exclusively at the position corresponding to 900 nucleotides. The concentration of the RNA hybridizing with pλRM7 probe increased with seed development from 8 DAF through 29 DAF. The increase in hybridizing RNA corresponds with one of the 13 kDa prolamin polypeptides that is strongly expressed in the late stage of seed development.

Genomic organization of the 13 kDa prolamin genes

Rice genomic DNA was digested with *Bam*HI, *Eco*RI, and *Hind*III restriction endonucleases for Southern blot analysis (Fig. 7). The ^{32}P-labeled probe for the cDNA insert of pλRM7 was hybridized with the blotted filter prepared from each digest. The presence of several additional hybridizing fragments suggests that the 13 kDa prolamin consists of a multi-gene family. The number of genes for the 13 kDa prolamin is large (7–10 copies/haploid genomes) under high stringency conditions.

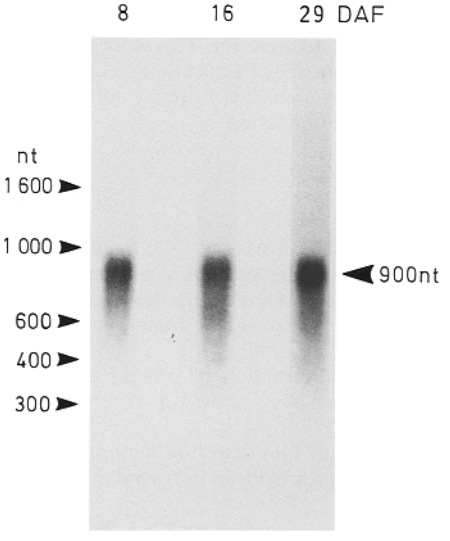

Fig. 6. Northern blot analysis of rice seed mRNA with a cDNA probe. Ten micrograms of total RNA isolated from 8, 16, and 29 DAF developing endosperm were separated by agarose gel electrophoresis in the presence of formaldehyde, and transferred to a nitrocellulose filter. The blot was probed with a ^{32}P-labeled cDNA insert of pλRM7. The length of hybridizing RNAs was estimated using the RNA molecular weight marker III (Boehringer Mannheim)

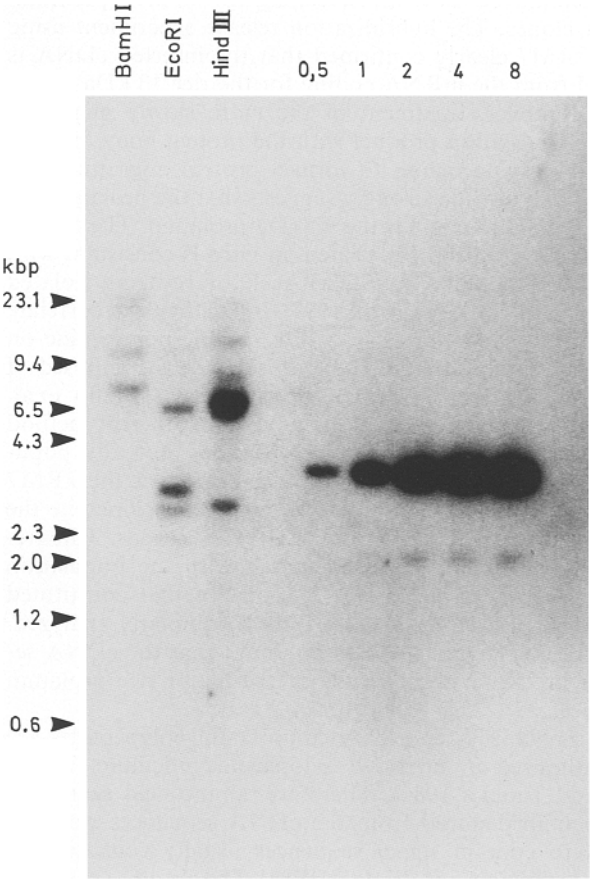

Fig. 7. Southern blot analysis of rice genomic DNA. Ten micrograms of rice germ DNA was digested with *Bam*HI, *Eco*RI, and *Hind*III, separated by agarose gel electrophoresis, and transferred to a nitrocellulose filter. The blot was probed with a ^{32}P-labeled cDNA insert of pλRM7. The number of gene copies of the 13 kDa prolamin indicated are 0.5, 1, 2, 4, and 8 copies/haploid genome. The numbers at the left indicate the position of DNA size markers in kb

