Multiple protein factors bind to a rice glutelin promoter region

Soo Young Kim and Ray Wu*
Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received September 14, 1990; Revised and Accepted October 31, 1990 EMBL accession no. X52153

ABSTRACT

Our goal is to identify *cis*-acting elements in the regulatory region of the major seed storage protein gene in rice. A glutelin gene (pGL5-1) has been cloned by screening a rice genomic DNA library with synthetic oligonucleotides and with an amplified DNA fragment. A transient expression assay using immature rice seeds shows that its 5′ flanking sequence can direct the synthesis of β-glucuronidase (GUS) when fused upstream of the GUS coding region. Gel-retardation assays were performed to study protein-DNA interactions between putative regulatory sequences of pGL5-1 and nuclear proteins from immature rice seeds. We demonstrate that at least six protein-DNA complexes are formed between the 5′ flanking sequence of pGL5-1 (−677 to −45) and nuclear protein factors. By subsequent DNase I-footprinting analyses we defined several protein-binding regions. Two of the protein−binding sequences contain the TGAGTCA motif, which is also present in the −300 element found in the 5′ flanking sequences of several storage protein genes of other crop plants, and to which the transcription factors jun and GCN4 bind.

INTRODUCTION

Glutelin is the major seed storage protein of rice, constituting 80% of the total endosperm protein of rice seeds (1). The mature glutelin protein is composed of two subunits, one acidic and one basic, that are generated by the posttranslational cleavage of a large precursor molecule (1−3). While its synthesis occurs at the rough endoplasmic reticulum, the glutelin appears to be processed through Golgi apparatus, and the mature protein is deposited in type II protein bodies (4). Since rice is one of the major cereal crops, the study of rice glutelin gene expression is important not only in understanding the basic mechanism of tissue-specific gene expression but also in improving the nutritional value of the protein in this major food source.

cDNA (5−11) and genomic DNA (7, 12) sequences of rice glutelin genes have been previously reported. The genes can be classified into several subfamilies according to their DNA sequences, and each subfamily is composed of several copies (7, 8). The Gt1 and Gt2 subfamilies are closely related and show more than 87% DNA sequence identity to each other, whereas the Gt3 subfamily shows significant divergence from the other two subfamilies (7). Analyses of cDNA clones show that the deduced amino acid sequence of the protein has 30−35% identity to those of leguminous 11S globulins, such as pea legumin and soybean glycinin (5−7).

Very little is known about the regulation of glutelin gene expression, except that mRNA levels are regulated differentially during seed development (7). In a recent study Leisy et al. introduced a chimeric gene composed of Gt3 promoter and the chloramphenicol acetyltransferase gene into tobacco (13). Their results show that 980 bp of the 5′ flanking sequence of Gt3 is sufficient to give the highest level of expression in seeds of transgenic tobacco, even though expression in other tissues was also observed in some transgenic plants.

In this paper we describe the cloning of a rice glutelin gene that appears to belong to the Gt2 subfamily. We set out to identify 5′ regulatory sequences of this gene by gel retardation and DNase I-footprinting assays, and show that several proteins bind to the 5′ flanking region of this cloned gene.

MATERIALS AND METHODS

DNA manipulation

Preparation of plasmid and phage DNAs and other manipulations of DNA were performed according to standard procedures (14).

Screening of a rice genomic DNA library

A rice (*Oryza sativa* L. cv. Labelle) genomic DNA library, which was constructed by ligating HindIII-digested rice genomic DNA into lambda 2001, was provided by R. Okagaki and S. Wessler. For the primary and secondary screenings, hybridization was performed with a ^{32}P-labeled synthetic oligonucleotide, glu-1 (5′-ACATCAAAAGATCCACCGT-3′), in 6× SSC (1× SSC: 0.15 M NaCl, 0.015 M sodium citrate), 5× Denhardt's solution, 0.05% sodium pyrophosphate, 0.5% sodium dodecylsulfate (SDS), and 50 μg/ml tRNA, at 45°C. Filters were washed first at room temperature for several hours and then at 45°C for 10 min in 2× SSC, 0.1% SDS. For the tertiary screening, duplicate filters were prepared and one set was screened with the glu-1 probe and the other set with another ^{32}P-labeled synthetic oligonucleotide, glu-2 (5′-TTGCCTAGGGCTAAGTTGA-3′). Final rounds of screening were performed with the glu-2 probe,

* To whom correspondence should be addressed

and with a nick-translated polymerase chain reaction (PCR) amplified fragment (pAGL) of the glutelin gene.

Polymerase chain reaction

Rice genomic DNA (11 μg) and 100 pmoles of primers, glu-1 and glu-2, were mixed in 100 μl of reaction volume containing 1.5 mM dNTPs, 10 mM Tris/Cl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, and 10 mM dithiothreitol (DTT). The mixture was boiled for 2−3 min, chilled on ice for 1 min, and incubated at 37° or 45°C to allow the primers to anneal to the genomic DNA template. After 2−4 min, Klenow fragment of DNA polymerase (1 unit) was added to start the primer extension reaction and incubation was continued for 4 min. The cycle of boiling, chilling on ice, annealing, and primer extension reaction was repeated 30 times. After completing the 30th cycle, DNA was isolated by phenol/chloroform extraction and ethanol precipitation. The amplified DNA was recovered from a low melting point agarose gel, cloned into the SmaI site of M13 by blunt end ligation, and then identified by sequencing. The resulting glutelin clone is referred to as pAGL.

DNA sequence analysis

DNA sequencing was performed according to the dideoxynucleotide chain termination method (15), using either M13 or double-stranded plasmid DNA.

