Plant Physiol. (1988) 88, 649–655
0032-0889/88/88/0649/07/$01.00/0

Structure, Expression, and Heterogeneity of the Rice Seed Prolamines[1]

Received for publication April 19, 1988 and in revised form June 17, 1988

WOO TAEK KIM AND THOMAS W. OKITA*
Graduate Program in Plant Physiology (W.T.K.) *and Institute of Biological Chemistry* (T.W.O.), *Washington State University, Pullman, Washington 99164-6340*

ABSTRACT

By screening two rice (*Oryza sativa* L.) seed cDNA libraries, recombinant cDNA clones encoding the rice prolamine seed storage protein were isolated. Based on cross-hybridization and restriction enzyme map analyses, these clones can be divided into two homology classes. All clones contain a single open reading frame encoding a putative rice prolamine precursor (molecular weight = 17,200) possessing a typical 14-amino acid signal peptide. The deduced primary structures of both types of prolamine polypeptides are devoid of repetitive sequences, a feature prevalent in other cereal prolamines. Clones of these two homology classes diverge mainly by insertions/deletions of short nucleotide stretches and point mutations. An isolated genomic clone about 15.5 kilobases in length displays a highly conserved 2.5-kilobase *Eco*RI fragment, repeated in tandem four times, each containing the prolamine coding sequence. Close homology is exhibited by the coding segments of the genomic and cDNA sequences, although the 5′ ends of the untranslated regions are widely divergent. The sequence heterogeneity displayed by these genomic and cDNA clones and large gene copy number (~80–100 copies/haploid genome) indicate that the rice prolamines are encoded by a complex multigene family.

Prolamines, typified by their solubility in alcohol solutions, are the major seed storage proteins in most of the cereals. These proteins accumulate during endosperm development and serve as a source of nitrogen, carbon, and sulfur for the young developing seedling (12, 26). The rice prolamines have molecular sizes of about 12 to 17 kD and, as seen for other cereal prolamines, contain a high mole percentage of glutamine residues and low levels of lysine, histidine, cysteine, and methionine (18, 22). They are initially synthesized around 10 DAF (17, 31) and are deposited in protein bodies formed by direct dilation of the RER lumen (13). SDS-PAGE analysis of *in vitro* translation products purified by immunoprecipitation using a rice prolamine antibody revealed the synthesis of a 16 kD precursor form presumably containing a single peptide (14, 32).

Recently, we showed (21) that the rice prolamines are immunologically distinct from other cereal prolamines. DNA sequence analysis of a single near full length prolamine cDNA clone revealed that the derived primary sequence of the rice prolamine did not exhibit significant homology to prolamines from the major cereals (11). In this study, we report that the rice prolamines are encoded by at least two gene classes whose respective mRNA transcripts differ by about 3- to 4-fold in abundance levels during the early stages of endosperm development. Southern blot analysis of rice DNA and sequence data of recombinant DNA clones suggest that the rice prolamines are encoded by a complex family of genes.

[1] Supported in part by a grant from the Rockefeller Foundation and by Project 0590, Agricultural Research Center, College of Agriculture and Home Economics, Washington State University, Pullman, WA 99164.

MATERIALS AND METHODS

Plant Material. Rice (*Oryza sativa* L. cv Biggs M-201) was grown in an environment controlled chamber as described in detail (14). Panicles were tagged on the day of anthesis and harvested at different stages of seed development. The seeds were frozen immediately in liquid N_2 and stored at −80°C.

RNA Isolation. Total RNA was obtained by a method adapted from the established protocols reviewed by Lizzardi (16) with slight modifications as described (24). The total RNA was precipitated overnight at 4°C by addition of 5 volumes of 4 M LiCl (3). Poly(A)$^+$RNA was obtained by oligo(dT)-cellulose chromatography.

cDNA Library Construction and Screening. The preparation of double-stranded DNA complementary to rice seed poly(A)$^+$RNA and its ligation to lambda gt 11 arms or to plasmid vector were performed as described by Huynh *et al.* (9) and Heidecker and Messing (6), respectively. Lambda gt 11 recombinants were plated on host strain Y1090 at a density of 3×10^4 plaque-forming units/90-mm plate and screened with a partially purified rice prolamine antiserum as described (9). Screening of the cDNA library with a radiolabeled cDNA insert was performed by an established procedure as described (19).

Dot Blot Hybridization. Selected plasmids containing prolamine cDNA inserts were immobilized onto Zeta-Probe membrane filters (Bio-Rad, Richmond, CA) and incubated with prolamine cDNA inserts radioactively labeled by random priming (5) according to the manufacturer's recommendations. The hybridization buffer contained 50% formamide, 0.25 M sodium phosphate (pH 7.2), 0.25 M NaCl, 7% (w/v) SDS, 1 mM EDTA, 2× Denhardt's solution (19), and 7.5% (w/v) PEG 8000. Filters were then washed with 0.2× SSPE (19) and 0.1% SDS for 45 min at various temperatures as indicated in the text (19).

DNA Sequencing. cDNA inserts were isolated from cDNA clones and subcloned in M13mp18 and M13mp19. These cDNA inserts were serially deleted by exonuclease III, treated with mung bean nuclease, religated, and subcloned into *Escherichia coli* JM101 (7). This clustered set of overlapping cDNA inserts was then sequenced using the dideoxynucleotide chain termination method (25).