Discussion

The importance of improving the protein of PB-I has been described elsewhere (Kumamaru et al. 1988). Knowledge of the molecular characteristics of polypeptides accumulating in PB-I is indispensable to accomplish this objective. However, deduction of the entire amino acid sequence directly from the isolated polypeptide is essentially impossible because of the difficulty of isolating the individual polypeptides. The only currently known method to determine the amino acid sequence of plant storage proteins involves using a cDNA for the corresponding mRNA of the polypeptide.

Kim and Okita (1988a, b) recently reported a cDNA sequence and the deduced amino acid sequence for a major rice prolamin. They did not, however, describe the corresponding polypeptide observed in SDS-PAGE analysis. The hybridization-released translation product accumulated in the present work, using λRM7, clearly illustrates that the polypeptide coded by λRM7 is a 13 kDa type, as shown in Fig. 1. This is the first report of a cDNA sequence elucidated for the rice 13 kDa prolamin.

Rice endosperm 13 kDa prolamin seems to be expressed strongly in the late stage of seed development. Therefore, a cDNA library constructed of mRNAs from late stage seeds should be used for screening for 13 kDa prolamin cDNA clones. The hybridization release experiment using clone λRM7 clearly confirmed that the inserted cDNA is derived from the mRNA coding for the rice 13 kDa prolamin. Moreover, treatment of the more slowly migrating in vitro translation product with the protein body fraction from developing maize, to form a protein migrating as a 13 kDa polypeptide strongly suggests that the prolamin encoded by λRM7 clone is the 13 kDa prolamin. The formation of a large precursor protein in vitro is consistant with the slower mobility on SDS-PAGE of proteins isolated from rice PB-I (Ogawa et al. 1987). It is difficult to correlate the cloned cDNA sequence with a specific polypeptide on SDS-PAGE, because the 13 kDa prolamin is composed of at least 7 different types of polypeptides (Hibino et al. 1989). The screening technique used is an efficient method for finding strongly expressed polypeptides. Actually, population analysis showed that the occurrence of the λRM7 insert sequence is more than 1.8% of the clones in the cDNA library prepared from 22 DAF rice seeds. This frequency seems quite large compared with the population number of one glutelin cDNA sequence that constituted only 0.9% of a 10 DAF cDNA library similarly analyzed. These results suggest there is no doubt that the cDNA sequence in λRM7 encodes one of the major rice prolamin sequences.

As previously reported, rice prolamin polypeptides are biosynthesized on the rough endoplasmic reticulum (Yamagata and Tanaka 1986). Therefore, amino acid sequences of prolamin deduced from the cDNA sequences were expected to contain signal sequences usually consisting of about 20 amino acids (Watson 1984). The cleavage site from the prolamin precursor to form the mature prolamin polypeptide can be predicted as described by Heijne (1986). Ideally, amino acid sequence information from the N-terminal of the isolated target prolamin is required to determine the signal sequence. Although the polypeptide sample subjected to N-terminal sequencing was not pure, the information obtained from the polypeptide mixture was extremely helpful. As shown in Fig. 3, the N-terminal sequences of 19 amino acids were obtained. The result suggests that the prolamin sample subjected to sequencing was composed mainly of two major sets of polypeptides. When the deduced N-terminal amino acid sequence was compared with those actually determined on the major rice 13 kDa prolamin fraction, 14 of the 19 amino acids were found in the N-terminals of the rice 13 kDa prolamins. This result conclusively locates the signal-sequence cleavage site between Ala^{19} and Arg^{20} as predicted in Fig. 2. However, the N-terminal amino acid is Gln or Ala when chemically determined. This suggests that there must be a cDNA clone that corresponds to prolamin polypeptides with Gln or Ala as the N-terminal amino acid of the mature polypeptide. Kim and Okita (1988b) proposed that the signal-peptide cleavage site of the amino acid sequence deduced from rice prolamin cDNA clones (pProl. 14 and 17) was between Ala^{14} and Cys^{15}. However, in consideration of the present results, the cleavage site should be between Ala^{19} and Gln^{20} and the N-terminal amino acids of the two clones isolated by Kim and Okita (1988b) i.e., pProl. 14 and 17 must be Gln.

