Plant Molecular Biology **32:** 621–630, 1996.

The same nuclear proteins bind to the 5′-flanking regions of genes for the rice seed storage protein: 16 kDa albumin, 13 kDa prolamin and type II glutelin

Masayuki Nakase[1], Takehisa Yamada[1], Takahiro Kira[1], Junji Yamaguchi[2], Naohito Aoki[1], Ryo Nakamura[1], Tsukasa Matsuda[1] and Takahiro Adachi[1,*]

[1]*Department of Applied Biological Sciences, School of Agricultural Sciences (*author for correspondence) and* [2]*Bioscience Center, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan*

Received 3 January 1996; accepted in revised form 20 May 1996

Key words: rice seed storage protein, albumin, gene expression, glutelin, prolamin

Abstract

Expression of rice seed storage-protein genes is dramatically regulated over a short period of seed maturation. To characterize the expression mechanism of the rice seed storage protein genes, their expression of major storage protein genes (16 kDa albumin, 13 kDa prolamin and type II glutelin) were compared by RNA blot analysis. Their coordinate expression suggests that the transcriptional regulatory machinery is shared among the glutelin, prolamin and albumin-genes. We isolated two novel genomic genes for prolamins (PG5a and PG5b) and obtained the promoter region of the glutelin gene by PCR. The 5′-flanking regions of these three rice seed storage protein genes were found to contain some similar conserved sequences. Nuclear extract partially purified from maturing rice seeds was used for the gel shift assay of the 5′ region of the RA gene. We identified two DNA sequences of RA gene which were recognized by independent DNA-binding proteins. The complexes of these DNA sequences and DNA-binding proteins were inhibited by the fragments containing the 5′ regions of the prolamin and glutelin genes, suggesting that these three genes share transcription factors.

Introduction

The major function of plant seed storage proteins is to provide nitrogen, sulfur and carbon sources to the seed embryo during germination. Rice seeds contain a large amount of storage proteins classified into glutelin, prolamin, globulin and albumin, and their genes have been cloned and characterized in this decade [1, 15, 17, 23, 30, 33]. Some proteins belonging to these classes of globulin and albumin have been identified as allergenic proteins [19, 21, 28]. We have isolated several 14–16 kDa albumins (RAs) belonging to the α-amylase/trypsin inhibitor family, which were recognized by sera from patients allergic to rice [1, 15]. Besides rice RA, in wheat, barley and rye, several proteins belonging to this family have been identified as allergens, suggesting that proteins belonging to this family might be potential allergens [5, 11, 21]. In addition, 2S storage proteins reveal structural homology to the α-amylase/trypsin inhibitor family [9, 27], of which castor bean 2S storage protein has been reported as allergenic protein [31]. They might share a common ancestor [18], and these proteins seem to exhibit prominent allergenicity. So far there is no good therapy for patients allergic to these proteins. It may be helpful for them if a hypo-allergenic crop would be generated by gene engineering techniques. To achieve this objective, it is necessary to further understand the properties of allergenic proteins including their gene expression mechanisms.

We have already characterized some properties of RAs at the molecular level [1, 15]. RAs are homologous proteins derived from a multigene family; several cDNAs and two genomic genes (RAG1 and RAG2) in this family have been isolated [1, 4, 15]. Comparison of the cloned cDNA sequences allows a classi-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers D73383 (PG5a) and D73384 (PG5b).

fication of RAs into at least four subfamilies (RA5, RA14, RA16 and RA17) with identities in the nucleotide sequences among and within subfamilies of 70–85% and above 95%, respectively [4]. Based on the results of the cDNA and protein sequences obtained so far, the RA14 subfamily seems to be dominant. Furthermore, although functional similarity is yet unclear, a region conserved among RA subfamilies is homologous to that of lipid transfer proteins [4]. The 5′-flanking regions of the genomic genes (RAG1 and RAG2) contain several direct-repeat sequences (consensus: ATGCAAA/T) [1] similar to those of glutelin gene [23] (consensus: TTGCAAAA), implying that these gene expression mechanisms resemble each other. Transcriptional regulatory factors have been studied in maturing soybean, tobacco, and maize seeds [3, 8, 13, 26], and recently in maturing rice seeds [14]. Although some *cis*-acting elements of the glutelin gene have been characterized [16, 34, 42, 43], those of other rice storage protein genes have not been analyzed so far.

In this study we attempted to identify the sequence in the 5′-flanking region of the RA gene that is recognized by DNA-binding proteins in maturing rice seed. In addition, we also analyzed the promoters of other rice seed storage protein genes.

Materials and methods

Genomic DNA cloning and DNA sequencing

The rice genomic DNA library used has been described previously [1]. Rice 13 kDa prolamin cDNA was used as a probe for the screening of the genomic library. The DNA sequence was determined by the dideoxy-chain termination method [25].

PCR amplification

Two primers (5′-TTCTGTAGTACAGACAA-3′ and 5′-CTCAAATGCTTGCAAC-3′) for the amplification of the 944 bp glutelin (Gt1) gene fragments (nucleotide number from −740 through +204 in Okita *et al.* [23] were synthesized in an ABI 381A DNA synthesizer (Applied Biosystems). PCR was performed with *Taq* DNA polymerase under standard conditions in the DNA Thermal Cycler (Perkin Elmer Cetus). The PCR product was subcloned into pUC plasmid vector.

RNA preparation and RNA blot analysis

Preparation of total RNA from maturing rice seed and RNA blotting were performed as described previously [1]. ^{32}P-labeled cDNA fragments of RA [15] and prolamin [17] were used as probes. In the case of glutelin, the ^{32}P-labeled PCR fragment described above was used as a probe.

Preparation of nuclear proteins

Ground rice seeds 15 days after flowering (DAF) were homogenized in NB2 buffer (1 M hexylene glycol, 10 mM PIPES-KOH pH 7.0, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.1 mM PMSF) by polytron and lysed in the NB2 buffer containing 0.5% Triton X-100. After removing the debris by filtration with Miracloth, nuclei were collected by centrifugation at 100 × *g*. Nuclei pellets were suspended in the NB2 buffer and washed three times. Then, nuclei were isolated by percoll density stepwise gradient (40%, 60% and 95%) centrifugation at 3000 × *g*. The collected nuclei were washed with the NB2 buffer, suspended in 1 ml of the NB2 buffer containing 20% glycerol, and stored at −80 °C until use. Nuclear proteins were extracted from nuclei by adding 9 ml NLB buffer (100 mM KCl, 15 mM HEPES-KOH pH 7.6, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 5 μg/ml antipin and 5 μg/ml leupeptin) and KCl (final concentration 0.5 M). After removing debris by centrifugation at 50 000 × *g*, nuclear proteins were precipitated by $(NH_4)_2SO_4$ (final concentration 80%), dissolved in 0.5 ml of the NEB buffer (40 mM KCl, 25 mM HEPES-KOH pH 7.6, 0.1 mM EDTA, 1 mM DTT and 20% glycerol), and dialyzed against the same buffer.