Isolation of High Molecular Weight DNA. Each gram of rice leaf pulverized under liquid N_2 was suspended in 2.5 mL of extraction buffer (8.0 M urea, 50 mM Tris-Cl [pH 7.5], 20 mM EDTA, 350 mM NaCl, 2% [w/v] Sarkosyl, 5% [v/v] phenol, and 20 mM β-mercaptoethanol). After successive extractions with phenol:chloroform and then chloroform, the aqueous phase was concentrated by ethanol precipitation and centrifugation. The

pellet was resuspended in 10 mM Tris-Cl (pH 7.5) and 1 mM EDTA adjusted to a density of 1.5 g/mL by the addition of saturated CsCl, and the DNA was banded overnight in a vertical rotor at 200,000*g*. The DNA band was collected, extracted with 1-butanol, and dialyzed extensively against 10 mM Tris-Cl (pH 7.5) and 1 mM EDTA.

Southern and Northern Blot Hybridization. DNA was cleaved with various restriction enzymes and then concentrated by ethanol precipitation. Cleaved DNA (5 μg) was separated on a 0.4% agarose gel alongside gene copy number standards. Total RNA was resolved on 1.2% agarose-formaldehyde gels (19). Both types of gel were transferred by capillary blot to membrane filters and hybridized with cDNA inserts as described above.

Genomic Library Construction and Screening. Rice leaf DNA was partially digested with *Eco*RI and fractionated on a 5 to 20% (w/v) sucrose gradient. Fragments between 10 and 20 kb[2] were pooled and concentrated by ethanol precipitation. The genomic fragments were ligated into purified *Eco*RI lambda Charon 35 arms, packaged into phage virions, and transfected into *E. coli* KH802 (17). The libraries were screened by established procedures using radioactively labeled cDNA probes (19).

DNA and Protein Sequence Analysis. The nucleotide and derived amino acid sequences of the prolamine clones were analyzed by the computer programs developed by the University of Wisconsin Genetics Computer Group (4). Hydrophobicity of the prolamine primary sequences was calculated according to Kyte and Doolittle (15) using an average window span of 7 residues.

RESULTS

Isolation and Classification of Prolamine cDNA Clones. By antibody screening of a cDNA lambda gt 11 library, several putative prolamine cDNA clones were obtained. Restriction enzyme digests of DNAs isolated from these clones indicated that all, except pProl 7, contained inserts of about 400 to 500 bp, approximately 50% of the mRNA transcript size as estimated by Northern blotting (data not shown). To obtain full length inserts representative of the prolamine genes, a second cDNA library was constructed using the methods of Heidecker and Messing (6). From analysis of about 500 cDNA clones with a radiolabeled prolamine cDNA insert, nine positive prolamine clones were obtained. Subsequent analysis by restriction enzyme digestion and agarose gel electrophoresis revealed that all of the clones contained inserts at least 700 bp in length. Two clones, pProl 14 and pProl 17, were 870 bp in length and were estimated to contain almost all of the sequence information of the prolamine transcript.

These near full length cDNA clones were further characterized by restriction endonuclease mapping (Fig. 1). Single restriction sites for *Fok*I, *Kpn*I, *Sph*I, *Pst*I, *Xba*I, and *Eco*RV were evident at the same relative position for both pProl 7 and pProl 14. In contrast, pProl 17 shared only the unique *Pst*I site and lacked sites for *Kpn*I, *Sph*I, and *Xba*I. Identical analysis of the remaining recombinant DNA clones indicated that they could be classified either with pProl 7 or pProl 17 by these criteria.

The extent of homology among cDNA clones was examined by the dot blot hybridization technique. Purified, random-primed cDNA inserts were hybridized at 45°C to Zeta-Probe blotting membrane filters containing 250 ng of each plasmid, and two types of cross-hybridization pattern were obtained. Figure 2 represents the dot blot cross-hybridizations of the three near full length cDNA clones. When membrane filters were washed at 45°C after hybridization, pProl 7 and pProl 14 inserts cross-hybridized extensively to each other, but these two inserts hybridized only weakly to pProl 17 (Fig. 2A, lanes 1 and 2). When the insert of pProl 17 was used as a probe, strong hybridization signals were obtained to its own DNA but cross-hybridization to pProl 7 and pProl 14 occurred to a lesser degree (lane 3). When these filters were washed at 65°C, these differences in cross-hybridization patterns were even more evident (Fig. 2B). Overall, these results, together with those obtained by physical mapping of restriction enzyme sites, indicate that the prolamines are encoded by at least two classes of genes.

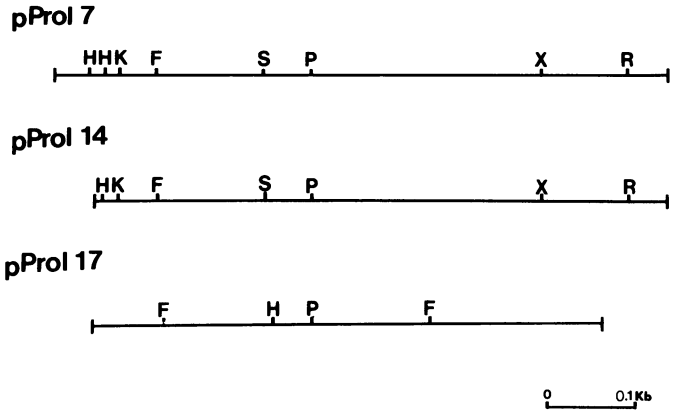

FIG. 1. Restriction enzyme map analysis of three near full length prolamine cDNA clones. F, *Fok*I; H, *Hae*II; K, *Kpn*I; P, *Pst*I; S, *Sph*I; R, *Eco*RV; X, *Xba*I.

