Mol Gen Genet (1994) 245:177–186

ORIGINAL PAPER

Fernando A. de Freitas · José Andrés Yunes
Marcio J. da Silva · Paulo Arruda · Adilson Leite

Structural characterization and promoter activity analysis of the γ-kafirin gene from sorghum

Received: 26 January 1994 / Accepted: 12 April 1994

Abstract A genomic clone encoding the γ-kafirin gene from sorghum was isolated and sequenced. A 2938 bp sequenced fragment includes an intronless open reading frame of 636 nucleotides encoding a putative polypeptide of 212 amino acids. Comparison of the deduced amino acid sequence of γ-kafirin with the published sequences of γ-prolamins of maize, and *Coix* revealed highly conserved domains. The N-terminal region of these proteins contains the conserved hexapeptide PPPVHL, which is repeated eight times in γ-zein, four times in γ-kafirin and three times in γ-coixin. The number of PPPVHL repeats accounts predominantly for the differences in the molecular weights of γ-prolamins. Several putative regulatory sequences common to the γ-kafirin and γ-zein genes were identified in both the 5′ and the 3′ flanking regions. Putative *GCN4*-like regulatory sequences were found at positions −192 and −476 in the 5′ flanking region of γ-kafirin. In the 3′ noncoding region, three putative polyadenylation signals, two AATAAT and one AATGAA, were found at positions +658, +716, and +785, respectively. In order to investigate the role of the putative *GCN4*-like motifs and other possible *cis*-acting element(s) of the γ-kafirin promoter, a series of deleted and chimeric promoter constructs were introduced into maize, *Coix* and sorghum tissues by particle bombardment. Histochemical analysis of β-glucuronidase (GUS) activity in different tissues indicated that the element(s) responsible for tissue specificity is probably located in the 285-bp proximal region of the promoter, while the remaining promoter sequence seems to carry the element(s) responsible for the quantitative response.

Key words Gene expression · Tissue specificity · Seed storage protein · Sorghum · γ-Kafirin

Communicated by H. Saedler

F. A. de Freitas · J. A. Yunes · M. J. da Silva
P. Arruda · A. Leite (✉)
Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Cidade Universitária "Zeferino Vaz", 13081, Campinas, SP, Brazil

P. Arruda
Departamento de Genética e Evolução, IB, Universidade Estadual de Campinas, 13081, Campinas, SP, Barzil

The nucleotide sequence data reported in this paper appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number X62480

Introduction

Prolamins are the major storage proteins of many cereals (reviewed by Shewry and Tatham 1990). They constitute a group of alcohol-soluble proteins that are deposited in protein bodies in the developing starchy endosperm. These proteins generally account for about half of the proteins in mature grains, and contain high levels of proline and glutamine, and low levels of lysine and tryptophan. Based on their solubility, zeins, the prolamins of maize, have been classified into three distinct classes: α-, β-, and γ-prolamins (Esen 1986). This classification was further extended to the prolamins of sorghum (DeRose et al. 1989) and *Coix* (Leite et al. 1990). However, solubility alone does not always reflect common structural features of such proteins. For example, Kirihara et al. (1988) sequenced a methionine-rich zein that has the solubility properties of β-prolamins but its structure led them to introduce a fourth prolamin class, the δ-zein. The 17 kDa coixin, one of the prolamins, of *Coix*, was first described as an α-prolamin based on its solubility properties (Leite et al. 1990). However, after cDNA cloning, its sequence displayed high similarity to the 14 kDa β-zein (Leite et al. 1992).

In maize, γ-prolamins account for 5% to 10% of the total zein (Esen 1986) and are composed of two polypeptides (Prat et al. 1987), one of 28 kDa and the other of 16 kDa. The 16 kDa γ-zein has solubility properties similar to the β-prolamins (Esen 1986). The 28 kDa

γ-zein is encoded by one or two genes, while the 16 kDa γ-zein is encoded by a single gene (Gallardo et al. 1988; Das and Messing 1987; Das et al. 1991a, b).

In *Coix*, γ-coixin represents about 15% of total prolamins and comprises a single polypeptide of 22 kDa, highly homologous to the 28 kDa γ-zein (Leite et al. 1990, 1991).

Sorghum endosperm contains a lower concentration of γ-prolamin than do maize and *Coix*. γ-Kafirin, the γ-prolamin of sorghum, accounts for 2% to 5% of total prolamin and is composed of a single polypeptide of 27 kDa (Evans et al. 1987; Taylor et al. 1989).

The overall molecular data have shown a high degree of similarity between the genes encoding the prolamins of maize, *Coix* and sorghum, suggesting that these genes may have arisen from a common ancestor and that divergence occurred after species separation. The comparison of structural and regulatory sequences of these genes might allow a more precise definition of conserved domains important for regulation of expression and protein folding.

In this work, the characterization of a γ-kafirin genomic clone is described. Sequences related to the protein structure and putative regulatory elements were identified by comparison with γ-prolamin genes from *Coix* and maize. The ability of the γ-Kafirin promoter to drive β-glucuronidase (*gus*) gene expression in sorghum, maize and *Coix*, and its efficiency and tissue specificity in maize were studied by means of transient gene expression analysis.

Materials and methods

Plant material

Coix lacryma-jobi var. Adlay, *Sorghum vulgare* cv. INRA 450, and *Zea mays* F352 single hybrid were obtained from the germplasm collection of the Department of Genetics, State University of Campinas (Campinas, Brazil), and from Sementes Agroceres (São Paulo, Brazil). Immature seeds were obtained from field-grown plants. For transient gene expression experiments, immature seeds were surface-sterilized for 45 min with 5% commercial bleach and rinsed four times with distilled water. Seeds were sectioned longitudinally and placed in 100 mm petri dishes containing 15 ml of solid MS medium (Murashige and Skoog 1962) with the sliced surface facing upward. Embryos were isolated from sterile immature maize seeds and flattened in 100 mm petri dishes containing 15 ml of solid MS medium. Coleoptiles were dissected from seedlings 3 days after germination, sterilized for 5 min with 5% commercial bleach and rinsed four times in distilled water. Leaves were dissected from 7-day-old seedlings and sterilized as described for coleoptiles.

Purification of γ-prolamins and N-terminal sequencing

The γ-prolamins were extracted from endosperms dissected from mature seeds of maize, *Coix* and sorghum and separated from other prolamins by differential solubility (Esen 1986). The purity of the γ-prolamin fractions was verified by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli 1970) followed by silver staining (Moreno et al. 1985). Molecular weight was estimated as described by Plikaytis et al. (1986). N-terminal sequencing was performed on a 477A Protein Sequencer. The resulting phenylthiohydantoin derivatives were identified using an on-line phenylthiohydantoin analyzer 120A (Applied Biosystems, Foster City, Calif., USA).

DNA extraction

Seeds of sorghum were germinated in the dark for 10 days. DNA extraction from leaves was performed according to Rivin et al. (1982).

Selection of a recombinant clone encoding γ-kafirin

Approximately 1.5×10^6 phages from a *Sau*3AI sorghum genomic library constructed in lambda EMBL-4 were plated out on *Escherichia coli* DL-538. Nitrocellulose blots were prepared as described by Benton and Davis (1977). Filters were prehybridized in 6 × SSC, 0.05 × BLOTTO (Johnson et al. 1984), and 50% formamide for 4 h at 42° C. Hybridization was carried out overnight under the same conditions, using the 28 kDa γ-zein cDNA clone pME119 (Prat et al. 1985) as a probe. The *Pst*I fragment of the cDNA clone was labeled with α-[^{32}P]dCTP by random primer extension using the Megaprime kit (Amersham). The filters were washed once in 1 × SSC, 0.1% SDS for 20 min at room temperature, and twice in 0.1 × SSC, 0.1% SDS for 15 min at 60° C. After autoradiography, a unique positive plaque designated 10S was purified by two additional cycles of replating and hybridization.