Gel mobility shift assay

Nuclear extracts from maturing rice seeds of 15 DAF were prepared as above, and then purified partially by Heparin Sepharose column with a linear gradient of KCl (40–1000 mM). The proteins eluted from the column with 400–500 mM KCl had binding activity to the 5′ region of the RA gene (RA-b fragments in Fig. 3). Furthermore, to optimize the interaction of the DNA/protein complex by using the RA-b probe, the effect of $MgCl_2$ concentration (0–300 mM) on the DNA-binding specificity was examined and strong activity was observed at the concentration of 20–50 mM of $MgCl_2$. For further experiments, the binding buffer containing 30 mM of $MgCl_2$ was used.

DNA fragments, RA-a and RA-b, G-a, and P-a (Fig. 3), were prepared from the 5′ regions of RAG1 (for RA-a and RA-b), Gt1 (for G-a), and PG5a (for P-a). In addition, each set of complementary oligonucleotides, 5′-GATCCTAACCATCTGTTGTCCAG-3′ and 5′-GATCCTGGACAACAGATGGTTAG-3′ for RA-I, 5′-GATCCTTGCAAAAG-3′ and 5′-GATCCTTTTGCAAG-3′ for RA-II, and 5′-GATCCT TTCGTGTACG-3′ and 5′-GATCCGTACAGAAAG-3′ for GLU-I were synthesized, tandemly ligated and subcloned into pUC vectors. Inserted fragments of those vectors together with the prepared fragments above were used for gel shift assay. ^{32}P-labeled probes were incubated with nuclear extracts in binding buffer (25 mM HEPES pH 7.6, 40 mM KCl, 30 mM $MgCl_2$, 1 mM EDTA, 0.4 μg/ml poly[dI-dC] and 10% glycerol) for 15 min at room temperature. Reaction mixtures were loaded on a 4% polyacrylamide gel and electrophoresed in TAE buffer (40 mM Tris-acetate pH 8.0, 1 mM EDTA). The gel was dried and exposed to X-ray film. Competition assay was carried out by mixing normally 200-fold molar excess amount of unlabeled competitor with nuclear extract for 15 min before adding probe.

Results

Preparation of the potential promoter fragments of the rice storage protein genes

RAG1 gene [1] and the PCR fragment of Gt1 gene [23] were used as RA and glutelin genes, respectively. In addition, we tried to clone a 13 kDa prolamin gene from a rice genomic library. A rice genomic DNA library of about 2×10^5 clones was screened by plaque hybridization with the prolamin cDNA probe. Sixteen positive clones were obtained, and one clone possessing two prolamin genes was further characterized. Two prolamin genes termed PG5a and PG5b are tandemly arranged as shown in Fig. 1A. The nucleotide sequences of two prolamin genes were determined and shown in Fig. 1B. Each of PG5a and PG5b genes encodes an open reading frame consisting of 149 amino acid residues. Neither of the genes contain introns as the prolamin gene reported before [17]. Although there are some deletions, insertions, and replacements, the coding regions of these two genes are highly similar to each other (about 98% identity). The PG5a gene was used for further analysis. In the 5′-flanking region of the PG5a gene, putative TATA and CAAT boxes (−31 through −28 and −88 through −85, respectively) are present. The 5′-flanking region of PG5a gene including the putative transcriptional initiation site is almost identical with that of LProl 4a [17], so that PG5a gene seems to be functional.

Stage-specific expression of rice seed storage protein genes

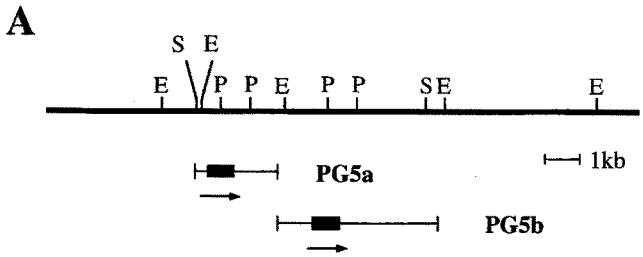
To obtain a clue in the transcriptional regulation of the rice storage protein genes, total RNA was prepared from maturing rice seeds at different stages (5, 10, 15, 20 and 25 DAF). RNA blot analysis was performed and the results are shown in Fig. 2. Although the probes used here could not distinguish the subfamilies of each gene, all of their genes were expressed strongly at 15 DAF and their expression patterns were almost the same. This result suggests that transcriptional regulation is shared among RA, prolamin and glutelin genes.

Structure of 5′ regions of rice seed storage protein genes

As mentioned in an earlier report [1], the 5′-flanking regions of RAG1 and RAG2 contain several conserved sequences (consensus: ATGCAAAA); similar sequences are also found present in the 5′ regions of Gt1 (consensus: TTGCAAAA, −303/−296 and −204/−197) and PG5a (consensus: ATCCAAAA, −572/−565 and −153/−146), though such sequences did not exist in the 5′-flanking region of the α-globulin gene [22]. Several fragments containing putative *cis* elements from rice storage protein genes were prepared as shown in Fig. 3.