Transient expression assay

Rice seeds were collected 2 weeks after flowering. The seed coat and the pericarp were removed by hand using a sharp-pointed forceps. The seeds were then placed on a wet filter paper in a petri dish, and particle-gun bombardment was performed according to Klein et al. (16). The seeds were bombarded twice, once on each side. To measure β-glucuronidse (GUS) activity the GUS substrate solution (0.1 M sodium phosphate, pH 7.0, 30% Ficoll, 5 mM $K_4Fe(CN)_6$, 5 mM $K_3Fe(CN)_6$, 0.005% Triton X-100, 0.3% X-Gluc) was placed on each seed 2 days after bombardment, and blue spots were counted 10 days later.

Preparation of nuclear extract for protein binding assays

Sixty grams of 10 to 14 day-old immature seeds was ground with a mortar and pestle in 200 ml of nuclei isolation buffer (0.4 M mannitol, 10 mM HEPES, pH 7.6, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% bovine serum albumin, 1% polyvinylpyrrolidone-40, 5 mM β-mercaptoethanol) and filtered through cheese cloth. The filtrate was centrifuged for 10 min at 200 g to isolate nuclei, and the pellet was resuspended in 50 ml of nuclei lysis buffer [15 mM HEPES, pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 1 mM phenylmethylsulfonylfluoride (PMSF)]. The nuclei were lysed by adding 1/10 volume of 4 M $(NH_4)_2SO_4$ and gently shaking for 30 min. The lysate was then cleared by centrifugation at 40,000 rpm for 1 hour in a Ti 50 rotor, and the supernatant was fractionated with solid ammonium sulfate. The proteins precipitated between 30−75% saturation were resuspended in 5 ml of nuclei extraction buffer (25 mM HEPES, pH 7.6, 100 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 mM DTT, 1 mM PMSF) and dialysed overnight against 1 liter of the same extraction buffer. Following dialysis, the extract was adjusted to 0.3 M KCl and loaded onto a pre-equilibrated heparin-agarose column (0.4 ml) and eluted with 0.6 M KCl. The protein concentration of each fraction was determined according to the method of Bradford (17). Peak fractions were pooled and dialysed against 500 ml of nuclei extraction buffer for 2 hours. Finally, the extract was divided into small aliquots, quick-frozen in liquid nitrogen, and stored at −70°C.

Preparation of DNA fragments for protein binding assays

To prepare DNA fragments for the binding assays, the 1.3 kb SpeI fragment from the glutelin gene, pGL5-1, was cloned into the XbaI site of pUC13 and then digested with NsiI. The resulting 180 bp and 450 bp fragments were recloned into the PstI site of pUC13. In later experiments the 450 bp fragment was further digested with RsaI to give 290 bp and 160 bp fragments, and cloned into Bluescript II KS (+), using HindIII/SmaI and SalI/EcoRV sites, respectively.

For gel-retardation assays the DNA fragments were cut out from the plasmid DNAs with HindIII/SalI (to give the 180 bp fragment) or with BamHI/SalI (to give the 290 and the 160 bp fragments), purified from low melting point agarose gels, phenol extracted, and then labeled with ^{32}P-dTTP using the Klenow fragment of DNA polymerase I. For DNase I-footprinting analyses plasmid DNA containing either the 180 bp fragment or the 160 bp fragment was digested with an appropriate restriction enzyme, labeled with three or four ^{32}P-dNTPs using the Klenow fragment, ethanol-precipitated, and digested with another restriction enzyme to excise the single-end labeled fragment from the vector. The mixture was then electrophoresed on a 5% polyacrylamide gel and the desired DNA fragment was eluted from a gel slice containing the DNA fragment.

For the preparation of the DNA fragments containing the synthetic oligonucleotides GT1-14 or GT3-13, each monomeric unit was self-ligated, and ligation products were separated on a 10% polyacrylamide gel. The 3-, 4- or 5-mers of the oligonucleotides were separately eluted from the gel, cloned into the BamHI (GT1-14) or into the EcoRV site (GT3-13) of the Bluescript II KS (+) and then sequenced. The clones which had the GT1-14 or the GT3-13 repeat units in the tandem head-to-tail orientation were selected for the binding study.

Gel-retardation assay

The labeled DNA fragment (1−3 ng) was mixed with 3 μl of purified nuclear extract in 20 μl of reaction volume containing 1−4 μg of poly[d(I−C)], 10% glycerol, 50 to 100 mM KCl, and 2 μl of 10× binding buffer (0.25 M HEPES, pH 7.6, 5 mM EDTA, 10 mM DTT, 50 mM $MgCl_2$). The mixture was incubated for 15 min at room temperature to allow binding to occur. Then 2 μl of loading dye (0.1% bromophenol blue and 0.1% xylene cyanol in binding buffer with 10% glycerol) was added and the mixture was immediately loaded onto a 5% polyacrylamide (acrylamide/bisacrylamide, 29:1) gel. The gel was pre-run for at least 1 hr, and electrophoresis was performed at 200 V in 0.5× TBE (45 mM Tris base, 45 mM boric acid, pH 8.3−8.5, 1 mM EDTA). Following electrophoresis, gels were dried for 30 min and autoradiographed.

DNase I-footprinting

To 15 μl of nuclear extract 2 μl (1 μg/μl) of poly d[(I-C)] and 1 μl of $MgCl_2/CaCl_2$ (100 mM and 20 mM, respectively) were added and the mixture was placed on ice. After 20 min the mixture was transferred to room temperature, 1−5 ng of labeled DNA was added, and incubation continued for another 20 min. DNase I digestion was started by adding 1 μl (0.025 to 0.1 units) of the enzyme and stopped by adding 80 μl of stop buffer (15 mM EDTA, 0.15% SDS, 0.1 M NaCl, and 70 μg/ml tRNA) after 15 seconds to 1 min. The optimal amount of DNase I to

get relatively uniform digestion pattern was empirically determined for each reaction mixture. Proteins were removed from the mixture by phenol/chloroform extraction, followed by precipitation of DNA with 2.5 volumes of ethanol and electrophoresis on a sequencing gel (7% polyacrylamide/7 M urea). The gel was dried for 1 hr at 80°C and autoradiographed.