Mapping and sequencing the γ-kafirin genomic clone

The γ-kafirin genomic clone was mapped by the *cos* mapping technique (Rackwitz et al. 1984) using the λ-Mapping System (Amersham). The recombinant phage was sub-cloned into the pBluescript KS + vector (Stratagene) and sequenced on both strands using the T7 DNA sequencing kit (Pharmacia/LKB). The sequence data were analyzed using the DNASIS sequence analysis software (Pharmacia/LKB). The alignments were primarily performed on CLUSTAL V software (Higgins and Sharp 1989) and the best alignment was further arranged manually.

Plasmid constructs

For all the constructs, the sequence positions are given relative to the ATG initiation codon.

pPKGUS

The 5′ flanking sequence of the 27 kDa γ-kafirin gene was isolated from the genomic clone 10S. The *Taq*I fragment spanning positions −1198 to −8 (see Fig. 1) was cloned into the *Cla*I site of pBluescript KS + (Stratagene), resulting in the pKTT1. The *Sma*I-*Sal*I fragment isolated from pKTT1 was inserted into the pRT103 vector (Töpfer et al. 1987) digested with *Hin*cII and *Xho*I, replacing the CaMV 35S promoter. This construction was designated pPK103. Further, the β-glucuronidase (*gus*) coding region was excised from pJII140 (Gallie et al. 1987) by digestion with *Nco*I and *Bam*HI and fused into the pPK103 clone in the corresponding sites between the γ-kafirin promoter and the CaMV 35S polyadenylation signal.

pPK285GUS

The −285 to −8 γ-kafirin promoter fragment was excised from pKTT1 by digestion with *Eco*RV and *Sal*I. This fragment was

ligated into pRT103 digested with *Hin*cII and *Xho*I, replacing the CaMV 35S promoter, producing pPK285. As described above, the *gus* coding region was further inserted into the pPK285 *Nco*I-*Bam*HI sites.

pRT103GUS

This construct was obtained by inserting the *gus* coding region into the *NcoI/Bam*HI sites of pRT103.

pRT90GUS

The *gus* coding region was inserted into the *Nco*I/*Bam*HI sites of pRT90, which was obtained by recircularization of the pRT103, which had previously been digested with *Eco*RV and *Hin*cII.

pPK90GUS

The −1190 to −285 γ-kafirin promoter fragment excised from pKTT1 by *Eco*RV digestion was inserted into pRT103 digested with *Eco*RV and *Hin*cII, producing pPK90. Then the *gus* coding region was inserted into the *Nco*I/*Bam*HI sites of pPK90.

Transient expression assays

All tissues were transiently transformed by particle bombardment performed with a home-made helium particle delivery device. Five micrograms of CsCl-purified DNA were used to coat 3 mg of 1 to 3 μm gold particles according to the method of Ye et al. (1990). Petri dishes containing samples of all tested tissues, arranged in a randomized complete block design, were bombarded twice with 0.5μg of the plasmid. After bombardment, samples were incubated in the dark at 27° C for 48 h. The tissues were then developed for histochemical GUS activity according to Jefferson (1987). The relative GUS activity was calculated from the ratio between the number of blue spots obtained for each construct and the number obtained for pRT103GUS.

Results and discussion

Isolation and characterization of a genomic clone encoding γ-kafirin

In order to isolate the γ-kafirin gene, a sorghum genomic library was screened with the γ-zein cDNA clone pME119 (Prat et al. 1985). One of the hybridizing phage, designated 10S and containing a 10 kb DNA insert, was characterized. The restriction map of the 10S genomic clone as well as the 2938-bp fragment that was completely sequenced are shown in Fig. 1. Figure 2 shows the nucleotide sequence of the γ-kafirin gene. The 5′ and 3′ flanking regions consist of 1511 bp and 789 bp, respectively. The open reading frame of 636 nucleotides encodes a polypeptide of 212 amino acids.

The nucleotide sequence of the γ-kafirin gene exhibits only a few differences from the cDNA clone sorgH described by Barros et al. (1991). The coding region can be perfectly aligned with the sorgH clone, except for five nucleotides, four of which represent silent transitions, and a non-synonymous transversion of G to C at position + 236. A deletion of one T in the 3′ flanking region was also observed in the γ-kafirin

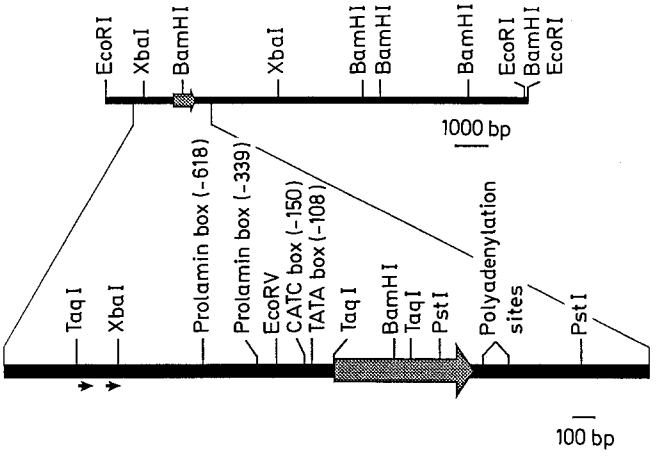

Fig. 1 Restriction map of the 10S γ-kafirin genomic clone.The *upper map* shows the entire insert in the EMBL-4 lambda vector, with the structural region of the γ-kafirin gene represented by an *arrow* indicating the transcriptional orientation. The *lower map* shows the sequenced region of the insert. The positions of the TATA box, CATC box, prolamin box and putative polyadenylation signals are indicated. The *small arrows* indicate the localization of the 39-nucleotide repeats

gene. Sequence comparison with the γ-kafirin cDNA clone also revealed the absence of introns in the genomic sequence. Reconstruction experiments indicated that γ-kafirin, like the 28 kDa γ-zein protein (Gallardo et al. 1988), is encoded by one or two genes (data not shown).

Conserved putative *cis*-acting regulatory elements were found in the 5′ region of the γ-kafirin gene (Figs. 1 and 2). The TATA box is located at position −108 whereas a putative CATC box was located 42 bp upstream from the TATA box, at position −150. Two copies of the sequence TGTAAAGT, the prolamin box, or " −300 element", are present in the γ-kafirin promoter. As in other cereal storage protein genes, one copy is located −340 nucleotides from the ATG (Kreis et al. 1986; Ottoboni et al. 1993), while the second is located at position −618. Upstream of the more distal prolamin box, a repeated 39-bp motif was found at positions −1052 and −1162. In the 3′ noncoding region, three putative polyadenylation signals, two AATAAT, and one AATGAA, were found at positions +658, +716, and +758, respectively.