Analysis of DNA-binding proteins

We prepared nuclear proteins as described in Materials and methods and performed a gel shift assay by using RA-a and RA-b as probes. Each probe formed the DNA protein complex (lanes 2 in Fig. 4A and B). When excess unlabeled probes were added into reaction mixtures, signals were disappeared depending on their dose (lane 3 and lanes 3 and 4 in Fig. 3A and 3B, respectively). To confirm specific binding activity, pUC was added in the reaction mixtures in addition to poly (dI-dC) for gel shift assay. Results show that there was no inhibition in the formation of the DNA/protein complex of RA-a (lane 4 in Fig. 4A) and RA-b (lane 7 in Fig. 4B) by the addition of pUC DNA, indicating that the binding is sequence-specific. Gel shift assays



B

```
a                                                                                                    AAGCTTGCGCGCGGAATACGGTGGTGGAGATGGGTTGGG -603

a AACCCTGGATTCCAAACACAGCCCAAGTCTATCCAAAATGTTTAGACAAGAAAATACGTAACAAGTTGGTTTACAGAAATAGCAATTAGATCAATCCTGCACTACAAGTAGAGTAAAGTG -483

a GTGATTTCTCTTAAATCTCTCGAATGGTGATTTAAGAATTCAGTGCAAACCAAATCCTTGCTATAATCAAATGTTCGGTACCCCATCAACGGAACAATAAAAAGCGCCTGGCTACCATAA -363

a TTTTGTCATTCTTCTTCAATTTGTAATTTAAGATGCATGAGGCCACACGACCTTAATGTTCAACGTGTCATGCATTAGTGAAATAATAGCTCACAAAACGCAACAAATATAGCTAGATAA -243

a CGGTTGCAATCCTTACCAAACTAACGTATAAAGTGAGCGATTAGTCATATCATTATCTCCCGCCTGCTAACCATCGTGTACACCATCCGATCCAAAAATGACAACTTCTAGGGATGAACC -123

a TGGACAAGGTTTAGGGTTTAGGGATGAATCTGGACAATGATTGTTCAGGTTCATCCCTAGATGTTGCTTTCTCCTTACGGGACGGAGGGAGTATATGTGATGGACACAAAAGTTACTTTC   -3
  *
a ATGATGAAAGGAAAGGGGATTTGTTGGGGCACTAATAGAACATCTGTCCAAATGGCATGACTCACTTATATCCTAATAGGACATCCAAGAAAAACTAACACTCTAAAGCAACCGATGAGG  118

a AATTGAAAGAAAATACGTGCCACCGCATCTATAAATGGACAAGCGCAATGGAAACCCTCCTCATCGTTCACACAGTTCAAGCATTATACAGCAAAATAGAAAGATCTATGTCCCAGAGCA  238

  M K I I F V F A L L A I A A C S A S A Q F D V L G Q S Y R Q Y Q L Q S P V L L Q   40
a ATGAAGATCATTTTCGTCTTTGCTCTCCTTGCTATTGCTGCATGCAGCGCCTCTGCGCAGTTTGATGTTTTAGGTCAAAGTTATAGGCAATATCAGCTGCAGTCGCCTGTCCTGCTACAG  358
b ..........................................................................................................................

  Q Q V L S P Y N E F V R Q Q Y G I A A S P F L Q S A A F Q L R N N Q V W Q Q L A   80
a CAACAGGTGCTTAGCCCATATAATGAGTTCGTAAGGCAGCAGTATGGCATAGCGGCAAGCCCCTTCTTGCAATCAGCTGCGTTTCAACTGAGAAACAACCAAGTCTGGCAACAGCTCGCT  478
b .....T.....................................................................................ACT.G....
     H                                                                                        H R

  G G A Q S H Y Q D I N I V Q A I A Q Q L Q L Q Q F G D L Y F D R N L A Q A Q A L  120
a GGTGGCGCACAATCTCACTATCAGGACATTAACATTGTTCAGGCCATAGCGCAGCAGCTACAACTCCAGCAGTTTGGTGATCTCTACTTTGATCGGAATCTGGCTCAAGCTCAAGCTCTG  598
b ......CA..................................................................................................
        Q

  L A F N V P S R Y G I Y P R Y Y G A P S T I T T L G G V L *                     149
a TTGGCTTTTAACGTGCCATCTAGATATGGTATCTACCCTAGGTACTATGGTGCACCCAGTACCATTACCACCCTTGGCGGTGTCTTGTAATGAGTTTTAACAGTATAGTGGTTCGGAAGA  718
b ..........................................................................................................

a TAAAAATAAGCTCAGATATCAT-ATATGTGACATGTGAAACTTTGGGTGATATAAATAGAAAAAAAGTTGTCTTTCATATTTAAATACCATGCCTCTATAAGGATATATCCTAGCACGTTG 838
b ....................C.......G....................T.....G.....................................G.............T......

a TCGTGACTAATTACCATCATCGGTACTCTACAATTTTACTGTGTGCTTACATTCGATCCGAAGCTACTTTGTTTTTTAAGATAGAAATGGAGCGTATAAAGGATGTCCGTCCTTTTATTC  958
b ..........................,.................................................................................

a CAATAAGAACAAACACAAAGCACATATGGAAAATACTAAATCCACTACAACAAAATCATAGACCATACAAAGCAGGTAAAGATGTAGCAATCCACATTGATTCTTCATCCACTACTACAA 1078

a CAGGCGCATATGAAAGGCCGCATATGCATGATCATTGTGAATATGACATGCCCTTGTCGAGCAAACGTGTCACTTCATGCAAATATGTTATTTCAAGCGACATCAATATATGGTTACAAA 1198

a ATTTTAAACATTGGTATATTAATGACTTTCATATATGCATATTAATTTCAATTGAAAGACATCTTGTTGTCCATCATTCCTAACAAGAGGTAGTGTACAAGGCATTATTTAGTTGAGCTC 1318

a TCGCTCACTTTGCATCACAGAAGCATAACCTAGACATAGGTAATTCATTATAGTATCAAAAACTGCAG                                                          1386
```

Figure 1. Structure and nucleotide and amino acid sequences of the PG5a and PG5b genes. A. Shaded bars indicate the coding regions for prolamins. Horizontal arrows indicate the direction of transcription. Abbreviation: E, *Eco*RI; P, *Pst*I; S, *Sal*I. B. Amino acid sequence deduced from the PG5a and PG5b are shown above and under the nucleotide sequences, respectively. 'a' and 'b' on the left side of the nucleotide sequences stand for those of PG5a and PG5b genes, respectively. Identical nucleotide sequences of PG5b to those of PG5a are denoted by dots (·) and only different amino acid residues from those of PG5a are shown. The nucleotide sequence data of PG5a and PG5b will appear in the EMBL, GenBank and DDBJ Databases under the accession numbers D73383 and D73384, respectively. Nucleotide numbers are based on the putative transcriptional initiation site, which is denoted by an asterisk. TATA and CAAT boxes are boxed. Conserved sequences with the 5′-flanking regions of glutelin (Gt1) and albumin (RAG1) genes are overlined.