RESULTS

Cloning of a rice glutelin gene

As a first step toward studying the expression of rice glutelin genes, we cloned one of the glutelin genes by screening a rice (*Oryza sativa L.* cv. Labelle) genomic DNA library. we employed the available cDNA sequence information (5) to prepare DNA probes. Two primers, named glu-1 and glu-2, were selected based on their sequence uniqueness and chemically synthesized. The two primers are each 19 bases in length and are 151 bp apart from each other. Using these primers and total rice DNA as the template, we synthesized a 151 bp DNA fragment using the polymerase chain reaction (18). The PCR amplified product was cloned in M13 and was named pAGL.

A rice genomic library was first screened with the glu-1 primer. Thirty four positive plaques were obtained, 22 of which survived the secondary and tertiary screenings where both the glu-1 and the glu-2 were used separately as probes. Final screening of the 22 clones was carried out using the PCR amplified fragment, pAGL. Ten positive clones were obtained. Further analysis showed that six of them had the same 3.6 kb HindIII insert.

We chose one of the six clones, pGL5-1, for further analysis and characterized it by restriction mapping and DNA sequencing. Subsequent alignment of the DNA sequence of pGL5-1 with the published glutelin cDNA sequence (5), and later with a glutelin genomic DNA sequence (pREE771, 12), indicated that the cloned HindIII fragment contains 2.2 kb of the 5′ flanking sequence and 1.3 kb of the coding region of glutelin (Figure 1a). The last HindIII site occurs in the middle of the third intron of the glutelin gene. Thus, a part of the intron and the last exon of the glutelin gene is missing in this clone. The first three exons of pGL5-1 and pREE771 show 99% DNA sequence identity, but the two genes show significant divergence in their 5′ flanking sequences. Recently, several other genomic DNA sequences have been reported (7). Comparison of restriction maps and DNA sequences suggests that pGL5-1 is closely related to the Gt2 subfamily of the glutelin gene family. pGL5-1 shows 98% DNA sequence identity to Gt2, both in the coding region and in the 5′ flanking sequence. Both clones have several insertions in their 5′ flanking sequences compared to Gt1 and Gt3. However, as shown in Figure 1b, where a partial DNA sequence of pGL5-1 is presented, we find that one of the insertions in pGL5-1 is 100 bp longer (from positions −349 to −250) than the corresponding one in Gt2. The long direct repeat found in the 5′ flanking sequences of pREE771 and Gt2 is also present in pGL5-1 (LR-1, LR-2).

Promoter activity of the pGL5-1

The high degree of sequence identity between pGL5-1 and Gt2 strongly suggests that pGL5-1 is also an active gene, and thus, its 5′ flanking sequence is expected to have promoter activity in immature rice seeds. To ensure that the putative promoter of pGL5-1 is active we did a transient expression assay. First, we constructed two expression vectors, pGLG1 and pGLG2 (Figure 2). Both vectors are composed of 2.2 kb of the 5′ flanking sequence of pGL5-1 and the GUS gene as a reporter. In the construct pGLG1 the glutelin sequence ends in the middle of the transcribed, but untranslated region (position +24). In pGLG2 the glutelin sequence contains the first 31 amino acid residues (position +129) including the putative signal peptide (6, 7). Thus,