Characterization of γ-prolamins

The γ-prolamins of *Coix*, sorghum, and maize were purified to homogeneity as demonstrated by silver straining after SDS-PAGE (Fig. 3). The apparent molecular weights of γ-coixin, γ-kafirin, and γ-zein were estimated to be 22, 27, and 28 kDa, respectively. The shorter 16 kDa γ-zein was not purified in the same fraction as the 28 kDa γ-zein, as previously demonstrated by Esen (1986). Figure 4 shows a comparison of

Fig. 2 Nucleotide and deduced amino acid sequence of the 10S γ-kafirin genomic clone. Nucleotide positions are numbered 5′ to 3′ and are indicated on the left side of each line. The 39-nucleotide repeats (−1161 and −1052), prolamin boxes (−618 and −340), CATC box (−150), TATA box (−108), as well as the putative polyadenylation signals (+658, +716 and +758), are *underlined*

```
-1511  AATATGCCTCAGTATTATATTAAAACATGAGAAATTTGATGCTAGCATTATTTTTTAATTTAGATTAAGATTGCTAT
-1434 GAAATATGAATCAAGCATCACATTAAATAGGCTACCATTAAATTTTTATAATAATTGGAGCAAGATAACTATAATTTT
-1356 AATTTTACAGTAAAAAGCATAGGCAAGCATTTTCAAAAAAAAAGTGTACTAAAATTTATGATATTTTTTGTTTACCGCA
-1278 TGGATCTACATATGTGGAGCTGAACAAAATTTGTTTTATAATTTTTAGATTTTTGTATGAATTATTATGTATTTATTA
-1200 ATTTTCAGTCGATTTAAAAAATAAAAGAAAAGTAAATTGGAGATTTTACATTGGGAACCTAGAAAATTTTTTATTTTT
-1122 TTTCTCTCTTCCATCGTAACGATTCTGGGCTGATTGGGCCTACAGAGAGGAGAGCGAAAGCGAAGTAATATGAGATTT
-1044 TACATTAGGAACCCTAGAATATTTTTTATTTTGTTTTTCCCCTTCTAGAAAAGTTGGATGGAGGAGGCCATCGGTAAA
 -966 CCGAACACGGACAAGATCAGGGGCAAAAGAAAATATTTGGCAAACTAAAATTTTGGCTCTTTATAGATAGGTATAGAT
 -888 TTGGGTTAAGAACAGTAGTCGGACATTCCACGCTTTGTTAAGTTCTTTTTTTACATAACAAGAATAGAATCACGCTCT
 -810 GTAGGCTCGTCAGACCGTACCTTTCTAAGGAAGTCGCTTTGGGTAGTTCAGTTGGCGAGAAAAGCCTTCCTACCTTTG
 -732 CAGGTCCATCGGGCCGACTACAACCCGTGGCTCAATCCCGGTTCTTGCGGTGTCTTGGCAACATTCTTGTTGGAAGAT
 -654 ACCAGAAGGTTGCTCCACGGGTAATCTTGACACGTATGTAAAGTGATGAGGAACATTGAACGAACATTGGCATGTAAG
 -576 CTCTATAATTGGTGTTATCCATAACAACGTCGCAGAACATCACAAATTGCACGTCAAGGGATTGGGTCAGAAACAAAT
 -498 CGTCTCCGTGTACAACGAAGTGGTGAGTCATGAGCCATGTTGATCTGATATATACATAGCACACACGACATCACAAAC
 -420 AAGTCATACTACATTACAGAGTTAGTTTCACCTTTCAAGTAAAAACAAAGTAGGCCGGAGAGAGGACAATAATCCTTG
 -342 ACGTGTAAAGTGAATTTACAAAGCCATATATCAATTTATATCTAATTCGTTTCATGTAGATATCAACAACCTGTAAAA
 -264 GGCAACAAATTGAGCCACGCAAAATTACAAGTGAGTCCAAATAAACCCTCACATGCTACATAAAAGTGAATGATGAGT
 -186 CATGTATATCTGGCAAGAAACTGTAGAAGCTACAGTCATCGGTAGCAAAGAAACACAAGAAAATGTGCTAATAAAAGC
 -108 TATAAATAACCCTCGTACGCCTATGCACATCTCCATCACCACCACTGGTCTTCATTCAGCCTATTAACTTATATCTAT
  - 30 CTACTCCAGAGCAGACAAGAACTCGACACCATGAAGGTGTTGCTCGTTGCCCTCGCTCTCCTGGCTCTCGCGGCGAGC
    1                                  MetLysValLeuLeuValAlaLeuAlaLeuLeuAlaLeuAlaAlaSer
   49 GCCGCCTCCACGCTTACAACCGGCGGCTGCGGCTGCCAGACACCTCATCTACCACCACCGCCGGTTCATCTGCCGCCG
   17 AlaAlaSerThrLeuThrThrGlyGlyCysGlyCysGlnThrProHisLeuProProProProValHisLeuProPro
  127 CCGGTGCATCTGCCACCGCCGGTGCACCTGCCGCCGCCGGTTCACGTGCCACCGCCGCCACCACAATGCCACCCACAC
   43 ProValHisLeuProProProValHisLeuProProProValHisValProProProProProGlnCysHisProHis
  205 CCTACTCTACCGCCCCACCCACATCCATGCGCTACATACCCACCGCATCCAAGCCCGTGCCACCCAGGGCATCCCGGA
   69 ProThrLeuProProHisProHisProCysAlaThrTyrProProHisProSerProCysHisProGlyHisProGly
  283 TCCTGCGGTGTTGGCGGCGGCCCCGTCACCCCGCCGATCCTGGGCCAGTGCATCGAGTTCCTGAGGCATCAGTGCAGC
   95 SerCysGlyValGlyGlyGlyProValThrProProIleLeuGlyGlnCysIleGluPheLeuArgHisGlnCysSer
  361 CCGGCGGCGACGCCCTACTGCTCGCCACAGTGCCAGGCGTTGCGGCAGCAGTGCTGTCAGCAGCTCAGGCAGGTGGAG
  121 ProAlaAlaThrProTyrCysSerProGlnCysGlnAlaLeuArgGlnGlnCysCysGlnGlnLeuArgGlnValGlu
  439 CCGCTGCACCGGTACCAGGCGATCTTCGGCGTGGTCCTGCAGTCCATCCAGCAGCAGCAGCCGCAAGGCCAGTCGTCA
  147 ProLeuHisArgTyrGlnAlaIlePheGlyValValLeuGlnSerIleGlnGlnGlnGlnProGlnGlyGlnSerSer
  517 CCGCTCCCGGCGCTGATGGCGGCGCAAATAGCACAGCAACTGACGGCGATGTGCGGTCTAGGAGTGGGGCAGCCAAGT
  173 ProLeuProAlaLeuMetAlaAlaGlnIleAlaGlnGlnLeuThrAlaMetCysGlyLeuGlyValGlyGlnProSer
  595 CCCTGCGCTTCTTGCAGCCCTTTTGCCGGTGGTGTCCACTATTAAAGAAACTATCTATACTGTAATAATGTTGTATAG
  199 ProCysAlaSerCysSerProPheAlaGlyGlyValHisTyr
  673 CCGCCGGATAGCTAGCTAGTTAGTCATTCAGCGGCGATGGGTAATAATAAAGTGTCATCCATCCATCACCATGGGTGG
  751 CAACGTGAGCAATGACCTGATTGAACAAATTGAAATGAAAAGAAGAAATATGTTATATGTCAACGAGATTTCCTCATA
  829 ATGCCACTGACAACGTGTGTCCAAGAAATGTATCAGTGATACGTATATTCACAATTTTTTTATGACTTATACTCACAA
  907 TTTGTTTTTTTACTACTTATACTCGAACAATTTGTTGTGGGTACCATAACAATTTCGATCGAATATATATCAGAAAGT
  985 TGACGAAAGTAAGCTCACTCAAAAAGTTAAATGGGCTGCGGAAGCTGCGTCAGGCCCAAGTTTTGGCTATTCTATCCG
 1063 GTATCCACGATTTTGATGGCTGAGGGACATATGTTCGGCTTAAGCTGCAGCTTTGTAGTTAGTTTTGTTTTTGATTAT
 1141 ATTTAATACTCTATGCATGTGCACCAAGATTTTTCTGGTGAATTAAACAAGGCCTAATAACGTGAGTAGCGTATCTAA
 1219 CTGTGACCTATAAAGTAGAGCACCTTTTTAGAGTAGGGGCTCCTTTTTTTAGAACTCTATTTATTGCACCCAACTTCA
 1297 ATAAGGGTCTTTTCATCCAAAATTAAGAGTCCTTACATTACATCTAATCGTCTATTCATTGTCTATATTTTAATATAA
 1375 ATCTTACTGTATATCTTGTAGCACACTAGTATGCCTCAAAGCCGACAATAAAT
```