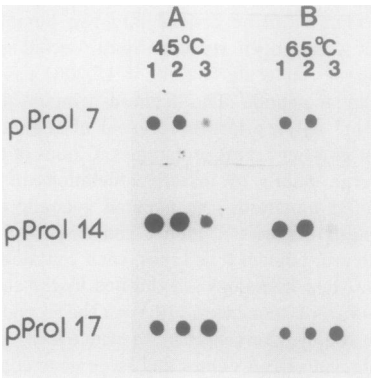

FIG. 2. Dot blot hybridization analysis of the prolamine cDNA sequences. Two hundred fifty ng of plasmid DNAs containing prolamine cDNA inserts were immobilized onto Zeta-Probe membrane filters and hybridized to ^{32}P-labeled prolamine cDNA inserts according to the manufacturer's recommendations. Filters were washed at 45°C (A) or 65°C (B) with 0.2× SSPE and 0.1% SDS. 1, pProl 7; 2, pProl 14; 3, pProl 17. The cDNA probes utilized are indicated at the left.

Structure and Primary Sequence of Rice Prolamine cDNA. DNA sequence analysis of pProl 14 and pProl 17 revealed that these two prolamine transcripts contain a single open reading frame encoding 148 and 149 amino acids, respectively (Fig. 3). Both pProl 14 and pProl 17 possess a relatively long 5′-untranslated region about 160 and 150 bases in length, respectively, followed by coding regions of 444 to 447 bp. The nucleotide sequence of the 5′-untranslated region of pProl 14 is identical to the analogous region of pProl 7 except for two point mutations and a single nucleotide insertion (11). In contrast, the 5′ end of pProl 17 is highly divergent from comparable sequences of pProl 7 or pProl 14. Despite the heterogeneity displayed by the 5′ untranslated segments of pProl 14 and pProl 17, both regions are highly enriched for G-C nucleotides and contain no potential AUG translational initiation codons. The 170-nucleotide 3′-untranslated region of pProl 14 contains two putative polyadenylation signals (AATAAG and AATAAA), whereas only a single polyadenylation signal (AATAAG) is present within the 40-base long 3′-untranslated region of pProl 17.

[2] Abbreviations: kb, kilobases; bp, base pairs; pProl, prolamine cDNA.