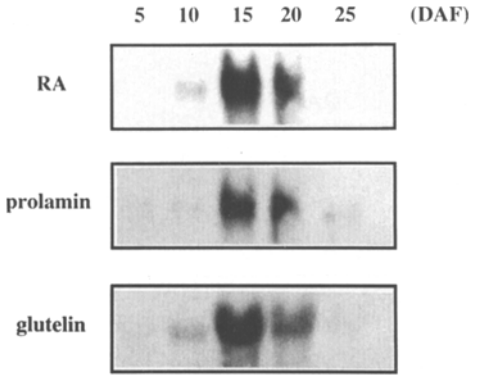
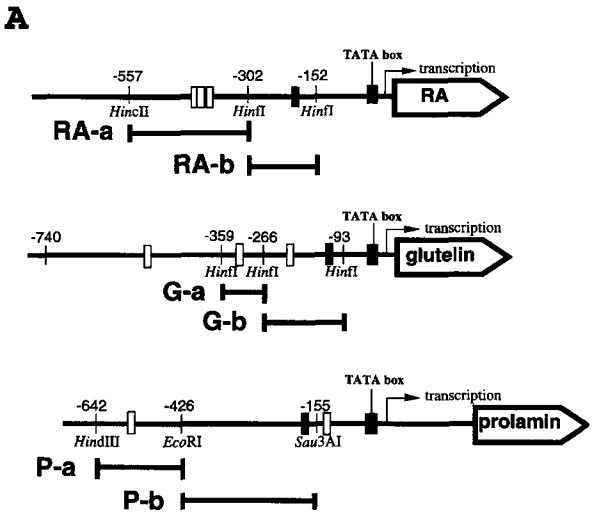


Figure 2. Coordinate expression of RA, glutelin and prolamin genes. Total RNA (10 μg) prepared from rice seeds at 5, 10, 15, 20 and 25 days after flowering (DAF) were used for the RNA blot analysis. The ^{32}P-labeled fragments of RA [15] and prolamin [17] cDNAs were used as probes. In the case of glutelin, the ^{32}P-labeled PCR fragment described in Materials and methods was used as a probe.



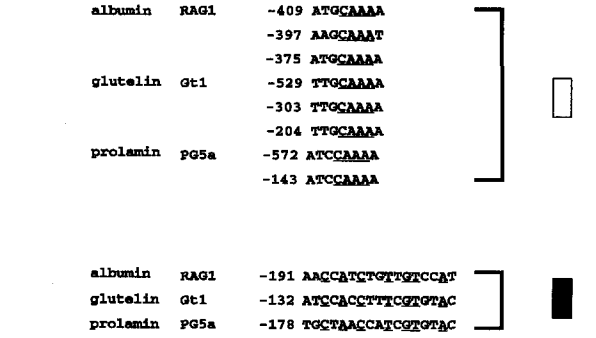


Figure 3. A. Schematic structures of the 5′-flanking regions of RAG1, Gt1 and PG5a genes. The sequences corresponding to B are shown by open box and hatched box. Numbers are based on the transcriptional initiation sites. DNA fragments used for gel shift assays are shown by bars below each gene. The coding regions are denoted by pentagons. Restriction enzyme recognition sites are indicated as follows; E, *Eco*RI; Hc, *Hinc*II; Hd, *Hin*dIII; Hf, *Hin*fI. B. Conserved motifs in the 5′-flanking regions of RAG1, Gt1 and PG5a genes. Numbers are based on the putative transcriptional initiation sites. Identical sequences among three genes are indicated by underlines.

in Fig. 4 showed that partially purified nuclear proteins contain DNA-binding factors that bound to the RA-a and RA-b fragments. In addition, the competition gel shift assay showed that the same nuclear protein binding to the RA-b fragments also bound to the G-b and P-b fragments (lanes 5 and 6 in Fig. 4B).

A comparison of the nucleotide sequences among RAG1, Gt1 and PG5a, led us to find several similar sequences in their 5′ regions, which are summarized in Fig. 3. We designed oligonucleoties as shown in Materials and methods to examine whether or not these conserved sequences are recognition sites for the DNA-binding proteins. Two complementary oligonucleotides, RA-I, containing the sequence similar to the *cis* element of the glutelin gene [34] were synthesized, and subcloned tandemly polymerized fragments were used for gel shift assay. Another fragment, RA-II, containing the conserved sequence (TGCAAAA) among the 5′ region of RAG1, Gt1, and PG5a genes, was prepared in the same way. Both fragments were recognized by the nuclear proteins (lanes 2 in Fig. 5A and 5B), and their complexes were not inhibited by the DNA fragments derived from the multiple cloning site of pUC (lanes 4 in Fig. 5A and 5B). The two sequences recognized by the DNA-binding proteins were localized within the sequences of RA-I and RA-II. Moreover, the DNA/protein complex of the RA-I was inhibited by RA-b, G-b or P-b (lanes 6, 7 and 8 in Fig. 5A, respectively), and that of RA-II was inhibited by RA-a, G-a, G-b or P-a (lanes 6, 7, 8 and 9 in Fig. 5B, respectively).

Together with these two fragments derived from oligonucleotides, GLU-1 containing the previously identified *cis* element for the glutelin *trans*-acting factor [34] was also prepared as described above to analyze whether or not this factor also binds to the 5′ region of RAG1. ^{32}P-labeled GLU-I and G-b were used for the gel shift assay. When labeled GLU-I was used for the assay, the DNA/protein complex was inhibited by the GLU-I and G-b fragments (lanes 3 and 4 in