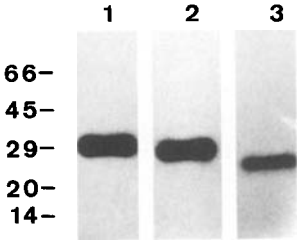

Fig. 3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of γ-prolamins of maize (lane 1), sorghum (lane 2) and *Coix* (lane 3) purified by differential solubility (Esen 1986). The numbers on the left indicate the positions of molecular weight markers in kilodaltons

the deduced amino acid sequence of the 27 kDa γ-kafirin with those of the 28 kDa and 16 kDa γ-zeins (Prat et al. 1985, 1987), and the 22 kDa γ-coixin (Leite et al. 1991). N-terminal sequencing of the first 6 residues of the mature γ-kafirin revealed the amino acids TLTTGG, which, when compared with the deduced amino acid sequence of the pre-protein, revealed the absence of the first 19 residues in the mature protein. This 19-amino acid fragment is highly conserved in all γ-prolamin sequences, and probably represents the signal peptide. After processing, the deduced amino acid sequence corresponding to the mature γ-kafirin polypeptide presents a calculated molecular weight of 20.3 kDa, which is 7 kDa smaller than the apparent molecular weight estimated by SDS-PAGE. This difference, also observed in the 28 kDa γ-zein (Wang and Esen 1986), can be explained by the presence of high levels of positively charged amino acid residues, like histidine (7.77%) and arginine (2.07%). An indication that this explanation is correct comes from the fact that the 16 kDa γ-zein, which contains lower levels of histidine and arginine, does not present such discrepancies between deduced and apparent molecular weight. Like the γ-prolamins of maize and *Coix*, γ-kafirin is characterized by high levels of the amino acids proline (22.8%), glutamine (11.9%), and cysteine (7.8%). The amino acid sequence alignment also shows that the positions of the ten cysteines are conserved for all the aligned γ-prolamins. The large number and conserved locations of such cysteines suggest that this amino acid

Fig. 4 Alignment of the deduced amino acid sequences of γ-prolamins. The amino acid sequences were deduced from the following clones: pME119 and pME177, (Prat et al. 1987) representing 28 and 16 kDa γ-zein; pBCX22.5 (Leite et al. 1991), representing γ-coixin and the 10S γ-kafirin genomic clone. The signal peptide at the N-terminus, as well as the repeat motifs are indicated. Identical residues and conservative substitutions are indicated by *asterisks* and *dots*, respectively. The gaps used to optimize the alignment are represented by *hyphens*

plays an important role in the formation of disulfide bridges during protein folding in the protein bodies of maize, sorghum and *Coix*.

The N-terminal region of γ-kafirin (Fig. 4) contains the same repetitive sequences as the 28 kDa γ-zein (Esen et al. 1982). The first common repeat motif, located adjacent to the conserved sequence CGCQ, is the hexapeptide PPPVHL. This motif is repeated eight times in γ-zein, four times in γ-kafirin and three times in γ-coixin. The difference in the number of PPPVHL repeats is responsible for the major differences in molecular weights observed for γ-prolamins in the three species. The short chain 16 kDa γ-zein contains only a few incomplete and modified versions of the PPPVHL motif. The fourth PPPVHL repeat of γ-kafirin contains a replacement of leucine by valine. This replacement is also observed in one 28 kDa γ-zein repeat. The second repeated motif of γ-zein is represented by the octapeptides QPHPCPCQ and QPHPSPCQ, which differ in only one amino acid (Wang and Esen 1986; Esen 1986). This duplicated motif is less conserved in γ-kafirin and γ-coixin proteins.

Comparison of the 5′ and 3′ regions of the γ-prolamin genes

Figure 5 shows the nucleotide sequence comparison of the γ-prolamin 5′ flanking regions. The γ-prolamin promoters share much similarity with the most conserved region comprising the proximal region surrounding the putative TATA box. On the other hand, it was not possible to identify a canonical CATC box. The 28 kDa γ-zein promoter exhibits a substitution of an A to a G, and one TC insertion relative to the others. Two copies of the prolamin box are present in both the 5′ flanking region of γ-kafirin and of the 28 kDa γ-zein genes. One of them is located at the same relative position in both genes, which is identical to the positions described for other prolamin gene classes of Andropogonea cereals, such as α-kafirin (DeRose et al. 1989), α-coixin (Ottoboni et al. 1993), α- and β-zeins (Brown et al. 1986; Pedersen et al. 1986; Liu and Rubenstein 1992). The presence of the prolamin box has also been observed in the prolamins of other cereals such as barley (Kreis et al. 1986), wheat (Summer-Smith et al. 1985; Colot et al. 1987), and rye (Hull et al. 1991). The γ-kafirin and the 28 kDa γ-zein genes also present one additional copy of the same sequence located in different but partially conserved regions of the promoters, while the 16 kDa γ-zein presents only truncated versions of the prolamin box.

Though present in many cereals, little is known about the role of the prolamin box in the regulation of gene expression. Functional analysis of the −300 region of zein promoters (Thompson et al. 1990; Quayle et al. 1991; Quayle and Feix 1992) has indicated that the prolamin box can stimulate the expression of prolamin genes and that this effect is position and orientation dependent.

Recently, Müller and Knudsen (1993) demonstrated that the transcription unit of a C-hordein gene from barley is responsive to exogenously supplied nitrogen. This response seems to be mediated by positive and negative regulation involving *GCN4*-like binding sites (Hill et al. 1986), and prolamin box sequences. The interaction between *GCN4*-like motifs and the prolamin box was also demonstrated by in vivo footprinting of a low molecular weight glutenin gene in wheat endosperm (Hammond-Kosack et al. 1993). A search for *GCN4*-like motifs (G(A/G)TGA(G/C)TCAT) in the γ-prolamin promoters revealed two copies at conserved position's (Fig. 5). The first one, located at position −192 of the γ-kafirin promoter, is completely conserved in the 16 kDa γ-zein promoter and presents only one nucleotide substitution in the 28 kDa γ-zein