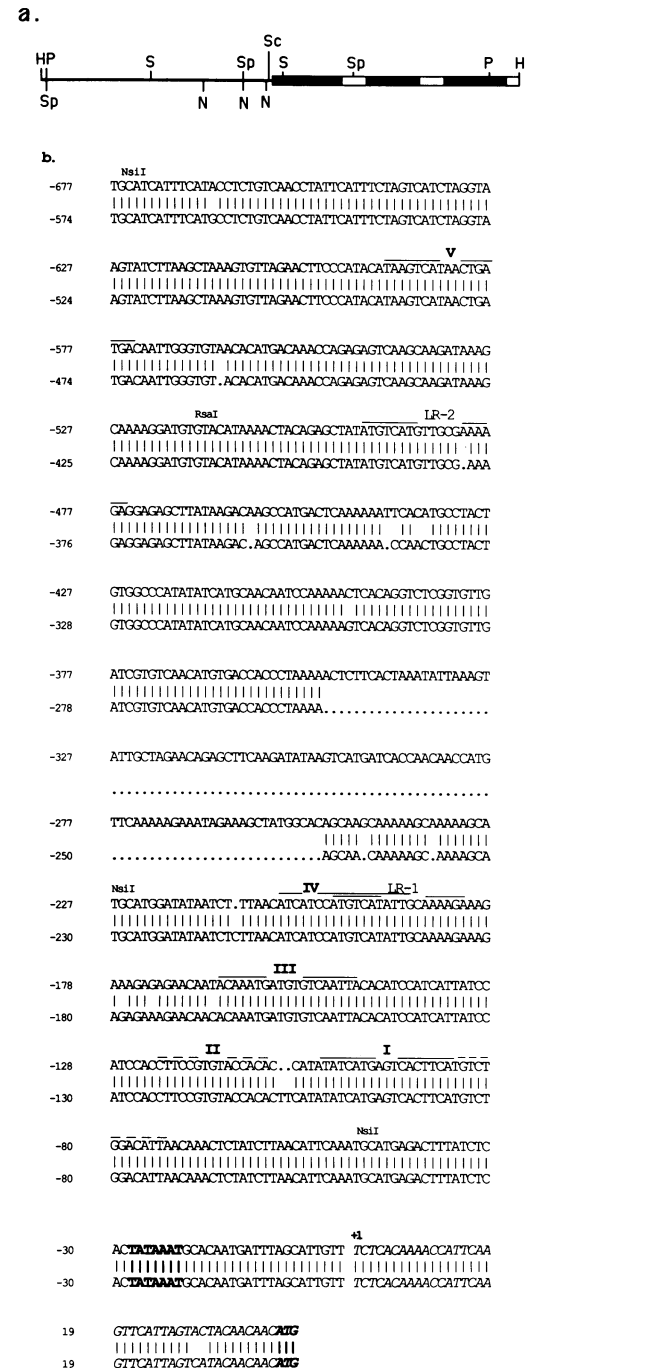

Figure 1. Restriction map and nucleotide sequence of pGL5-1. (**a**) Restriction map. Thin line, filled bar, and empty bar indicate 5′ flanking sequence, exon, and intron, respectively. H, HindIII; S, SpeI; Sp, SphI; Sc, ScaI; P, PstI; N, NsiI. (**b**) Partial DNA sequence of pGL5-1. The sequence from the translation initiation codon (ATG) to −677 from the transcription start site (+1) is presented. The transcription start site is from Okita et al. (7). Upper lines (in each set of two), sequence of pGL5-1; lower lines, corresponding sequence of Gt2. Identical nucleotides between pGL5-1 and Gt2 are shown with short vertical lines. Dots represent missing nucleotides. LR-1 and LR-2 are long direct repeats. TATA box and initiation codon are shown by bold letters. Roman numerals indicate the position of protein binding regions. Relevant restriction sites are also shown.

pGLG1 is a transcriptional fusion vector, whereas pGLG2 is a translational fusion vector. These constructs were then introduced into immature rice seeds by the biolistic method (16) and the GUS activity was examined by counting the number of blue spots.

We bombarded 12 seeds with pGLG1 and 14 seeds with pGLG2 in one experiment, and 67 blue spots were observed in the seeds bombrded with pGLG1. No blue spots were detected with pGLG2. In another experiment 48 blue spots were obtained from 14 seeds bombarded with pGLG1, and 3 blue spots from 18 seeds bombarded with pGLG2. In both experiments we did not get any blue spots either in unbombarded control seeds or in other tissues (root and leaf) bombarded with pGLG1 or pGLG2. Although the number of blue spots may not be considered to quantitatively represent GUS activity, two conclusions can be drawn from this result: a), the 5′ flanking sequence of pGL5-1 has promoter activity as expected, and b), the expression of the glutelin gene seems to be tissue-specific because it is expressed in immature seeds but not in roots and leaves. We do not know why the translational fusion vector pGLG2 gives fewer blue spots. It is possible that glycosylation of GUS resulted in the loss of the GUS activity (19), because the glutelin sequence in pGLG2 includes the signal peptide.

Binding of protein factors to the conserved region at the 5′ end of a rice glutelin gene

In spite of sequence divergence in further upstream regions, pGL5-1 and pREE771 show a high degree of sequence identity in the 5′ flanking sequences, between position −213 and the transcription start site (+1), which suggests that this region may play an important role in the regulation of glutelin gene expression. We probed this possibility by investigating protein-DNA interactions between nuclear proteins and this highly conserved region. A DNA fragment which includes the conserved region (fragment A in Figure 3a) was incubated with partially purified nuclear extracts prepared from immature rice seeds. Protein-DNA complex formation was then studied by gel-retardation (20, 21) and DNase I-footprinting (22) assays.

As shown in Figure 3b, four major retarded bands are observed in gel-retardation assays using the fragment A, indicating that four protein-DNA complexes (named B1, B2, B3 and B4) are formed (lane 2). The retarded bands are formed only after adding nuclear extract and appropriate amount of a non-specific competitor DNA, poly[d(I−C)] (lane 2). Addition of proteinase K to the DNA-extract mixture abolishes the formation of the bands (lane 7), showing that complexes B1, B2, B3 and B4 are

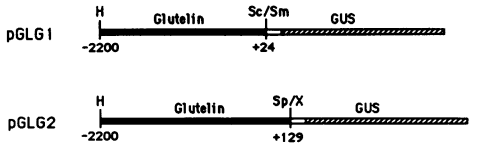

Figure 2. Expression vectors used in the study of promoter activity. The two expression vectors used in the transient expression assay are shown schematically. pGLG1 was constructed by replacing the 35S promoter-containing HindIII/SmaI fragment of pBI221 (from Clontech) with the HindIII/ScaI fragment (−2200 to +24) of pGL5-1. In pGLG2, the 35S promoter-containing HindIII/XbaI fragment of pBI221 was replaced by the larger HindIII/SpeI fragment (−2200 to +129) of pGL5-1. Filled bar, untranscribed sequence of pGL5-1; hatched bar, coding region of the GUS gene; stippled bar, transcribed sequence of pGL5-1; empty bar, sequence generated during construction. The translation initiation codon of pGL5-1 starts at nucleotide position +40. H, HindIII; Sc, ScaI; Sm, SmaI; Sp, SpeI; X, XbaI.