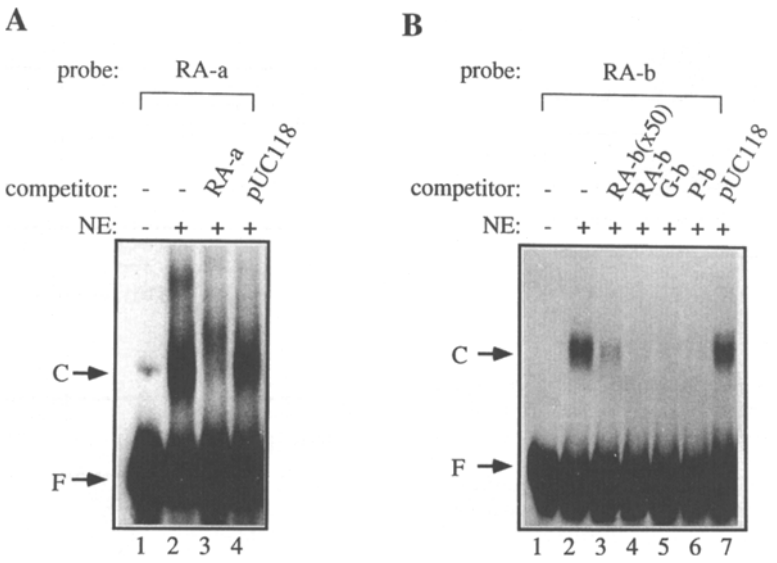


Figure 4. Nuclear protein binding to the 5′ region of RAG1. The DNA/protein complexes and free probes were denoted by C and F, respectively. A. The ^{32}P-labeled fragment, RA-a was incubated without nuclear extract (lane 1) or with nuclear extract of rice seed (lanes 2 to 4). As competitors unlabeled RA-a (lane 3) or pUC118 (lane 4) were incubated before adding probes. B. The ^{32}P-labeled fragment, RA-b was incubated without nuclear extract (lane 1) or with nuclear extract of rice seed (lanes 2 to 7). As competitors unlabeled RA-b (50-fold) (lane 3), RA-b (lane 4), G-b (lane 5), P-b (lane 6) or pUC118 (lane 7) was incubated before adding probes.

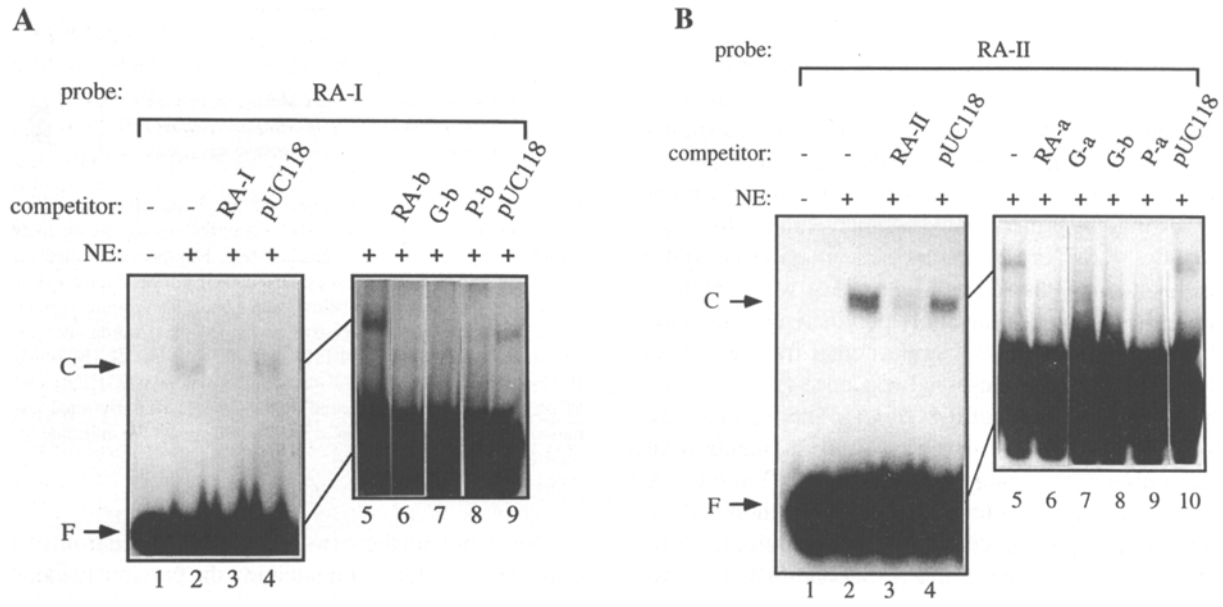


Figure 5. Determination of the nuclear protein-binding sites in the RAG1, Gt1 and PG5a genes. The DNA/protein complexes and free probes were denoted by C and F, respectively. A. The ^{32}P-labeled fragment, RA-I was incubated without nuclear extract (lane 1) or with nuclear extract of rice seed (lanes 2 to 9). As competitors unlabeled RA-I (lane 3), RA-b (lane 5), G-b (lane 7), P-b (lane 8) or pUC118 (lanes 4 and 9) were incubated before adding probes. B. The ^{32}P-labeled fragment, RA-II was incubated without nuclear extract (lane 1) or with nuclear extract of rice seed (lanes 2 to 10). As competitors unlabeled RA-II (lane 3), RA-a (lane 6), G-a (lane 7), G-b (lane 8), P-a (lane 9) or pUC118 (lanes 4 and 10) was incubated before adding probes.

Fig. 6A), indicating that GLU-I contains the nuclear protein binding site [34]. The DNA/protein complex was also inhibited by the G-b and RA-b fragments (lanes 4 and 5 in Fig. 6A), suggesting the RAG1 is also recognized by the same nuclear factors.

The DNA/protein complex of G-b was inhibited by GLU-I (lane 3 in Fig. 6B) and not by pUC (lane 5 in Fig. 6B), showing that this binding is specific as reported previously [34].

When the film was exposed longer, however, a minor band marked as C2 in Fig. 6B was also detected, and this complex was inhibited by RA-II, but not by pUC118 (lanes 8 and 9 in Fig. 6B, respectively). Furthermore, the DNA/protein complex of the labeled RA-II was inhibited by G-b (lane 8 in Fig. 5B). These results are consistent with the fact that G-b contains the conserved sequence (TGCAAAA) in the 5′-flanking region of the RA gene. This conserved sequence also exists in the P-a fragment of the PG5a gene as shown in Fig. 3. The DNA/protein complex of the labeled P-a fragment was observed (lane 2 in Fig. 6C), and this complex was inhibited by RA-II (lane 3 in Fig. 6C). From these results, it can be concluded that the DNA binding factors recognizing the sequence, TGCAAAA, are the same factors for the glutelin and prolamin genes.

Discussion

Storage proteins have been widely studied in seeds and, in the past years, genetic and molecular analyses have been employed to their corresponding genes. The *cis* or *trans* regulatory element of the rice seed storage protein genes, however, have not yet been studied in great detail, except for those of the glutelin gene [34, 37, 40, 41]. Therefore, in this study, DNA-binding proteins for other rice seed protein genes including the 16 kDa albumin, and 13 kDa prolamin were analyzed. We showed that the 5′-flanking regions of the RA (RAG1), type II glutelin (Gt1) and PG5a genes possess at least two nuclear protein-binding sites.