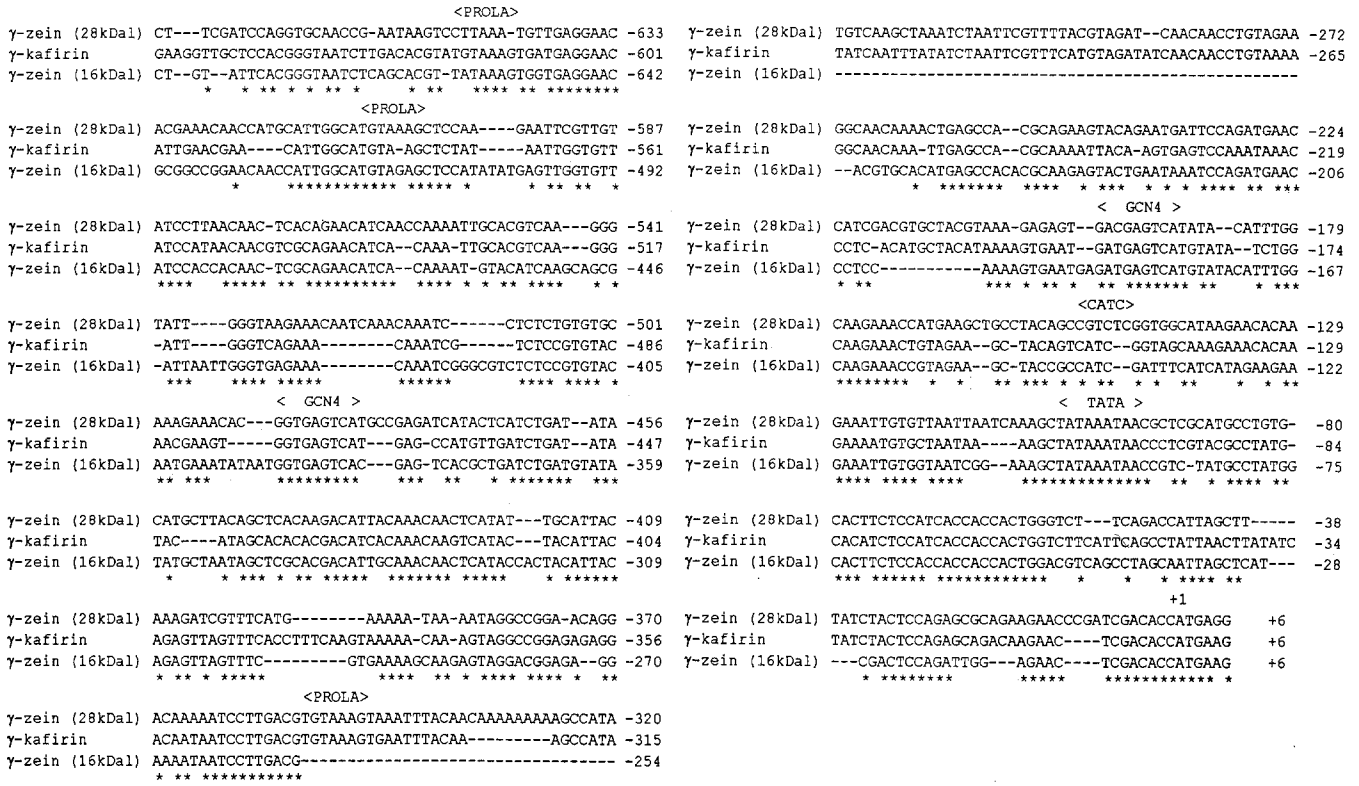

Fig. 5 Alignment of the 5′ flanking regions of γ-prolamins. The 16 and 28 kDa γ-zeins are represented by the nucleotide sequences of Zc1 (Reina et al. 1990a) and Zc2 (Reina et al. 1990b) genomic clones, respectively. The positions of the prolamin boxes (⟨PROLA⟩), GCN4-like motifs (⟨GCN4⟩), CATC box (⟨CATC⟩), and TATA box (⟨TATA⟩) are indicated. Identical residues are indicated by *asterisks*. The gaps used to optimize the alignment are represented by *hyphens*

promoter. The second one, located at position −476 of the γ-kafirin promoter, is completely conserved in the 28 kDa and presents only one nucleotide substitution in the 16 KDa γ-zein promoter. The conservation of these sequences at identical positions in the γ-prolamin promoters suggests that they may play an important role in the regulation of expression of these genes.

Comparison of the 3′ flanking regions of γ-prolamin genes revealed a high degree of similarity in the upstream but not in the downstream region relative to the poly(A) site. The comparison of the non-translated 3′ regions upstream of the poly(A) site is shown in Fig. 6. The locations of the poly(A) sites in the γ-prolamin genes were deduced from known γ-prolamin cDNA sequences (Prat et al. 1985; Barros et al. 1991; Leite et al. 1991). Based on the alignment with these sequences, five regions containing a putative polyadenylation signal, designated PS-1, PS-2, PS-3, PS-4, and PS-5 in Fig. 6, were identified. The polyadenylation signals of plants are not as well conserved as those of mammalian genes (Joshi 1987; Proudfoot 1991). However a canonical polyadenylation signal of mammals is present at the PS-2 region of the γ-coixin gene. The sequence AATAAT, described as the putative polyadenylation signal in α-zeins (Marks et al. 1985) and α-coixins (Ottoboni et al. 1993) is present at the PS-1 region of γ-kafirin, 16 kDa γ-zein and γ-coixin, and also at the PS-2 region of γ-kafirin and 28 kDa γ-zein genes. The region containing the sequences PS-3, PS-4 and PS-5, located 16 to 42 bp upstream of the poly(A) site, includes, different versions of the AATGAA motif, which was recently indicated as the major polyadenylation signal for the 28 kDa γ-zein (Wu et al. 1993). The AATGAC copy (PS-3), is completely conserved in all the γ-prolamin sequences and is followed by the less conserved PS-4, which is represented by ATTGAA, GATGAA and AATGAA in the γ-kafirin, γ-coixin and γ-zeins, respectively. The AATGAA copy (PS-5) is completely conserved in all the γ-prolamins except for the 16 kDa γ-zein, which contains the AATGGA version. The finding of multiple poly(A) sites within the PS-5 region of the γ-coixin cDNAs indicates that either the PS-3 or the PS-5 sites may act as polyadenylation signals (data not shown). A conserved 14 nucleotide motif containing the CATGG sequence is located upstream of PS-3 in all γ-prolamin genes. This motif has been found to be essential for 3′ end processing of the 28 kDa γ-zein mRNA (Wu et al. 1993).

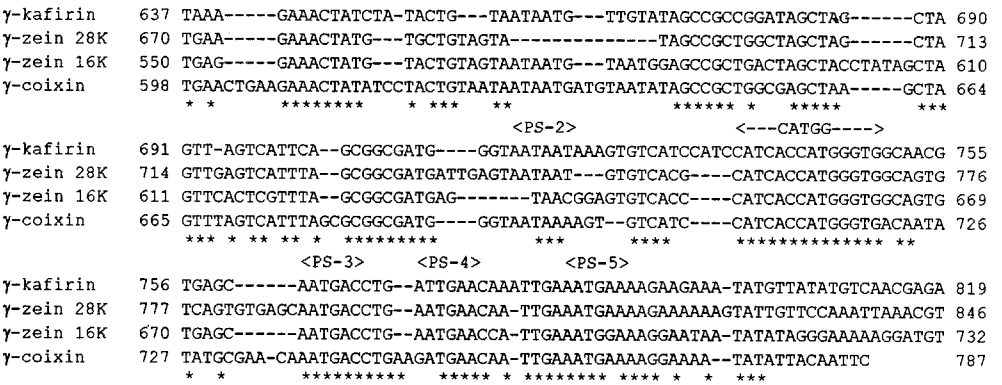

Fig. 6 Alignment of the 3′ flanking regions of γ-prolamins. The 16 and 28 kDa γ-zeins and γ-coixin are represented by the nucleotide sequences of the Zc1 (Reina et al. 1990a), and Zc2 (Reina et al. 1990b) genomic clones and pBCX22.5, respectively. The positions of five regions containing putative polyadenylation signals are indicated by ⟨PS-1⟩, to ⟨PS-5⟩, and ⟨CATGG⟩ indicates a conserved motif essential for mRNA 3′ end processing of 28 kDa γ-zein (Wu et al. 1993). Identical residues are indicated by *asterisks*. The gaps used to optimize the alignment are represented by *hyphens*