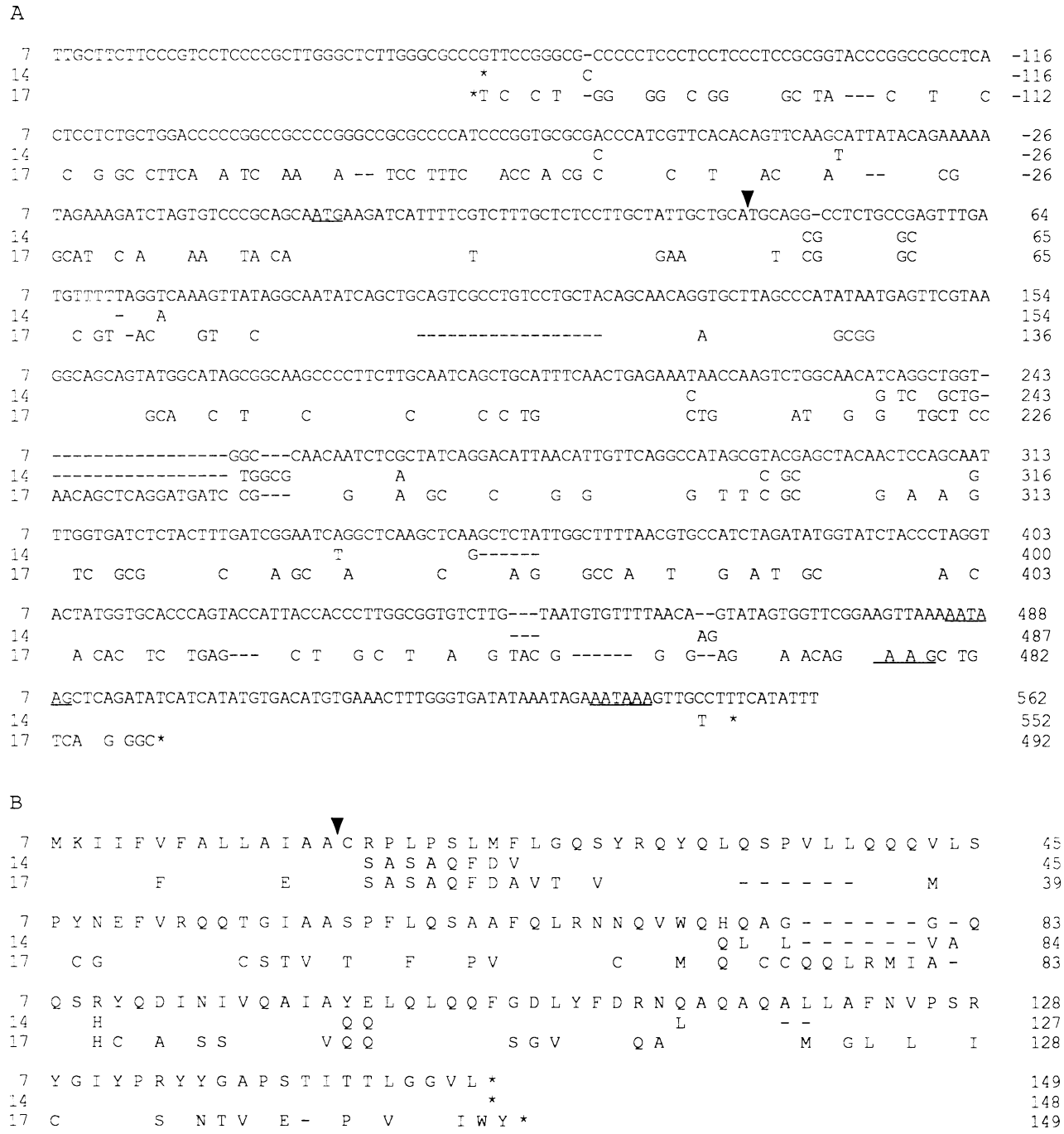

FIG. 3. Nucleotide (A) and deduced amino acid sequences (B) of prolamine cDNA clones. 7, pProl 7; 14, pProl 14; 17, pProl 17. Only the differences in the nucleotide and amino acid sequences among the cDNA clones are shown. Numbers on the right are in bp (A) or amino acid residues (B) relative to the translational start site. The putative translational initiation and polyadenylation signals are underlined. Dashed lines indicate gaps to align the nucleotide and amino acid sequences. The 5′- and 3′-ends of the cDNA clones are indicated by asterisks. The arrowhead represents the putative cleavage site of the signal peptide.

The encoded proteins of pProl 14 and pProl 17 are about 17.2 kD and this deduced molecular size is in good agreement with the value obtained by SDS-PAGE of the putative prolamine precursor (14). The composition of the first 14 amino acid residues of both proteins is typical of a signal peptide in displaying a basic amino acid, lysine, at residue 2 followed by a core of hydrophobic amino acids (Fig. 3B). These 14 amino acid residues are highly conserved in pProl 14 and pProl 17 except for Val/Phe and Ile/Glu exchanges at residues 6 and 12, respectively. Although the N terminus of the rice prolamine is blocked to Edman degradation (2), the cleavage site of the signal peptide is likely to occur between alanine (residue 14) and cysteine (residue 15), since it best conforms to von Heijne's rule for determining the signal peptide cleavage site (30). Consistent with the results from dot blot hybridization analysis, the coding sequence of pProl 14 is 95% homologous to pProl 7, whereas pProl 17 shares only 75% nucleotide sequence homology with pProl 7 or pProl 14. The predicted derived primary sequences of pProl 14 and pProl 17 share only about 63% homology. Short stretches (6–15 nucleotides) of insertions/deletions and numerous point muta-

Table I. *Amino Acid Composition (%) of Deduced Mature Prolamine Polypeptides*

	pProl 14	pProl 17
Ala	10.4	8.1
Val	6.7	8.1
Leu	11.9	6.7
Ile	4.5	5.2
Pro	4.5	5.2
Phe	5.2	5.2
Trp	0.7	0.7
Met	0.0	3.0
Lys	0.0	0.0
His	0.7	0.7
Gly	5.2	4.4
Thr	2.2	3.7
Cys	0.7	5.9
Tyr	7.5	4.4
Asx	7.5	3.7
Glx	20.1	23.7
Ser	7.5	8.2
Arg	4.5	3.0

tions are evident between pProl 14 and pProl 17, most of which result in amino acid replacements. In spite of the close DNA homology between pProl 7 and pProl 14, 8 of the 9 amino acid residues at the N terminus of the deduced mature proteins are dissimilar between these two polypeptides. This heterologous N terminus is due to a single nucleotide (C) insertion at +48 bp and a nucleotide (T) deletion at the +72 bp of pProl 14.

The deduced mature proteins of pProl 14 and pProl 17 are about 15.3 kD in size. The amino acid composition of the mature polypeptide of pProl 14 showed a high mole percentage of glutamine and hydrophobic amino acids and relatively low levels of charged residues such as lysine and histidine (Table I). The proline content of rice prolamine (5%), however, is lower than the 10 to 30 mole percentage evident in other cereal prolamines (12). The relatively higher proline content of prolamines from maize, wheat, barley, and rye (12, 26) is due in part to the presence of a tandemly repeated conserved peptide rich in this amino acid. Therefore, the lower content of proline in the rice prolamine is expected, since these proteins lack repetitive peptides prevalent in other cereal prolamines. No methionine and lysine residues are found in this rice polypeptide. This amino acid composition is in good agreement with the known rice prolamine composition (18, 22), supporting the identification of this clone. The deduced mature protein of pProl 17 also contains a high mole percentage of glutamine as well as hydrophobic amino acids. However, in contrast to pProl 7 or pProl 14, 4 methionine and 8 cysteine residues were found in this polypeptide (Table I). The presence of cysteine residues suggests that the pProl 17 polypeptide is likely to be soluble in alcohol solutions only in the presence of reducing agents.

The entire primary structures of the precursor polypeptides of pProl 14 and pProl 17 were analyzed by a function of hydropathy index (Fig. 4). The hydropathy patterns of pProl 14 and pProl 17 are similar except for three regions. These differences in hydropathy are due to changes in residues mediated by point and DNA segment mutations at these regions. Other than the hydrophobic N-terminal region containing the signal peptide, the hydropathy of the primary sequence of both prolamine polypeptides is quite neutral.