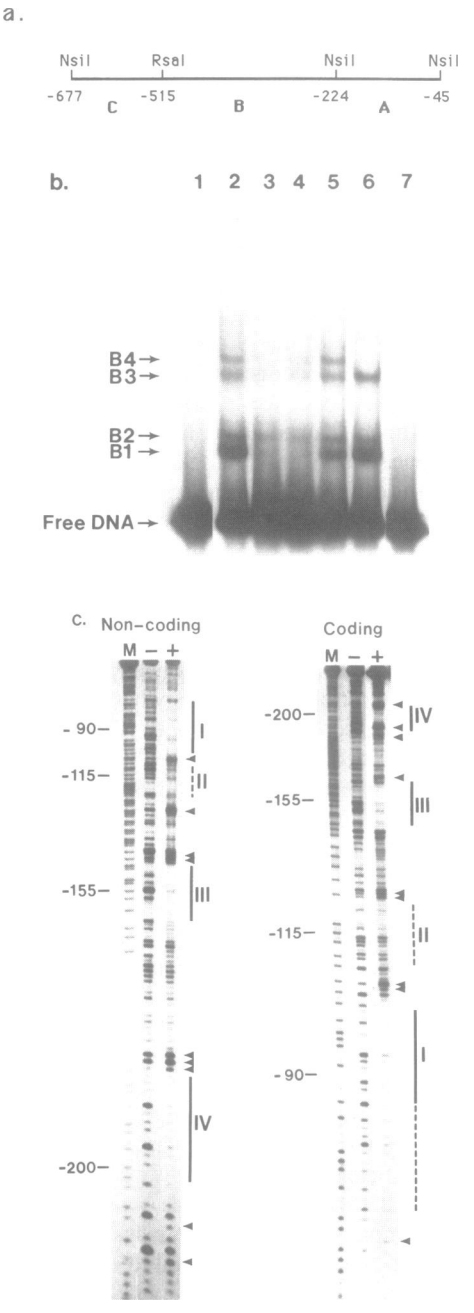

Figure 3. Binding of protein factors to the promoter region of pGL5-1. **(a)** Restriction map of the promoter fragments used in the binding studies. Numbers indicate nucleotide position from the transcription start site (taken as +1). Fragment A contains sequences conserved in the Gt1 and Gt2 subfamilies, including one copy of the long direct repeat. Fragment B contains the insertion sequence and one copy of the long direct repeat. **(b)** Gel-retardation assay of the conserved region (fragment A, position −45 to −223). Labeled fragment A (1 ng) was incubated with 4.5 μg of partially purified extract from immature rice seeds. Lanes: 1, free DNA; 2, plus extract; 3−5, 100-fold, 50-fold, and 10-fold excess of unlabeled fragment A, respectively; 6, 100 ng of α-amylase promoter fragment (260 bp long); 7, addition of 2 μl of a 1 mg/ml proteinase K. Lanes 2−7 contain 4 μg of poly [d(I-C)] as a nonspecific competitor DNA. The four major retarded bands are named as B1, B2, B3 and B4. **(c)** DNase I−footprinting of the conserved region. Extract containing 22.5 μg of total protein was incubated with 2 ng of labeled fragment A as described in Materials and Methods. DNase I digestion was for 30 seconds with 0.025 units of the enzyme (− lanes), or for 1 min with 0.1 units of the enzyme (+ lanes). M, (G+A) chemical sequencing ladder; −, without extract; +, with extract. The four protected regions are marked by thick bars and Roman numerals. Weakly protected sequences are indicated by broken bars. Arrowheads show DNase I-hypersensitive sites. The numbers at the left hand side show nucleotide positions relative to the transcription start site.

due to the binding of protein factors to the DNA fragment. The retarded bands are competed out by adding 100-fold or 50-fold excess of the unlabeled promoter fragment (lanes 3 to 4). In contrast, increasing the amount of non-specific competitor DNA, poly[d(I-C)], from 1 to 5 μg does not inhibit their formation significantly (data not shown). We also used a promoter fragment of a rice α-amylase gene (23) as a competitor. The fragment prevented formation of complex B4 completely (lane 6).

To localize protein-binding regions in more detail, we performed DNase I-footprinting analysis of the same DNA fragment (fragment A), and observed several protected regions (Figure 3c). The region showing the strongest protection (Box III) occurs between position −164 and −146. This region contains a sequence element, TGTGTCA, which is present in the −300 element, the sequence conserved in the 5′ flanking sequences of several seed storage protein genes of wheat and barley (see Discussion for details). Another strongly protected region (Box I) is centered at position −90. The protection in this region appears to be non-uniform: the upstream half of the sequence (−103 to −86) is more strongly protected than the downstream half (−85 to −74) in the coding strand. The upstream half of Box I shows significant sequence identity to the Box III sequence (Figure 5). In addition to the two clearly protected regions we detect a third region (Box IV) near the end of the conserved 5′ flanking sequence. The sequence (−206 to −189) is separated from Box III by a stretch of purines, and flanked by several DNase I-hypersensitive sites (marked by arrowheads). There appears to be a fourth protected region (Box II) between Box I and Box III (−122 to −108). The protection is weak and not so obvious as in other three regions. Nevertheless, we believe that Box II is another protected region for the following reasons. First, it is consistently observed and, occasionally, the protection is as strong as in Box I (not shown). Second, the α-amylase promoter fragment which competes out complex B4 in gel-retardation assays described above, contains a sequence element that is very similar to the Box II sequence not only in its DNA sequence but also in its relative position from transcription start site (Figure 7a). In addition, we observe DNase I-hypersensitive sites flanking Box II, as in other protected regions. These observations, when considered together, support the notion that Box II is protected from DNase I digestion. In summary, we detect three, or more likely four, protected regions in the DNase I-footprinting analysis of the conserved 5′ flanking sequence of pGL5-1.