Transient expression assays

Microprojectile bombardment of immature seeds was used as a transient expression system to investigate the role of the conserved *cis*-acting regulatory elements of the γ-kafirin promoter. The construct of pPKGUS, containing the *gus* reporter gene fused to a 1190 bp of the γ-kafirin promoter is shown in Fig. 7A. Histochemical analysis of GUS activity in bombarded seeds demonstrated that the γ-kafirin promoter was able to drive equivalent levels of *gus* expression in all three species studied (Fig. 8). Reporter gene activity was located exclusively in the endosperm, indicating that regulatory factors controlling tissue-specific expression of γ-prolamin genes are conserved in sorghum, maize and *Coix*. No expression was observed when seeds were bombarded with a construct containing the same promoter fragment in inverted orientation. Since the activity levels and tissue specificity driven by the γ-kafirin promoter were similar in the three species investigated, and because controlled pollination, a necessary condition for obtaining samples of synchronized seeds, is difficult to achieve in sorghum and *Coix*, further experiments using endosperms were performed in maize seeds.

To analyze the *cis* elements involved in the tissue-specific activity of the γ-kafirin promoter, a series of deleted and chimeric promoter constructs, summarized in Fig. 7, was delivered into coleoptiles, embryos, leaves, and seeds by particle bombardment. The constructs consisted of plasmids harboring the *gus* gene under the control of the intact γ-kafirin promoter (pPKGUS, Fig. 7A) or a deleted version of the γ-kafirin promoter containing only the first 285 nucleotides (pPK285GUS, Fig. 7B). This deleted version includes the putative TATA and CATC boxes and the proximal *GCN4*-like motif, but it lacks other possible upstream *cis*-acting regulatory elements, including the prolamin boxes and the distal *GCN4*-like motif. The level of expression driven by each construct was standardized by comparison with a reference construct containing the *gus* gene fused to the constitutive CaMV 35S promoter (pRT103-GUS, Fig. 7C) (Benfey and Chua 1990). The use of a reference construct should minimize errors caused by differences in area, cell viability, tissue consistency, and competence of cells to be transformed. As shown in Fig. 9, bombardment with the intact γ-kafirin promoter resulted in maximum relative GUS activity in the endosperm. Surprisingly, coleoptiles presented a relatively high level of GUS activity driven by the intact γ-kafirin promoter. The result could mean that at early stages of seedling development, conditions in the cells of the coleoptile are similar to those in the endosperm, thus allowing the expression of γ-prolamin genes. However, Northern blot analysis (data not shown) did not detect mRNA corresponding to γ-zein in the coleoptile. This contrasted with the observation of γ-kafirin promoter activity in this tissue. The promoter activity could be explained by a condition in which the regulatory sequences of the γ-prolamin promoter delivered into the cells are exposed to the *trans*-acting factors present in the coleoptile. Since the gene was amplified in a bacterial plasmid, modifications such as methylation will not occur. This would allow the transcriptional factors present in the coleoptile cells to recognize the regulatory sequences and promote the transcription of the reporter gene. The endogeneous gene might be silent owing to methylation of the target sequences present in the promoter. Embryo and leaf presented only minimal GUS activity close to background level.

The construct containing only the −285 promoter fragment (pPK285GUS) resulted in an approximately 80% reduction in relative GUS activity in endosperm. A similar reduction was also observed for coleoptile but the tissue specificity for endosperm and coleoptile was maintained. These results suggest that the *cis*-acting element(s) responsible for tissue specificity are located in the −285 fragment of the γ-kafirin promoter and may possibly include the *GCN4*-like motif. On the

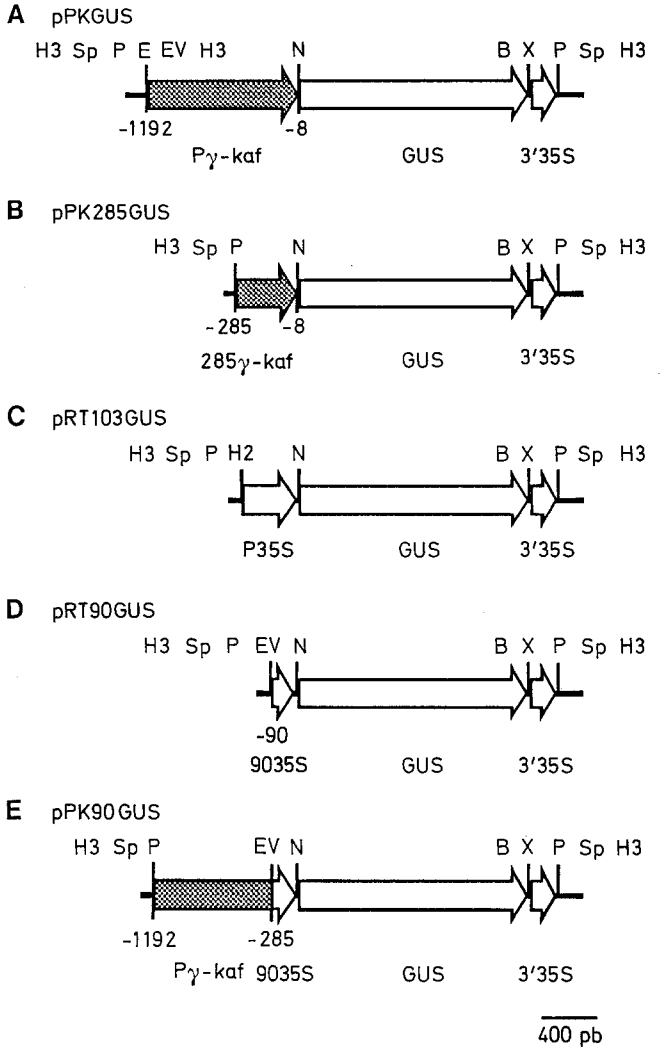

Fig. 7A–E Schematic representation of chimeric GUS (β-glucuronidase) constructs used in the transient expression assays. Distances are relative to the first residue of the initiator ATG codon. Unique restriction site enzymes are represented by: *B Bam*HI, *EV Eco*RV, *H2 Hin*cII, *H3 Hin*dIII, *N Nco*I, *P Pst*I, *Sp Sph*I, *X Xba*I. **A** pPKGUS, the GUS-3′35S reporter under the control of the 1192-nucleotide fragment of the γ-kafirin promoter. **B** pPK285GUS, the GUS-3′35S reporter under the control of the 285 nucleotide fragment of the proximal region of the γ-kafirin promoter. **C** pRT103GUS, the GUS-3′35S reporter under the control of the CaMV 35S promoter. **D** pRT90GUS, the GUS-3′35S reporter under the control of the 90-nucleotide fragment of the proximal region of the CaMV 35S promoter. **E** pPK90GUS, the GUS-3′35S reporter under the control of a chimeric powder constructed by fusion of the proximal region of the CaMV 35S promoter and the distal region of the γ-kafirin promoter

other hand, the elements responsible for the quantitative response seem to be located within the remaining 5′ upstream promoter sequence. To verify whether the reduction of the relative GUS activity of the deleted promoter was due to the absence of quantitative *cis*-acting elements we performed transient expression analysis with a chimeric promoter construct in which the distal promoter region deleted in the construct