One of these DNA sequences (CTTTCGTGTA) is similar to the promoter sequence of the glutelin gene reported before. This sequence that is located between −130 and −120 in the glutelin gene has been identified as a tissue-specific nuclear protein binding site, in which the nucleotide A at position −120 was shown to be important for the binding of the nuclear protein [34]. The nucleotide A at this position is conserved among RAG1, PG5a (Fig. 3B), and other glutelin and prolamin genes [17, 23]. As shown in Fig. 6B, the nuclear extract prepared in this study also contained the DNA-binding protein which recognizes the same sequence determined by Takaiwa *et al.* [34]. Moreover, we examined whether or not this DNA-binding protein can interact with other storage protein genes such as RA (RAG1) and prolamin (PG5a) genes, because we found that similar recognition sequences exist in their 5′ regions. Indeed GLU-I containing this sequence inhibited DNA/protein complexes of RA (RA-b) and prolamin (P-a), suggesting that the DNA-binding protein for the 5′-flanking regions of the RA and prolamin genes is the same with that of the glutelin gene. Although a sequence (CTTTGGTCTT) resembling the nuclear protein binding site of the glutelin gene is present in the RAG1 promoter region (−65/−56), this sequence was not recognized by this DNA-binding protein (data not shown).

Another DNA-binding protein that recognizes the sequence ATGCAAAA/T might be also involved in the regulation of the glutelin and prolamin gene expression because RA-II containing the sequence, ATGCAAAA, inhibited the DNA/protein complexes of G-b and P-a. Both fragments contain the nuclear protein-binding sites. The other RA gene (RAG2) also contains this sequence existing in two locations of its 5′ region [1]. Furthermore, similar sequences are also contained in the 5′ regions of other cereal storage protein genes [10, 23, 38]. This nuclear protein has lower affinity or is less abundant than the one which recognizes the sequence, CTTTCGTGTA, because only a minor DNA/protein complex of G-b fragment was competed with the RA-II fragment containing the ATGCAAAA sequence (Fig. 6B). The DNA/protein complex of RA-II could not be observed when the nuclear extract from shoot was used instead of that from the maturing seeds (data not shown), suggesting that the ATGCAAAA sequence is recognized by a nuclear protein specific for maturing seed. Moreover, this sequence shows some similarity to the consensus sequence, TG(T/A/C)AAA(G/T), which has been shown to be essential for endosperm-specific expression of the cereal prolamin genes in transgenic plants [6, 20], and this resemblance is agreeable to the proposed function of this sequence as an enhancer-like element in the pea legumin gene [29]. These sequences have been referred as the −300 element which is conserved in cereal prolamin gene promoters [7, 24, 39]. In addition, although the sequence was referred to as the CCAAT box, nuclear proteins (high mobility group proteins) from maize endosperm have been shown to interact with the sequence, TTGCAAAT, of the 19 kDa zein gene in maize [12]. This sequence is highly homo-