The polypeptide sorting which operates in the developing rice endosperm is particularly noteworthy. The rice endosperm cell is unique in containing two types of protein body: PB-I and PB-II. The mechanism by which PB-I accumulates only prolamin polypeptides and PB-II accumulates glutelin polypeptides remains to be elucidated. Similarities in the type of protein body and the highly homologus signal sequence of cereal storage proteins are demonstrated in Fig. 5. These observations suggest that the signal sequences may play an essential role in the process of polypeptide accumulation into the protein bodies.

Acknowledgements. We are grateful to Dr. Yanagi and Dr. Ogino of the Central Institute of Sumitomo Chemical Co. Ltd. (Osaka) and to Dr. Kozo Ohtsuki and Dr. Makoto Kawabata of Kyoto Prefectural University for sequencing of the N-terminal region of the mature rice 13 kDa prolamin. We are also grateful to Dr. Rusty J. Mans for careful reading of the manuscript. This work has been supported by grants from the Ministry of Education, Science, and Culture (Japan), the Ministry of Agriculture, Forestry, and Fisheries (Japan), and the Rockefeller Foundation.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) (1987) Preparation of genomic DNA from plant tissue. In: Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience, NY, pp 2.3.1–2.3.3

Bechtel DB, Juliano BO (1980) Formation of protein bodies in the starchy endosperm of rice (*Oryza sativa* L.): A reinvestigation. Ann Bot 45:503–509

Booner WM, Laskey RA (1974) A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels. Eur J Biochem 46:83–88

Brawerman G (1974) The isolation of messenger RNA from mammalian cells. Methods Enzymol 30:605–612

Chen EY, Seeburg PH (1985) Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4:165–170

Domoney C, Casey R (1985) Measurement of gene number for seed storage proteins in *Pisum*. Nucleic Acids Res 13:687–699