Relative Abundance of Prolamine mRNA Transcripts. The relative abundance levels for the two different types of prolamine mRNA transcripts were examined by Northern blot analysis (Fig. 5). Total RNA was isolated from rice seeds at 10 and 25 DAF and immobilized onto membrane filters. These filters were hybridized with ^{32}P-labeled pProl 14 (lane A) and pProl 17 (lane B) cDNA inserts, respectively. At 10 DAF of rice seed development, mRNA transcripts for pProl 17 are present at 3- to 4-fold higher levels than pProl 14 transcripts, whereas the relative abundance of these two mRNA transcripts is similar at 25 DAF. Overall, the expression levels of both prolamine mRNA transcripts are higher at 25 DAF than at 10 DAF.

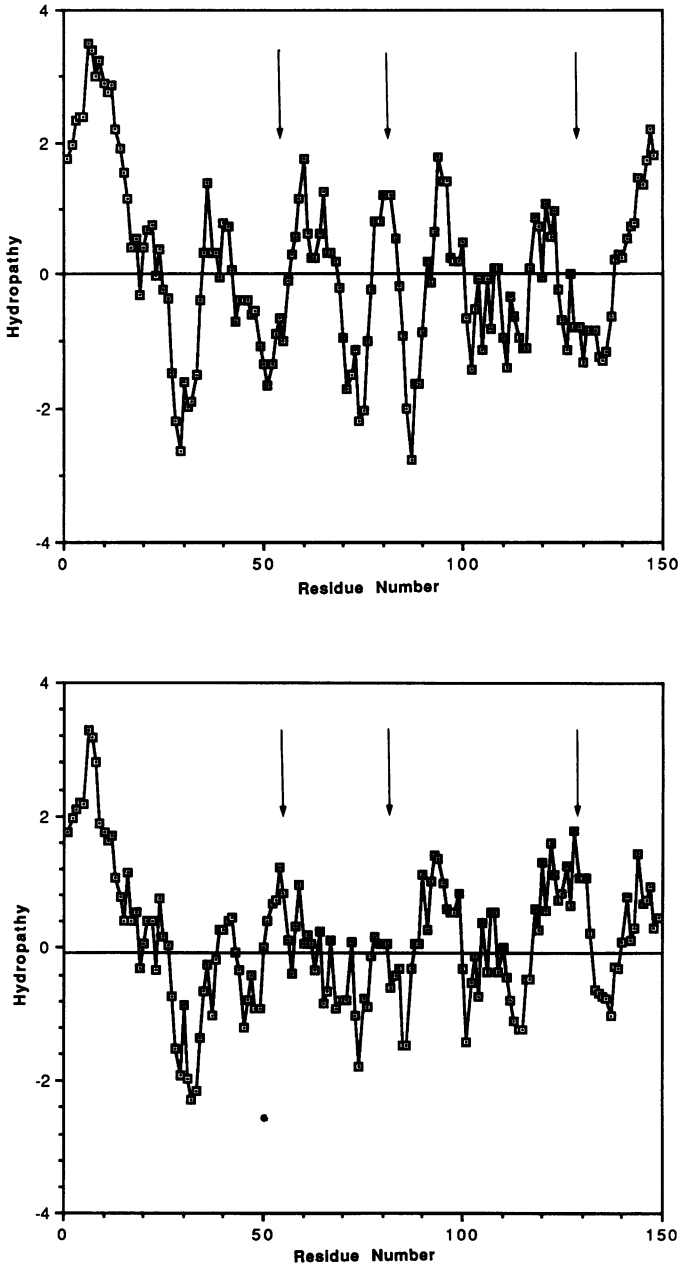

FIG. 4. Hydropathy graph of deduced prolamine precursor polypeptides. Top, pProl 14; bottom, pProl 17. The negative values of the vertical axis represent the hydrophilicity of the amino acids and the positive values show the hydrophobicity. The horizontal axis indicates the amino acid residue number. Arrows indicate the different hydropathy regions of the prolamine polypeptides.

Organization of the Rice Prolamine Genes. The organization and copy number of the prolamine genes were determined by using the genomic Southern blotting technique (Fig. 6). Rice leaf DNA was digested with *Eco*RI, *Hin*dIII, and *Bam*HI and probed with the ^{32}P-labeled pProl 14 insert. Amounts of pProl 14 equivalent to 1, 2, 10, and 20 copies/haploid genome were run on gel lanes adjacent to the rice DNA fragments. With all three enzymes, complex patterns of rice DNA restriction fragments containing the prolamine gene sequences were obtained. The *Eco*RI

FIG. 5. Northern blot analysis of rice seed RNA. Five μg of total RNA isolated from 10 and 25 DAF seeds were resolved on a 1.2% agarose-formaldehyde gel alongside identically treated *Eco*RI/*Sca*I cut pUC 19 fragments as molecular size standards. The gel was blotted onto membrane filter and the blot was hybridized to ^{32}P-labeled cDNA inserts of pProl 14 (lane A) and pProl 17 (lane B), respectively.