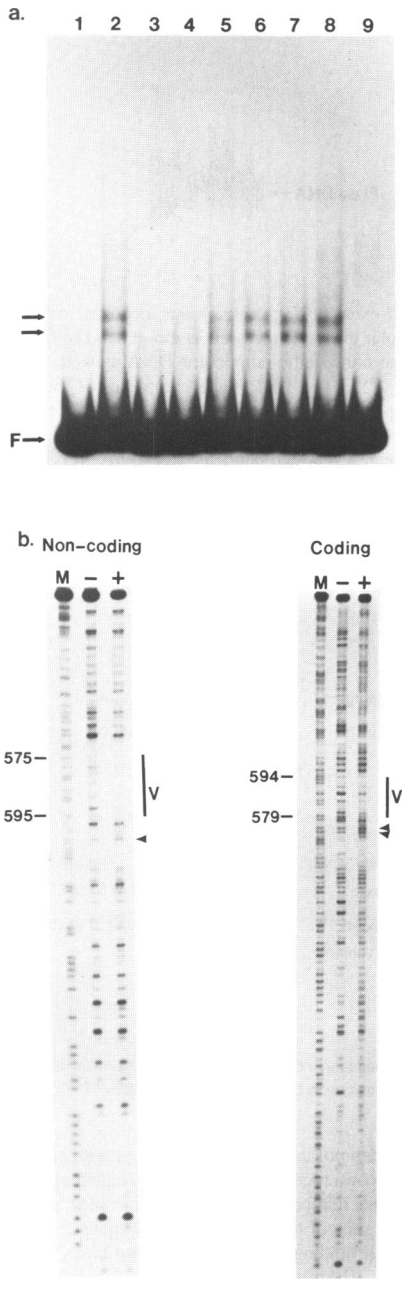

Figure 4. Binding of protein factors to fragment C of the upstream region of pGL5-1. **(a)** Gel-retardation assay of fragment C (Figure 3a). Labelled fragment C (3 ng) was incubated with 6 μg of extract. Lanes: 1, free DNA; 2 and 8, plus extract; 3−5, 100-fold, 50-fold, 15-fold excess of unlabeled fragment C, respectively; 6, 100-fold molar excess of fragment A; 7, 100-fold molar excess of fragment B; 9, 0.5 mg/ml of proteinase K. All the lanes contain 4 μg of the nonspecific competitor DNA. The two retarded bands are marked by arrows. F, free DNA. **(b)** DNase I−footprinting of fragment C. The assay was performed in the same way as in Figure 3c, using 22.5 μg of total protein and 2 ng of labeled fragment C. M, (G+A) chemical sequencing ladder; −, without extract; +, with extract. The protected region is marked by the thick bar and identified by the Roman numeral.

Binding of protein factors to the upstream sequence of pGL5-1

We extended the protein binding study beyond the immediate upstream region of the 5′ flanking sequence of pGL5-1, up to

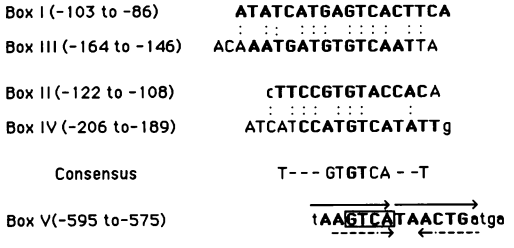

Figure 5. The protein-binding sequences in the promoter region of pGL5-1. Bold, upper case letters represent sequences protected in both DNA strands. Sequences protected only in the coding strand are shown by upper case letters, and those protected in the noncoding strand are shown by lower case letters. For Box I, only the strongly protected sequences are included. Numbers in the parentheses denote the position of the first and the last nucleotides relative to the transcription start site. The consensus sequence between Box I through IV is shown in line 5. The GT dinucleotides conserved in all four regions are shown by bold letters. The 7 bp imperfect direct repeat of Box V is indicated by arrows, and the 5 bp repeat by broken lines with arrowheads. The four nucleotides identical to part of the conserved sequence of other protected regions are boxed.

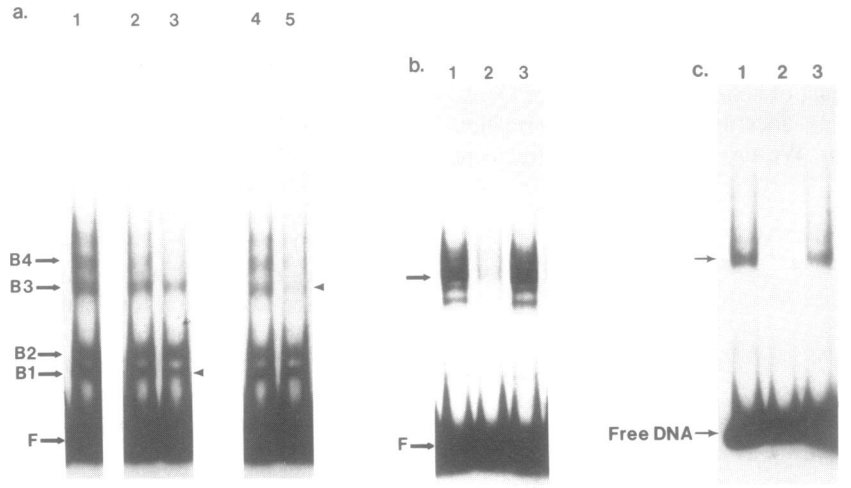

Figure 6. Binding study with synthetic oligonucleotides. (a) Gel-retardation competition assay. Nuclear extract (3 μg protein) was incubated with 2 ng of the labeled fragment A (Figure 3a). Lanes 1, control with no specific competitor; 2 and 3, 90-fold and 450-fold molar excess (expressed as monomers) of GT1-14, respectively; 4 and 5, 90-fold and 450-fold molar excess of GT3-13, respectively. All the lanes contain 3 μg of the nonspecific competitor DNA as well. Arrowheads indicate the position of the specific bands that are competed out by GT1-14 and GT3-13. F, free DNA. B1, B2, B3 and B4, same complexes as those in Figure 3b. **(b)** Gel-retardation assay of a cloned oligomer of GT1-14. Labeled DNA fragment (2 ng) containing four copies of GT1-14 in tandem head-to tail orientation was incubated with extract (4.5 μg of total protein) in the presence of 5 μg of poly[d(I-C)]. Lanes: 1, positive control with no specific competitor DNA; 2, 100 ng of fragment A as a competitor; 3, 120 ng of fragment C as a competitor. The arrow shows the position of the strongest retarded band. F, free DNA. **(c)**. Gel-retardation assay of a cloned oligomer of GT3-13. Labeled DNA fragment (2 ng) containing four copies of GT3-13 in tandem head-to-tail orientation was incubated with 3 mg of extract in the presence of 2 μg of poly[d(I-C)]. Lanes; 1, control with no specific competitor DNA; 2, 100 ng of fragment A as a competitor; 3, 120 ng of fragment C as a competitor. The arrow in the central part of the figure shows the retarded band position.

position −677. The fragment, from −224 to −677, was divided into two smaller fragments (fragments B and C in Figure 3a) for the convenience of gel-retardation and DNase I-footprinting assays. We did not detect any complex formation in gel-retardation assays using fragment B (data not shown). However, two closely-positioned retarded bands are evident in the gel-retardation assay using fragment C (Figure 4a). Formation of the complexes is sensitive to proteinase K (lane 9) and inhibited by addition of 100-fold or 50-fold excess of unlabeled fragment C (lanes 3 and 4), but neither fragment A nor fragment B shows any competition for the retarded bands (lanes 6 and 7). The results indicate that the two retarded bands are due to sequence-specific protein binding and that the sequence(s) which is responsible for the protein−DNA complex formation is unique to fragment C. In footprinting analysis one region (Box V) that is centered at position −585 is protected from DNase I digestion (Figure 4b).