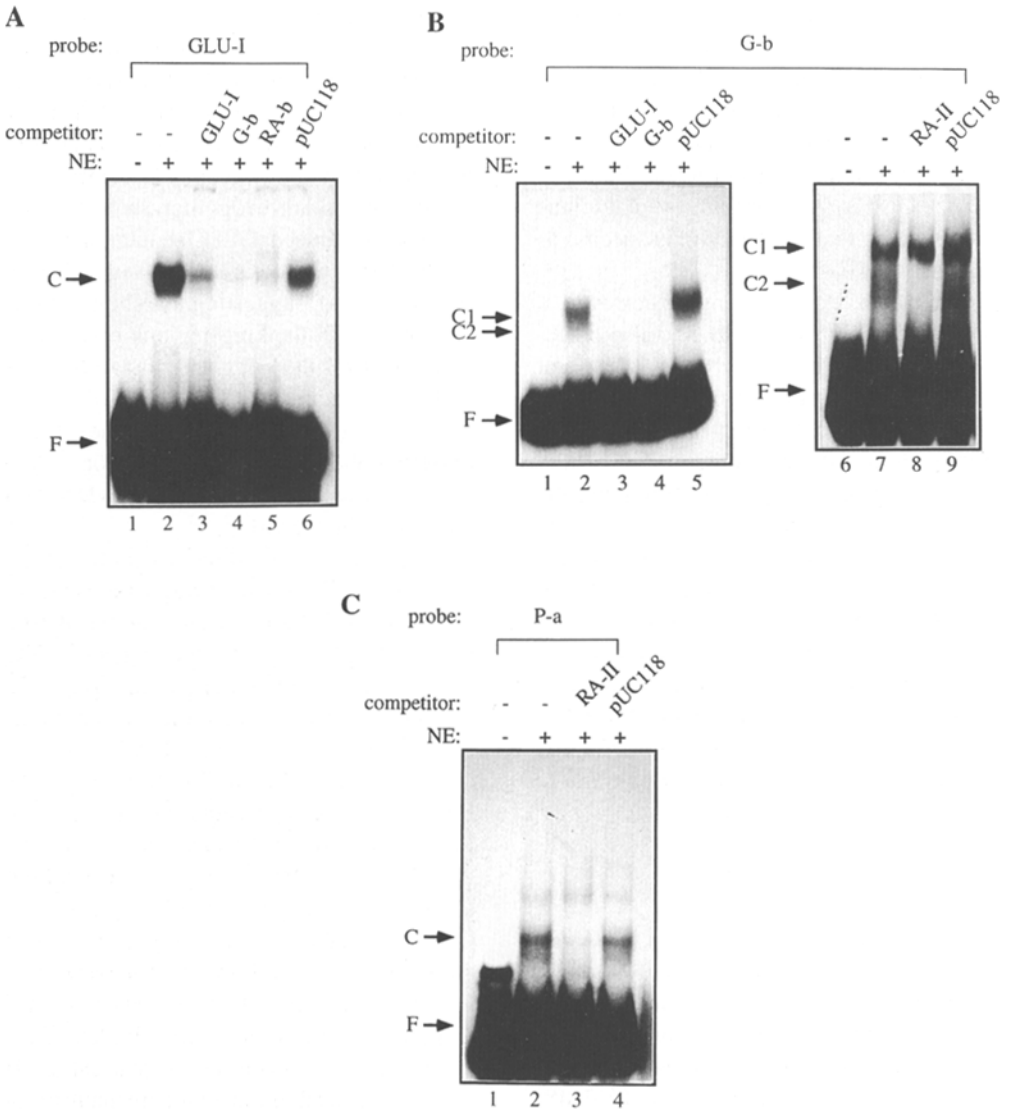


Figure 6. The DNA-binding proteins are shared among the RAG1, Gt1 and PG5a gene promoters. The DNA/protein complexes and free probes were denoted by C and F, respectively. A. Nuclear protein binding to GLU-I. The ^{32}P-labeled fragment, GLU-I was incubated without nuclear extract (lane 1) or with nuclear extract of rice seed (lanes 2 to 6). As competitors, unlabeled GLU-I (lane 3), G-b (lane 4), RA-b (lane 5) and pUC118 (lane 6) were used. B. Nuclear protein binding to G-b. The ^{32}P-labeled fragment, G-b was incubated without nuclear extract (lanes 1 and 6) or with nuclear extract of rice seed (lanes 2–5 and 7–9). As competitors, unlabeled GLU-I (lane 3), G-b (lane 4), RA-II (lane 8) and pUC118 (lanes 5 and 9) were pre-incubated with probes. C. Nuclear protein binding to P-a. The ^{32}P-labeled fragment, P-a was incubated without nuclear extract (lane 1) or with nuclear extract of rice seed (lanes 2 to 4). As competitors, unlabeled RA-II (lane 3) and pUC118 (lane 4) were used.

logous to the sequence that was identified as nuclear protein binding site in the 5′-flanking region of the RAG1. Thus these homologous sequences are present in the promoter of cereal storage protein gene, and the conservation among them suggests that the homologous sequence plays an important role for gene expression.

Glutelin is the most dominant protein (about 80%) in total rice seed proteins while albumin and prolamin are comprised in much lesser amounts. However, the copy number of the glutelin gene, suggested to be 5 to 8 [23], is less than those of the albumin and prolamin genes [1, 4, 17]. Considering this fact, glutelin gene should be expressed more efficiently than the albu-

min and prolamin genes. Hence, although these three genes share the same DNA-binding protein, the DNA sequence of the glutelin gene(s) might have stronger affinity to the DNA-binding protein than that of the RA and prolamin genes. The binding sequences of these three genes are not completely identical, so that probably they show different affinity to the DNA-binding protein. Another possibility is that there is an additional control element which elucidates the expression of the glutelin gene, for example, the AACA motif (AAACAACTCTATC), the conserved sequence common to all glutelin genes [36] but not present in RAG1 and PG5a promoters appears to be the glutelin-specific *cis* element [37, 40, 41]. Furthermore, despite the use of heterologous system, the region between -411 and -237 has been identified as an essential region for the expression of the glutelin gene [35]. From this result, the DNA-binding protein which recognizes the region between -130 and -120 might be not enough for the full promoter activity.

Recently we have isolated a genomic clone for α-globulin [22] in which the *cis* sequences identified in this study as the DNA-binding protein recognition sites do not exist. Although α-globulin is one of the major seed proteins in rice, its genetic organization is different from those of glutelin, prolamin and albumin, i.e., the globulin gene is a single copy whereas each of other three genes consists of a multigene family.

Altogether, we conclude that at least some of the transcriptional machinery is shared among RA, glutelin and prolamin genes. Thus it seems difficult to discriminate the expression of RA gene from those of the other storage protein genes. Judging from these results, antisense construct or gene targeting technique might be useful to generate a hypoallergenic rice. We have already examined the quantities of RA in cultivated and wild rice species wherein most of them contain certain amounts of albumins except for the wild species with poor endosperm [2]. However, even in glutelin genes, the expression of Gt3 does not coordinate with the expression of Gt1 and Gt2, showing that they are regulated differently [23]. Based on these results, there might still be certain differences in the transcriptional regulation between RA and other storage proteins. Further analysis will be necessary to clarify the gene expression mechanism for the rice seed storage proteins.

Acknowledgments

We are grateful to Dr K. Yamazaki (Nagoya University) for valuable discussion. We would like to thank Dr K. Tanaka (Kyoto Prefectural University) and Dr T. Tashiro (Nagoya University) for providing a rice genomic library and rice seeds, respectively. We also thank to Dr A. M. Alvarez (International Rice Research Institute) for the assistance in the preparation of the manuscript. This research was partly supported by Grants-in Aid for Scientific Research, Ministry of Education, Science and Culture of Japan.

References

1. Adachi T, Izumi H, Yamada T, Tanaka K, Takeuchi S, Nakamura R, Matsuda T: Gene structure and expression of rice allergenic proteins belonging to the α-amylase/trypsin inhibitor family. Plant Mol Biol 21: 239–248 (1993).
2. Adachi T, Alvarez AM, Aoki N, Nakamura R, Garcia VV, Matsuda T: Screening of rice strains deficient in 16-kDa allergenic protein. Biosci Biotechnol Biochem 59: 1377–1378 (1995).
3. Allen RD, Bernier F, Lessard PA, Beachy RN: Nuclear factors interact with a soybean β-conglycinin enhancer. Plant Cell 1: 623–631 (1989).
4. Alvarez AM, Adachi T, Nakase M, Aoki N, Nakamura R, Matsuda T: Classification of rice allergenic protein cDNAs belonging to the α-amylase/trypsin inhibitor gene family. Biochim Biophys Acta 1251: 201–204 (1995).
5. Barber D, Sanchez-Monge R, Gomez L, Carpizo J, Armentia A, Lopez-Otin C, Salcedo G: A barley flour of insect α-amylase is a major allergen associated with baker's asthma disease. FEBS Lett 248: 119–122 (1989).
6. Colot V, Robert LS, Kavanagh TA, Beven MW, Tompson RD: Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J 6: 3559–3564 (1987).
7. Forde BG, Heyworth A, Pywell J, Kreis M: 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. Nucl Acids Res 13: 7327–7339 (1985).
8. Foster R, Izawa T, Chua NH: Plant bZIP proteins gather at ACGT elements. FASEB J 8: 192–200 (1994).
9. Garcia-Olmedo F, Salcedo G, Sanchez-Monge R, Gomez L, Royo J, Carbonero P: Plant, proteinaceous inhibitors of proteinases and α-amylases. Oxford Surv Plant Cell Biol 4: 275–334 (1987).
10. Gioviazzo G, Manzocchi LA, Bianchi MW, Coraggio I, Viotti A: Functional analysis of the regulatory region of a zein gene in transiently tranformed protoplasts. Plant Mol Biol 19: 257–263 (1992).
11. Gomez L, Martin E, Hernandez D, Sanchez-Monge R, Barber D, Pozo V, Andres B, Armentina A, Lahoz C, Salcedo G, Palomino P: Members of the α-amylase inhibitors family from wheat endosperm are major allergens associated with baker's asthema. FEBS Lett 261: 85–88 (1990).