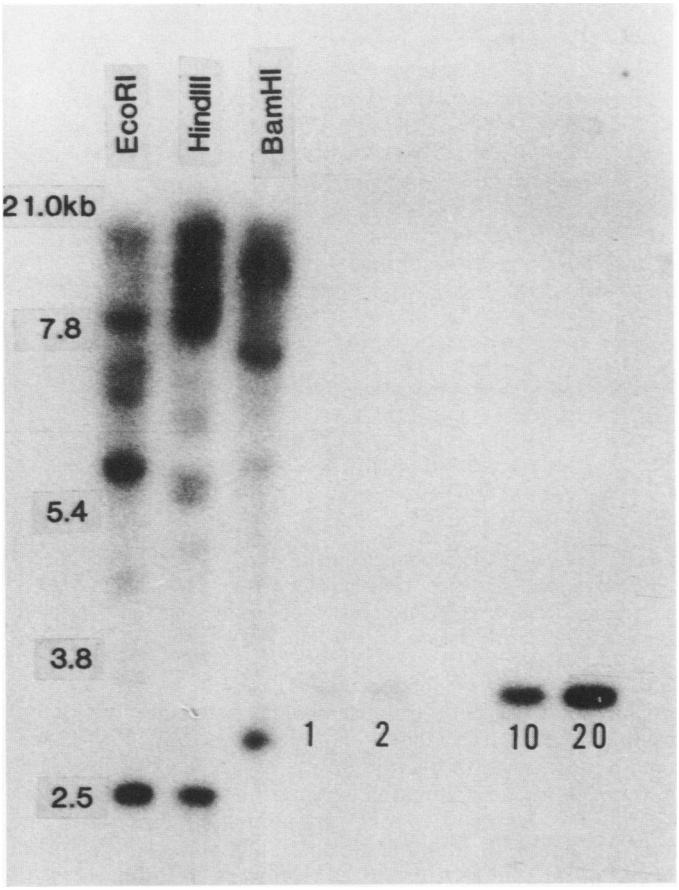

FIG. 6. Southern blot analysis of rice genomic DNA. Five μg of rice leaf genomic DNA were digested with *Eco*RI, *Hin*dIII, and *Bam*HI and resolved on a 0.4% agarose gel along with prolamine cDNA gene copy number standards. The genomic blot was hybridized with ^{32}P-labeled pProl 14 cDNA insert. Bands labeled 1, 2, 10, and 20 indicate the number of gene copies of the prolamine insert sequence/haploid rice genome.

and *Hin*dIII digestions of rice DNA gave about 10 to 20 copies of a 2.5-kb fragment when its autoradiographic signals are compared to those displayed by the internal standards. Several additional bands of 5 kb or larger size were also detected, all containing multiple copies of the prolamine sequence. Overall, the results from genomic Southern blot analysis are consistent with the notion that the number of genes for prolamines is extremely large (~80–100 copies/haploid genome).

Isolation and Characterization of the Prolamine Genomic Clones. A lambda Charon 35 rice genomic library, produced from DNA partially digested with *Eco*RI, was screened using a cDNA insert of pProl 14 as a probe, and two genomic clones were isolated. One of the isolated genomic clones, LProl 1b, contained an insert of about 12.0 kb and another, LProl 4a, possessed an insert about 15.5 kb in length. These clones were characterized by restriction enzyme mapping and Southern blot analysis using the cDNA insert as a probe. The LProl 1b clone contained two *Eco*RI fragments of 6.0 and 5.7 kb and the cDNA probe hybridized to a 3.0-kb *Xba*I-*Pst*I fragment within the 5.7-kb *Eco*RI fragment (Fig. 7A). Interestingly, LProl 4a genomic clone was found to contain four tandem repeats of a highly conserved 2.5-kb *Eco*RI fragment, each containing the prolamine gene sequences as shown in Figure 7B. The prolamine gene sequences resided in 1.1-kb *Sph*I-*Sst*I fragment within these conserved *Eco*RI fragments (Fig. 7B). One of the 2.5-kb *Eco*RI fragments was randomly subcloned into M13mp18 and M13mp19 and analyzed by DNA sequencing (Fig. 8). The encoded protein is 95% homologous to pProl 14 polypeptide. No introns are present in the coding region. Putative TATA and CAAT regulatory sequences are present at −271 and −330 bp from the translational codon. The 5′-flanking region of this genomic clone, however, contains several AUG codons in addition to the presumed authentic translational start. A protected band at −221 position from the AUG translational start signal was observed by an S1 protection experiment (results not shown).

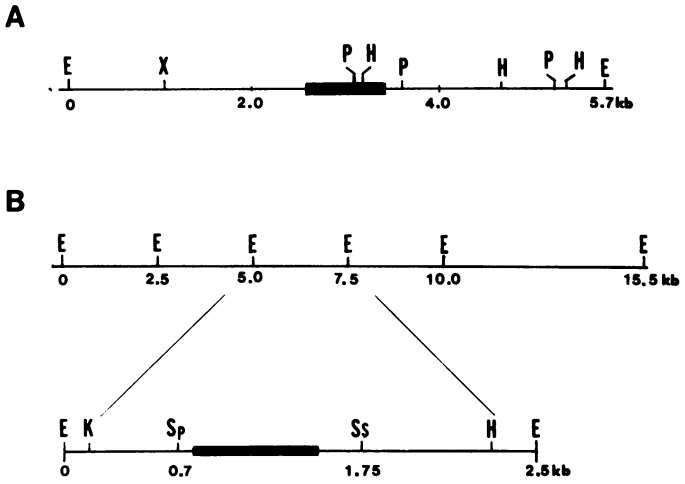

FIG. 7. Physical restriction enzyme map analysis of prolamine genomic clones. A, the 5.7-kb LProl 1b genomic fragment; B, the LProl 4a genomic clone, which contains a conserved 2.5-kb *Eco*RI fragment tandem repeated four times. E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; Sp, *Sph*I; Ss, *Sst*I; X, *Xba*I.