Edman P, Begg G (1967) A protein sequenator. Eur J Biochem 1:80–91

Esen A, Bietz JA, Paulis JW, Wall JS (1987) A 23.8-kD alpha-zein

with N-terminal sequence and immunological properties similar to 26.7-kD alpha-zeins. Plant Mol Biol 9:421–430
Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13. Addendum (1984) Anal Biochem 137:266–267
Forde BG, Heyworth A, Pywell J, Kreis M (1985) Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucleic Acids Res 13:7327–7339
Grunstein M, Hogness DS (1975) Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci USA 72:3961–3965
Heijne GV (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683–4690
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
Hunkapiller MW, Hood LE (1983) Protein sequence analysis: Automated microsequencing. Science 219:650–659
Huynh TV, Young RA, Davis RW (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. In: Glover DM (ed) DNA cloning, vol I. IRL Press, Oxford Washington DC, pp 49–78
Kim WT, Okita TW (1988a) Nucleotide and primary sequence of a major rice prolamin. FEBS Lett 231:308–310
Kim WT, Okita TW (1988b) Structure, expression, and heterogeneity of the rice seed prolamins. Plant Physiol 88:649–655
Kirihara JA, Hunsperger JP, Mahoney WC, Messing JW (1988) Differential expression of a gene for a methionine-rich storage protein in maize. Mol Gen Genet 211:477–484
Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR (1985) Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183:499–502
Kumamaru T, Satoh H, Iwata N, Omura T, Ogawa M, Tanaka K (1988) Mutants for rice storage proteins. Theor Appl Genet 76:11–16
Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685
Lásztity R (1984) The chemistry of cereal proteins. CRC Press, Boca Raton, Florida
Lutcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differs in plants and animals. EMBO J 6:43–48
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Marks MD, Larkins BA (1982) Analysis of sequence microheterogeneity among zein messenger RNAs. J Biol Chem 257:9976–9983
Masumura T, Mitsukawa N, Fujii S, Tanaka K (1988) Accumulation of foreign polypeptides into rice PB-I type protein bodies – in vitro trials. Rice Genet Newslett 5:113–118
Masumura T, Shibata D, Hibino T, Kato T, Kawabe K, Takeba G, Tanaka K, Fujii S (1989a) cDNA cloning of an mRNA encoding a sulfur-rich 10 kDa prolamin polypeptide in rice seeds. Plant Mol Biol 12:123–130
Masumura T, Kidzu K, Sugiyama Y, Mitsukawa N, Hibino T, Tanaka K, Fujii S (1989b) Nucleotide sequence of a cDNA encoding a major rice glutelin. Plant Mol Biol 12:723–725
Mills VC, Brandt A (1988) A γ-hordein gene. Plant Mol Biol 11:449–461
Nevins JR (1983) The pathway of eukaryotic mRNA formation. Annu Rev Biochem 52:441–466
Ogawa M, Kumamaru T, Satoh H, Iwata N, Omura T, Kasai Z, Tanaka K (1987) Purification of protein body-I of rice seed and its polypeptide composition. Plant Cell Physiol 28:1517–1527
Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of the α/β-type and γ-type gliadin DNA sequences. J Biol Chem 260:8203–8213
Pedersen K, Devereux J, Wilson DR, Sheldon E, Larkins BA (1982) Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell 29:1015–1026
Pedersen K, Argos P, Naravana SVL, Larkins BA (1986) Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of M_r 15000. J Biol Chem 261:6279–6284
Pernollet JC (1978) Protein bodies of seeds: Ultrastructure, biochemistry, biosynthesis and degradation. Phytochemistry 17:1473–1480
Prat S, Cortadas J, Puigdomenech P, Palau J (1985) Nucleic acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2. Nucleic Acids Res 13:1493–1504
Roberts BE, Paterson BM (1973) Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc Natl Acad Sci USA 70:2330–2334
Scheets K, Rafalski JA, Hedgcoth C, Söll DG (1985) Heptapeptide repeat structure of a wheat γ-gliadin. Plant Sci Lett 37:221–225
Shotwell MA, Afonso C, Davies E, Chesnut RS, Larkins BA (1988) Molecular characterization of oat seed globulin. Plant Physiol 87:698–704
Takaiwa F, Kikuchi S, Oono K (1986) The structure of rice storage protein glutelin precursor deduced from cDNA. FEBS Lett 206:33–35
Takaiwa F, Kikuchi S, Oono K (1987a) A rice glutelin gene family – A major type of glutelin mRNAs can be divided into two classes. Mol Gen Genet 208:15–22
Takaiwa F, Ebinuma H, Kikuchi S, Oono K (1987b) Nucleotide sequence of a rice glutelin gene. FEBS Lett 221:43–47
Yamagata H, Tanaka K (1986) The site of synthesis and accumulation of rice storage proteins. Plant Cell Physiol 27:135–145
Yamagata H, Sugimoto T, Tanaka K, Kasai Z (1982) Biosynthesis of storage proteins in developing rice seeds. Plant Physiol 70:1094–1100
Yamagata H, Tamura K, Tanaka K, Kasai Z (1986) Cell-free synthesis of rice prolamin. Plant Cell Physiol 27:1419–1422
Watson MEE (1984) Complication of published signal sequences. Nucleic Acids Res 12:5145–5164

Communicated by H. Saedler

Received April 19, 1989 / November 21, 1989