12. Grasser KD, Maier UG, Haass MM, Feix G: Maize high mobility group proteins bind to CCAAT and TATA boxes of a zein gene promoter. J Biol Chem 265: 4185–4188 (1990).
13. Hartings H, Maddaloni M, Lazzaroni N, Di Fonzo N, Motto M, Salamini F, Thompson R: The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J 8: 2795–2801 (1989).
14. Izawa T, Foster R, Nakajima M, Shimamoto K, Chua NH: The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. Plant Cell 6: 1277-1287 (1994).
15. Izumi H, Adachi T, Fujii N, Matsuda T, Nakamura T, Tanaka K, Urisu A, Kurosawa Y: Nucleotide sequence of a cDNA clone encoding a major allergenic protein in rice seeds: homology of the deduced amino acid sequence with members of α-amylase/trypsin inhibitor family. FEBS Lett 302: 213–216 (1992).
16. Kim SY, Wu R: Multiple protein factors bind to a rice glutelin promoter region. Nucl Acids Res 18: 6845-6852 (1990).
17. Kim WT, Okita TW: Structure, expression, and heterogeneity of the rice seed prolamines. Plant Physiol 88: 649–655 (1988).
18. Kreis M, Shery PR, Forde BG, Milfin BJ: Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surv Plant Mol Cell Biol 2: 253–317 (1985).
19. Limas GG, Salinas M, Moneo I, Fischer S, Wittman-Liebold B, Hendez E: Purification and characterization of ten new rice NaCl-soluble proteins: identification of four protein-synthesis inhibitors and two immunoglobulin-binding proteins. Planta 181: 1–9 (1990).
20. Maier UG, Brown JWS, Toloczyki C, Feix G: Binding of a nuclear factor to a consensus sequence in the 5′ flanking region of zein genes from maize. EMBO J 6: 17–22 (1987).
21. Matsuda T, Nomura R, Sugiyama M, Nakamura R: Immunochemical studies on rice allergenic proteins. Agric Biol Chem 55: 509–513 (1991).
22. Nakase M, Hotta H, Adachi T, Aoki N, Nakamura R, Masumura T, Tanaka K, Matsuda T: Cloning of the rice seed α-globulin-encoding gene: sequence similarity of the 5′-flanking region to those of the genes encoding wheat high-molecular-weight glutenin and barley D hordein. Gene 170: 223–226 (1996).
23. Okita TW, Hwang YS, Hnilo J, Kim WT, Aryan AP, Larson R, Krishnan HB: Structure and expression of the rice glutelin multigene family. J Biol Chem 264: 12573–12581 (1989).
24. Quayle T, Feix G: Functional analysis of the −300 region of maize zein genes. Mol Gen Genet 231: 369–374 (1992).
25. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
26. Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G: Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. Plant Cell 4: 689–700 (1992).
27. Sharief FS, Li SS: Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. J Biol Chem 257: 14753–14759 (1982).
28. Shibasaki M, Suzuki S, Nemoto H, Kuroume T: Allergenicity and lymphocyte-stimulating property of rice protein. J Allergy Clin Immunol 64: 259–265 (1977).
29. Shirsat A, Wilford N, Croy R, Boulter D: Sequence responsible for the tissue specific promoter activity of pea legumin gene in tobacco. Mol Gen Genet 215: 326–331 (1989).
30. Shorrosh BS, Wen L, Zen K, Huang J, Pan JS, Hermodson MA, Tanaka K, Muthukrishnan M, Reeck GR: A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice. Plant Mol Biol 18: 151–154 (1992).
31. Spies JR: Allergens. J Agric Food Chem 22: 30–36 (1974).
32. Sugiyama T, Rafalski A, Peterson D, Soll D: A wheat HMW glutenin subunit gene reveals a highly repeated structure. Nucl Acids Res 13: 8729–8737 (1985).
33. Takaiwa F, Ebinuma H, Kikuchi S, Oono K: Nucleotide sequence of a rice glutelin gene. FEBS Lett 221: 43–47 (1987).
34. Takaiwa F, Oono K: Interaction of an immature seed-specific *trans* acting factor with the 5′ upstream region. Mol Gen Genet 224: 289–293 (1990).
35. Takaiwa F, Oono K, Kato A: Analysis of the 5′ flanking region responsible for the endosperm-specific expression of a rice glutelin chimeric gene in transgenic tobacco. Plant Mol Biol 16: 49–58 (1991).
36. Takaiwa F, Oono K, Wing D, Kato A: Sequence of three members and expression of a new major subfamily of glutelin genes from rice. Plant Mol Biol 17: 875–885 (1991).
37. Takaiwa F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K: Characterization of common *cis*-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Mol Biol, in press 1996.
38. Thompson GA, Boston RS, Lyznik LA, Hodges TK, Larkins BA: Analysis of promoter activity from an α-zein gene 5′ flanking sequence in transient expression assays. Plant Mol Biol 15: 755–764 (1990).
39. Ueda T, Wang Z, Pam N, Messing J: Identification of a transcriptional activator-binding element in the 27-kilodalton zein promoter, the −300 element. Mol Cell Biol 14: 4350–4359 (1994).
40. Yoshihara T, Takaiwa F: *cis*-Regulatory elements responsible for quantitative regulation of the rice seed storage protein glutelin *GluA-3* gene. Plant Cell Physiol 37: 107–111 (1996).
41. Yoshihara T, Washida H, Takaiwa F: A 45-bp proximal region containing AACA and GCN4 motifs sufficient to confer endosperm-specific expression of the rice seed storage protein glutelin gene, *GluA-3*. FEBS Lett, in press (1996).
42. Zhao Y, Leisy DJ, Okita T: Tissue-specific expression and temporal regulation of the rice glutelin Gt3 gene are conferred by at least two spatially separated *cis* regulatory elements. Plant Mol Biol 25: 429–436 (1994).
43. Zheng Z, Kawagoe Y, Xiao S, Li Z, Okita T, Hau TL, Lin A, Murai N: 5′ distal and proximal *cis* acting regulator elements are required for developmental control of a rice seed storage protein glutelin gene. Plant J 4: 357–366 (1993).