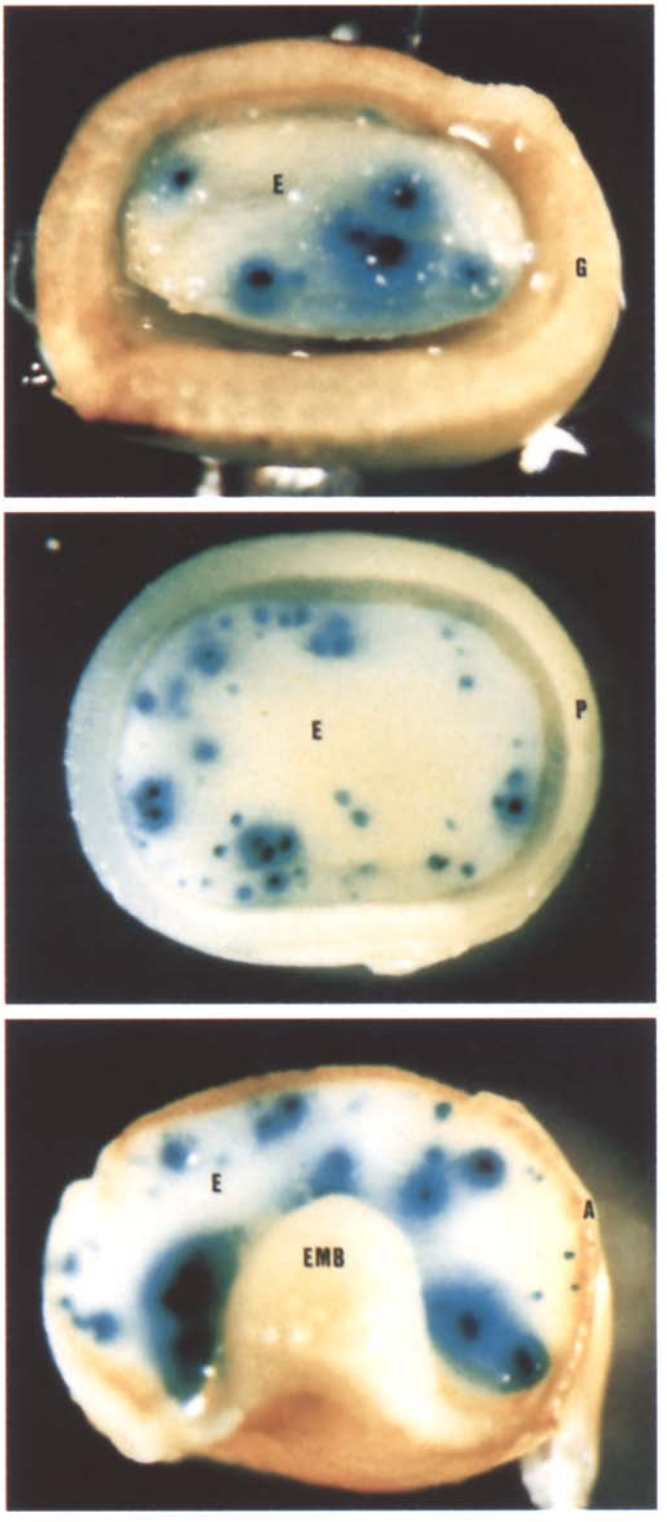

Fig. 8 Transient expression from the γ-kafirin promoter in immature seeds of sorghum (upper part), *Coix* (middle), and maize (lower part) following bombardment with the pPKGUS construct. (*A* aleurone, *E* endosperm, *EMB* embryo, *G* glume, *P* pericarp)

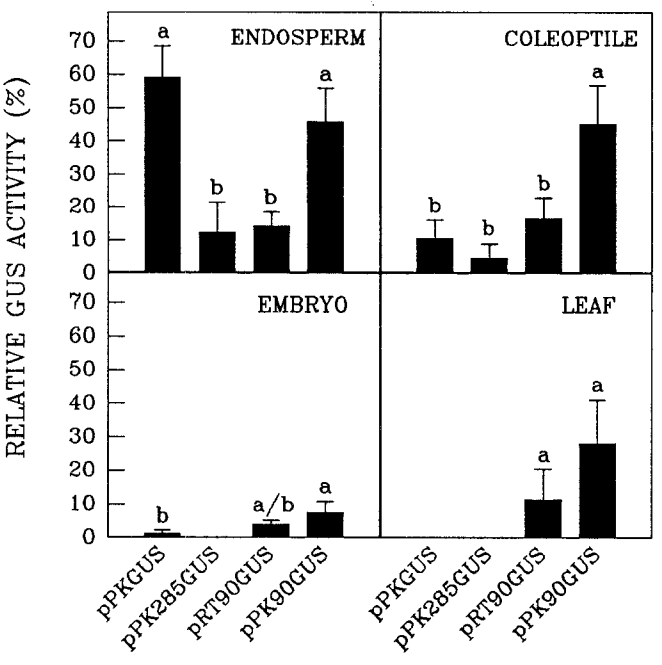

Fig. 9 Transient expression from whole, truncated and chimeric γ-kafirin promoters in different maize tissues. GUS activity was normalized relative to the expression of pRT103GUS in each tissue: 100% activity corresponded to 98 ± 16 blue spots for endosperm, 618 ± 182 for embryo, 829 ± 151 for coleoptile and 131 ± 54 for leaves. The values represent the mean of three independent experiments. *Lowercase letters* over *bars* represent significant differences ($P < 0.05$) as determined by ANOVA and SNK multiple range tests (Steel and Torrie 1980)

pPK285GUS was fused to the 90-bp proximal region of the CaMV 35S promoter (pPK90GUS, Fig. 7E). Figure 9 shows that the addition of the distal region of the γ-kafirin promoter resulted in lack of tissue specificity, and in a three- and twofold increase in the relative GUS activity in endosperms and other tissues, respectively, when compared with the construct containing only the 90-bp proximal region of the CaMV 35S promoter (pRT90GUS, Fig. 7D). Although the CaMV 35S-90 fragment does not correspond to the minimal promoter (Fang et al. 1989), the experimental conditions used gave rise to unequivocal results regarding the activity and tissue specificity of the γ-kafirin promoter. These results confirm that the region upstream of position −285 of the γ-kafirin promoter contains elements involved in quantitative transcriptional regulation.

Acknowledgments This work was financed by grants to P.A. from Financiadora de Estudos e Projetos (FINEP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)/Projeto Temático de Equipe (PTE)-90/3808-5. P.A. and A.L. received research fellowships from CNPq. J.A.Y. and M.J.S. received postgraduate fellowships from CNPq and Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). We thank Dr. Jose C. Novello of the Biochemistry Department of UNICAMP, Campinas, Brazil, for performing the N-terminal sequencing of the mature γ-kafirin. The authors are grateful to Dr. Claude Crétin (Université Paris XI, France) for supplying the sorghum genomic library. We would also like to thank Dr. Pedro Puigdomènech (CSIC, Barcelona, Spain) for the 16 kDa γ-zein cDNA clone (pME119), and Drs. Ana B. Garcia and Eugen S. Gander for reviewing this manuscript.