Summary of the protein binding studies

The results of the protein binding studies are summarized in Figure 5. Comparison of the four protected DNA sequences in the conserved 5′ flanking sequence (fragment A) with each other reveals that they can be classified into two groups. Box III and the strongly protected part of Box I are closely related to each other: 12 bases are conserved among the 18 protected bases when the two sequences are aligned for best match. Boxes II and IV form another group, 8 out of 15 protected bases showing identity to each other. In addition, several nucleotides appear to be conserved among all four protected regions (consensus). The dinucleotide sequence GT occurs in all four boxes at the same relative position and the four nucleotides surrounding it are present in three out of the four regions. Box V is composed of a 7 bp imperfect direct repeat and shares 4 (GTCA) of the 6 conserved nucleotides in the other four protected regions. Box V can also be considered to contain a 5 bp repeat facing each other (dotted lines with arrowheads).

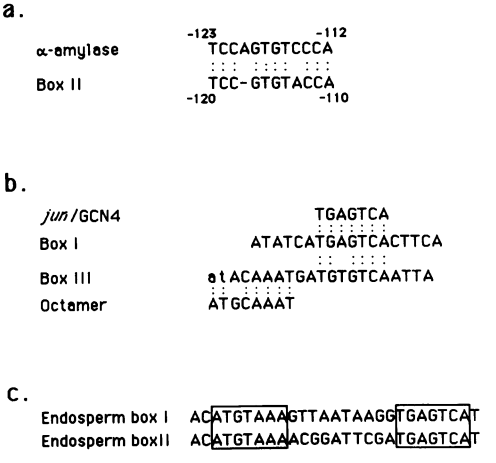

Figure 7. Comparison of the protected sequences with other sequence elements. (a) The sequence conserved in the 5′ flanking sequences of pGL5-1 and the α-amylase gene. The numbers indicate the nucleotide position relative to the transcription start sites of the two genes. **(b)** Sequence identity of Boxes I and III to other protein binding sequences. The two bases A and T located at the left border of Box III are not protected from DNase I digestion. **(c)** A typical example of the −300 element. The two sequences, present in the 5′ flanking sequence of one of the wheat LMW glutenin genes, are from Colot et al. (29). The sequences homologous to the octameric and the GCN4 binding sites are boxed. The endosperm Boxes I and II are centered at −250 and −525, respectively, relative to the transcription start site.

Protein binding to synthetic oligonucleotides

In order to confirm the results of the protein binding studies and to delineate sequences necessary for binding of specific protein

factors, we performed binding studies with synthetic oligonucleotides. Two sets of oligonucleotides, GT1-14 and GT3-13, were chemically synthesized for this purpose. GT1-14 (5′-gatcATCATGAGTCACTT-3′ and its complementary sequence) is composed of 14 bp of the upstream half of the Box I sequence and 4 bp Bam HI recognition sequence (lower case letters), whereas GT3-13 (5′-GATGTGTCAATTA-3′ and its complementary sequence) contains most (13 out of 19 bp) of the Box III sequence.

To examine whether any protein factors bind to the synthetic oligonucleotides, we first employed GT1-14 or GT3-13 as a competitor DNA in the gel-retardation analysis of the fragment A. Since we did not observe any competition in the initial assays with monomers, we used ligated GT1-14 and GT3-13 in subsequent assays. The synthetic oligonucleotides were phosphorylated and, after annealing, ligated to give an average number of 5 to 10 repeated units. The whole mixture of ligated GT1-14 or GT3-13 was then added to the assay mixture to test its ability to compete with fragment A. As shown in Figure 6a each of the ligated oligonucleotide mixtures shows a distinct competition pattern. Ligated GT1-14 competes with complex B1 (lane 3), whereas ligated GT3-13 competes with complex B3 (lane 5). Both of them, especially ligated GT1-14, also compete with complex B4. The fact that competition patterns of GT1-14 and GT3-13 are different from each other implies that different proteins may bind to the two protected regions, Boxes I and III, even though their sequences are closely related.

Next, we carried out direct gel-retardation assays using an oligomer of GT1-14 or GT3-13. Cloned DNA fragments containing four copies of GT1-14 or GT3-13 in tandem, head-to-tail orientation were labeled and used in the assays. With GT1-14 several protein-DNA complexes are observed (Figure 6b), and with GT3-13 a single complex is detected (Figure 6c). Formation of the complexes is inhibited by adding unlabeled fragment A (lanes 2 in both Figures 6b and 6c), which contains the GT1-14 and the GT3-13 sequences, but not by the unrelated fragment C (lanes 3).

DISCUSSION

In this paper, we describe the cloning of a rice glutelin gene (pGL5-1) and binding of putative regulatory proteins to the 5′ flanking sequence of the cloned gene. Our results show that several complexes are formed between nuclear proteins from immature rice seeds and the 5′ upstream sequence of pGL5-1.