DISCUSSION

Although extensive studies have been conducted on the major cereal prolamines, there has been limited information about these proteins from rice seeds. As the first step to study the rice prolamines, we isolated several prolamine cDNA clones. Based on cross-hybridization and restriction enzyme map analysis, our cDNA clones can be divided into two homology classes (Figs. 1 and 2). DNA sequence analysis of pProl 14 and pProl 17, representative of these two homology classes, revealed that these cDNAs share about 75% homology at the nucleotide level and 63% homology in their derived primary sequences (Fig. 3). The divergence between the prolamine classes is due to various insertions/deletions (6–15 nucleotides) and point mutations at codon positions 1 and 2. Despite the involvement of insertion/deletion events in prolamine gene evolution, both gene classes encode polypeptides of almost identical size (15.2 kD). This result is consistent with our earlier observation that only a single prolam-

FIG. 8. DNA and amino acid sequence comparison of the 2.5-kb *Eco*RI fragment of LProl 4a genomic clone to pProl 14. Only the sequence differences between the two clones are shown. Numbers are the bp relative to translational start site. The regulatory sequences, putative TATA, CAAT, and polyadenylation signal are underlined. Dashed lines indicate the insertion/deletion segments between the genomic and cDNA clones. The asterisk indicates the 5′- and 3′-end of pProl 14. The arrowhead represents the putative cleavage site of the signal peptide. The putative transcriptional start site is underlined with bold letters.

ine band is detected on SDS-PAGE (14, 21). The 5′- and 3′-untranslated regions are also highly heterologous. Even within the same homology class, the N termini of the deduced polypeptides differ substantially because of a nucleotide insertion followed 24 bases later by a base deletion. Furthermore, varying levels of mRNA transcripts of these two prolamine classes are detected during seed development (Fig. 5), suggesting that these two classes of genes are differentially regulated.

Based on the Osborne fractionation of seed proteins, several studies have shown that prolamines constitute only a minor portion (3–5%), whereas glutelin composed the predominant protein fraction (80%) of seed protein (10, 28). The reduced glutelin fraction, when resolved by SDS-PAGE, was found to be composed of four major groups of polypeptide bands having molecular sizes of about 51, 34 to 38, 21–22, and 14 kD (14, 29). Based on Western blotting analysis, the 14 kD polypeptide was found to be a prolamine and was the major contaminant of the glutelin fraction (14). The amino acid composition of deduced prolamine polypeptides of our cDNAs exhibits an amino acid composition typically observed for rice prolamine (Table I). The encoded protein of pProl 17, however, contains an unusually high mole percentage of cysteine (5.9%) and methionine (3.0%), suggesting that this polypeptide is not a typical alcohol-soluble prolamine. Because of this unusually high percentage of cysteine residues, it seems likely that prolamine encoded by pProl 17 will remain in the glutelin fraction during the serial extraction steps of the Osborne fractionation scheme (10, 28) and therefore account for the major 14 kD contaminant observed in the glutelin fraction. In addition, Northern blot analysis using pProl 14 and pProl 17 cDNA inserts as probes revealed that pProl 17 mRNA transcripts are 3 to 4 times more abundant than pProl 14 transcripts at 10 DAF of rice seed development (Fig. 5). These results suggest that the percentage of prolamine present in total seed protein fraction is probably underestimated by the serial extraction analysis (10, 28).

Southern blot analysis of rice leaf DNA suggests that the number of genes encoding the prolamine is extremely large (~80–100 copies/haploid genome). This result is consistent with the findings from other studies (12, 26) that cereal prolamine polypeptides are encoded by a complex family of genes. We have not discounted the possibility, however, that many of the hybridizable bands observed by Southern blot analysis may be due to prolamine-like sequences present in the rice genome and that the number of authentic genes is much smaller.

One of the isolated rice prolamine genomic clones contains a highly conserved 2.5-kb *Eco*RI fragment, repeated in tandem

four times. Hybridization studies indicate that the prolamine gene sequences are contained within this conserved *Eco*RI fragment (Fig. 7). Such a contiguous tandem arrangement of seed protein genes has been previously observed for those encoding wheat γ-gliadin (23) and maize zein (27) storage proteins. Southern blot reconstruction data indicate that this 2.5-kb *Eco*RI fragment is present at about 10 to 20 copies/haploid genome (Fig. 6), suggesting the presence of several sets of closely linked prolamine genes on the rice chromosome. This tight clustering of the prolamine genes may be a consequence of the relatively small genome size of rice (~0.6×10^9 bp/haploid genome) as compared to other cereals (1).

One of the randomly chosen 2.5-kb *Eco*RI fragments was shown by DNA sequence comparison to encode a prolamine which is highly homologous to pProl 7 and pProl 14 (Fig. 8). As shown for other prolamines, the rice gene was devoid of introns (12). The 5′-untranslated region, above −70 bp from the translational start site, however, was heterologous among genomic and cDNA clones. S1 nuclease protection study indicated that the genomic clone corresponded to an RNA whose 5′ end mapped to −221 bp from the AUG translational start signal, but the S1 nuclease signal obtained was very faint after prolonged autoradiographic exposure. Moreover, the TATA box showed only weak identity to the plant consensus sequence (20). These results and the presence of several additional potential AUG translational starts in the 5′-untranslated region indicate that this gene may be only weakly expressed and does not represent a major prolamine transcript.

The alcohol solution-soluble prolamines, the predominant storage proteins of cereal seed, share a number of common features. They are encoded by multigene families and contain high amounts of glutamine and proline residues and low levels of basic amino acids such as lysine, which account for their solubility in alcohol solution (12, 26). The rice prolamines, however, are unique in lacking repetitive sequences, a common structural feature of other major cereal prolamines (12, 26). In addition, no significant homology is detected between rice prolamines and prolamines from other cereals (11). The absence of repetitive sequences and homology to other cereal prolamines at both the DNA and protein level suggests that the rice prolamine genes may have evolved from an ancestral gene distinct from those that gave rise to the prolamines of the other major cereals. This distinctive property of the rice prolamine as well as the close sequence homology evident between the rice glutelin and legume 11S storage proteins (8) may indicate that rice diverged very early during the evolution of the grasses.

LITERATURE CITED

1. Bennett MD, JB Smith 1976 Nuclear DNA amount in angiosperms. Philos Trans R Soc Lond B 274: 227–273
2. Bietz JA 1982 Cereal prolamine evolution and homology revealed by sequence analysis. Biochem Gene 20: 1039–1053
3. Cathala G, J-F Savouret, B Mendez, BL West, M Karin, JA Martial, JD Baxter 1983 A method for isolation of intact, translationally active ribonucleic acid. DNA 2: 329–335
4. Devereux J, P Haeberli, O Smithies 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Res 12: 387–395
5. Feinberg AP, B Vogelstein 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13
6. Heidecker G, J Messing 1983 Sequence analysis of zein cDNAs obtained by an efficient mRNA cloning method. Nucleic Acid Res 11: 4891–4906
7. Henikoff S 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359
8. Higuchi W, C Fukuzawa 1987 A rice glutelin and a soybean glycinin have evolved from a common ancestral gene. Gene 55: 245–253
9. Huynh TV, RA Young, RW Davis 1984 Constructing and screening cDNA libraries in lambda gt 10 and lambda gt 11. *In* D Glover, ed, Cloning Techniques: A Practical Approach. IRL Press, Oxford, pp 49–78
10. Juliano BO, D Boulter 1976 Extraction and composition of rice endosperm glutelin. Phytochemistry 15: 1601–1606
11. Kim WT, TW Okita 1988 Nucleotide and primary sequence of a major rice prolamine. FEBS Lett 231: 308–310
12. Kreis M, PR Shewry, BG Forde, J Forde, BJ Miflin 1985 Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. *In* BJ Miflin, ed, Oxford Surveys of Plant Molecular and Cell Biology, Vol 2. Oxford University Press, Oxford, pp 253–317
13. Krishnan HB, VR Franceschi, TW Okita 1986 Immunochemical studies on the role of the Golgi complex in protein-body formation in rice seeds. Planta 169: 471–480
14. Krishnan HB, TW Okita 1986 Structural relationship among the rice glutelin polypeptides. Plant Physiol 81: 748–753
15. Kyte J, RF Doolittle 1982 A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132
16. Lizzardi PM 1983 Methods for the preparation of messenger RNA. Methods Enzymol 96: 24–38
17. Luthe DS 1983 Storage protein accumulation in developing rice (*Oryza sativa* L.) seed. Plant Sci Lett 32: 147–158
18. Mandac BE, BO Juliano 1978 Properties of prolamine in mature and developing rice grain. Phytochemistry 17: 611–614
19. Maniatis T, EF Fritsch, J Sambrook 1983 Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY
20. Messing J, D Geraghty, G Heidecker, N-T Hu, J Kridl, I Rubenstein 1983 Plant gene structure. *In* T Kosuge, C Meridith, A Hollaender, eds, Genetic Engineering of Plants. Plenum Press, NY, pp 211–227
21. Okita TW, HB Krishnan, WT Kim 1988 Immunological relationships among the major seed proteins of cereals. Plant Sci 57: 103–111
22. Padhye VW, DK Salunkhe 1979 Extraction and characterization of rice proteins. Cereal Chem 56: 389–393
23. Rafalski JA 1986 Structure of wheat gamma-gliadin gene. Gene 43: 221–229
24. Reeves CD, HB Krishnan, TW Okita 1986 Gene expression in developing wheat endosperm. Plant Physiol 82: 34–40
25. Sanger F, S Nicklen, AR Coulson 1977 DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
26. Shotwell MA, BA Larkins 1988 The biochemistry and molecular biology of seed storage proteins. *In* A Marcus, ed, The Biochemistry of Plants: A Comprehensive Treatise, Vol. 15 (in press)
27. Spena A, A Viotti, T Pirrotta 1983 Two adjacent genomic zein sequence: structure, organization and tissue-specific restriction pattern. J Mol Biol 169: 799–811
28. Tecson EMS, BV Esmanan, LP Lontok, BO Juliano 1971 Studies on the extraction and composition of rice endosperm glutelin and prolamin. Cereal Chem 48: 181–186
29. Villareal RM, BO Juliano 1978 Properties of glutelin from mature and developing rice grain. Phytochemistry 17: 177–182
30. von Heijne G 1986 A new method for predicting signal sequence cleavage sites. Nucleic Acid Res 14: 4683–4690
31. Yamagata H, T Sugimoto, K Tanaka, Z Kasai 1982 Biosynthesis of storage protein in developing rice seeds. Plant Physiol 70: 1094–1100
32. Yamagata H, K Tamura, K Tanaka, Z Kasai 1986 Cell-free synthesis of rice prolamine. Plant Cell Physiol 27: 1419–1422