According to the study of Okita et al. (7), subfamilies of rice glutelin genes are differentially expressed in accordance with the sequence divergence in the 5′ flanking sequences. mRNAs of the closely related Gt1 and Gt2 subfamilies continue to be accumulated for at least 20 days during seed development. In contrast, the mRNA level of the Gt3 subfamily reaches a peak about 10 days after flowering, and gradually decreases afterwards. This differential expression of glutelin genes implies the presence of subfamily-specific regulatory element(s) in addition to those common to all glutelin genes. From sequence comparison we observe that Boxes III, IV and V are located in regions where Gt3 shows significant divergence from Gt1 and Gt2. Thus, the function of Boxes III through V might be specific to the Gt2 and Gt1 subfamilies. As mentioned earlier, complex B4 of fragment A was competed out by a rice α- amylase promoter fragment (Figure 3b). We searched for sequence identity between the two DNA fragments, and the result shows that a Box II-like sequence is also present in the α-amylase promoter fragment (Figure 7a). The similarity between the two sequence elements is not only in their DNA sequences, but also in their relative position to the transcription start sites.

It is noteworthy that the TGAGTCA binding motif (Figure 7b), to which transcription factors jun and GCN4 are known to bind (24), is located in the middle of the Box I and the Box III sequences. The seven base pair element shows perfect identity to the binding motif in Box I, and has one base mismatch in Box III. The binding site is also found in the 5′ flanking sequences of several seed storage protein genes of wheat and barley as a part of the −300 element (25, 26). In a wheat LMW glutenin gene the element is repeated twice, and one of the elements is located in the middle of a 160 bp sequence which confers tissue-specificity to the expression of the glutenin gene in transgenic tobacco (27, 28). Thus, the TGAGTCA motif might play an important role in tissue-specific expression of the glutelin gene.

Also noteworthy is the presence of the octameric binding site (29), ATGCAAAT, in Box III with one base mismatch (Figure 7b). This binding site is separated from the TGAGTCA motif by two bases in pGL5-1, and thus, Box III consists of the two binding motifs arranged in tandem. Interestingly, the −300 element found in storage protein genes of other cereal plants also has a similar structure (Figure 7c), but the two binding motifs are separated from each other by 8 to 10 bp of non-conserved sequences.

ACKNOWLEDGMENT

We are grateful to J. Cao for the bombardment of rice seeds, and R. Okagaki and S. Wessler for providing us with a rice genomic DNA library. We thank V. Shaff for typing, I. Tanida, H. Moore, and D. McElroy for reading the manuscript. The work was supported by grant RF 84066, allocation No. 3, from the Rockefeller Foundation.

REFERENCES

1. Yamagata, H., Sugimoto, T., Tanaka, K. and Kasai, Z. (1982) Plant Physiol., 70, 1094−1100.
2. Zhao, W.M., Gadehouse, J.A. and Boulter, D. (1983) FEBS Lett., 162, 96−102.
3. Wen, T.N. and Luthe, D.S. (1985) Plant Physiol., 78, 172−177.
4. Yamagata, H. and Tanaka, K. (1986) Plant Cell Physiol., 27, 135−145.
5. Takaiwa, F., Kikuchi, S. and Oono, K. (1986) FEBS Lett., 206, 33−35.
6. Higuchi, W. and Fukazawa, C. (1987) Gene, 55, 245−253.
7. Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R. and Krishnan, H.B. (1989) J. Biol. Chem., 264, 12573−12581.
8. Takaiwa, F., Kikuchi, S. and Oono, K. (1987) Mol. Gen. Genet., 208, 15−22.
9. Wang, C.-S., Shastri, K., Wen, L., Huang, J.-K., Sonthayanon, B., Muthukrishnan, S. and Reeck, G.R. (1987) FEBS Lett., 222, 135−138.
10. Masumura, T., Kidzu, K., Sugiyama, Y., Mitsukawa, N., Hibino, T., Tanaka, K. and Fujii, S. (1989) Plant Mol. Biol., 12, 723−725.
11. Takaiwa, F., Kikuchi, S. and Oono, K. (1989) Nucleic Acids Res., 17, 3289.
12. Takaiwa, F., Ebinuma, H., Kikuchi, S. and Oono, K. (1987) FEBS Lett., 221, 43−47.
13. Leisy, D.J., Hnilo, J., Zhao, Y. and Okita, T.W. (1990) Plant Mol. Biol., 14, 41−50.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. (USA), 74, 5463−5467.
16. Klein, T.M., Wolf, E.D., Wu, R. and Sanford, J.C. (1987) Nature, 327, 70−73.
17. Bradford, M.M. (1976) Anal. Biochem., 72, 248−254.
18. Mullis, K.B. and Faloona, F.A. (1987) In Wu, R. (ed.), Methods in Enzymol., Academic Press, Vol. 155, pp. 335−350.

19. Iturriaga, G., Jefferson, R.A. and Bevan, M. W. (1989) Plant Cell, 1, 381–390.
20. Garner, M.M. and Rezvin, A. (1981) Nucleic Acids Res., 9, 3047–3060.
21. Fried, M. and Crothers, D.M. (1981) Nucleic Acids Res., 9, 6505–6525.
22. Galas, D. and Schmitz, A. (1978) Nucleic Acids Res., 5, 3157–3170.
23. Ou-Lee, T.-M., Turgeon, R. and Wu, R. (1988) Proc. Natl. Acad. Sci. (USA), 85, 6366–6369.
24. Curran, T. and Franza, B.R. Jr. (1988) Cell, 55, 395–397.
25. Forde, B.G., Heyworth, A., Pywell, J. and Kreis, M. (1985) Nucleic Acids Res., 13, 7327–7339.
26. Halford, N.G., Forde, J., Anderson, O.D., Greene, F.C. and Shewry, P.R. (1987) Theor. Appl. Genet., 75, 117–126.
27. Colot, V., Robert, L.S., Kavanagh, T.A., Bevan, M.W. and Thompson, R.D. (1987) EMBO J., 6, 3559–3564.
28. Colot, V., Bartels, D., Thompson, R. and Flavell, R. (1989) Mol. Gen. Genet., 216, 81–90.
29. Garcia-Blanco, M.A., Clerc, R.G. and Sharp, P.A. (1989) Genes Dev., 3, 739–745.