References

Barros EG, Takasaki K, Kirleis AW, Larkins BA (1991) Nucleotide sequence of a cDNA clone encoding γ-kafirin protein from *Sorghum bicolor*. Plant Physiol 97:1606–1607

Benfey PN, Chua N-H (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250:959–966

Benton WD, Davis RW (1977) Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196:180–182

Brown JWS, Wandelt C, Feix G, Neuhaus G, Schweiger HG (1986) The upstream regions of zein genes – Sequence analysis and expression in the unicellular alga *Acetabularia*. Eur J Cell Biol 42:161–170

Colot V, Robert LS, Kavanagh TA, Bevan MW, Thompson RD (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J 6:3559–3564

Das PD, Messing JW (1987) Allelic variation and differential expression at the 27-kilodalton zein locus in maize. Mol Cell Biol 7:4490–4497

Das PO, Ward K, Ray S, Messing J (1991a) Sequence variation between alleles reveals two types of copy correction at the 27-kDa zein locus of maize. Genomics 11:849–856

Das PO, Poliak E, Ward K, Messing J (1991b) A new allele of the duplicated 27kD zein locus of maize generated by homologous recombination. Nucleic Acids Res 19:3325–3330

DeRose RT, Ma D-P, Kwon I-S, Hasnain SE, Klassy RC, Hall TC (1989) Characterization of the kafirin gene family from sorghum reveals extensive homology with zein from maize. Plant Mol Biol 12:245–256

Esen A (1986) Separation of alcohol-soluble proteins (zeins) form maize into three fractions by differential solubility. Plant Physiol 80:623–627

Esen A, Bietz JA, Paulis JW, Wall JS (1982) Tandem repeats in the N-terminal sequence of a proline-rich protein from corn endosperm. Nature 296:678–679

Evans DJ, Schüssler L, Taylor JRN (1987) Isolation of reduced-soluble protein from sorghum starchy endosperm. J Cereal Sci 5:61–65

Fang R-X, Nagy F, Sivasubramaniam S, Chua N-H (1989) Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell 1:141–150

Gallardo D, Reina M, Rigua J, Boronat A, Palau J (1988) Genomic organization of the 28 kDa glutelin-2 gene from maize. Plant Sci 54:211–218

Gallie DR, Sleat DE, Wats JW, Turner PC, Wilson TMA (1987) A comparison of eukaryiotic viral 5′-leader sequences as enhancers of mRNA expression *in vivo*. Nucleic Acids Res 15:8693–8711

Hammond-Kosack MCU, Holdsworth MJ, Bevan MW (1993) *In vivo* footprinting of a low molecular weight glutenin gene (LMWT-1D1) in wheat endosperm. EMBO J 12:545–554

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5:151–153

Hill DE, Hope IA, Macke JP, Struhl K (1986) Saturation mutagenesis of the yeast *his3* regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. Science 234:451–457

Hull GA, Halford NG, Kreis M, Shewry PR (1991) Isolation and characterization of genes encoding rye prolamins containing a highly repetitive sequence motif. Plant Mol Biol 17:1111–1115

Jefferson RA (1987) Assaying chimeric genes in plants: The GUS gene fusion system. Plant Mol Biol 5:387–405

Johnson DA, Gautsch JW, Sportsman JR, Elder JH (1984) Improved technique utilizing non-fat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal Techniques 1:3–8

Joshi CP (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucleic Acids Res 15:9627–9640

Kirihara JA, Husperger JP, Mahoney JW (1988) Differential expression of a gene for a methionine-rich storage protein in maize. Mol Gen Genet 211:477–484

Kreis M, Williamson MS, Forde J, Schmutz D, Clark J, Buxton B, Pywell J, Marris C, Henderson J, Harris N, Shewry PR, Forde BG, Miflin BJ (1986) Differential gene expression in the developing barley endosperm. Philos Trans R Soc Lond [Biol] 314:355–356

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–684

Leite A, Ottoboni LMM, Targon MLPN, Silva MJ, Arruda P (1990) Phylogenetic relationships of zeins and coixins as determined by immunological cross-reactivity and Southern Blot analysis. Plant Mol Biol 14:743–751

Leite A, Freitas FA, Yunes JA, Arruda P (1991) Nucleotide sequence of a cDNA clone encoding γ-coixin from *Coix lacryma-jobi* seeds. Plant Physiol 97:1604–1605

Leite A, Yunes JA, Turcinelli SR, Arruda P (1992) Cloning and characterization of a cDNA encoding a sulfur-rich coixin. Plant Mol Biol 18:171–174

Liu C-N, Rubenstein I (1992) Molecular characterization of two types of 22 kilodalton α-zein genes in a gene cluster in maize. Mol Gen Genet 234:244–253

Marks MD, Lindell JS, Larkins BA (1985) Nucleotide sequence analysis of mRNAs from maize endosperm. J Biol Chem 260:16451–16459

Moreno MR, Smith JF, Smith RV (1985) Silver staining of proteins in polyacrylamide gels: Increased sensibility through a combined Coomassie Blue-silver stain procedure. Anal Biochem 151: 466–470

Müller M, Knudsen S (1993) The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and prolamin box. Plant J 4:343–355

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol Plant 15:473–497

Ottoboni LMM, Leite A, Yunes JA, Targon MLPN, Souza Filho GA, Arruda P (1993) Sequence analysis of 22 kDA-like α-coixin genes and their comparison with homologous zein and kafirin genes reveals highly conserved protein structure and regulatory elements. Plant Mol Biol 21:765–778

Pedersen K, Argos P, Naravana SLV, Larkins BA (1986) Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of M_r 15000. J Biol Chem 14:6279–6284

Plikaytis BD, Carlone GM, Edwards P, Mayer W (1986) Robust estimation of standard curves for protein molecular weight and linear-duplex DNA based pair number after gel electrophoresis. Anal Biochem 152:346–364

Prat S, Cortadas J, Puigdomenech P, Palau J (1985) Nucleic Acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2. Nucleic Acids Res 13:1493–1504

Prat S, Pèrez-Grau L, Puigdomènech P (1987) Multiple variability in the sequence of a family of maize endosperm proteins. Gene 52:41–49

Proudfoot N (1991) Poly(A) signals. Cell 64:671–674

Quayle T, Feix G (1992) Functional analysis of the −300 region of maize zein genes. Mol Gen Genet 231:369–374

Quayle TJA, Hetz W, Feix G (1991) Characterization of a maize endosperm culture expressing zein genes and its use in transient transformation assays. Plant Cell Rep 9:544–548

Rackwitz HR, Zehetner G, Fischauf AM, Lehrach H (1984) Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene 30:195–200

Reina M, Guillén P, Ponte I, Boronat A, Palau J (1990a) DNA sequence of the gene encoding the Zc1 protein from *Zea mays* W64 A. Nucleic Acids Res 18:6425

Reina M, Ponte I, Guillén P, Boronat A, Palau J (1990b) Sequence analysis of a genomic clone encoding a Zc2 protein from *Zea mays* W64 A. nucleic Acids Res 18:6426

Rivin CJ, Zinner EA, Walbot V (1982) Isolation of DNA and -1DNA recombinants from maize. In: Sheridan WF (ed) Maize for biological research. University Press, University of North Dakota, Grand Forks, pp 161–164

Shewry PR, Tatham AS (1990) The prolamin storage of cereal seeds: structure and evolution. Biochem J 267:1–12

Steel RGD, Torrie JH (1980) Principles and procedures of statistics. A biometrical approach. McGraw-Hill, New York, pp 172–194

Summer-Smith M, Rafalski JA, Sugyama T, Stoll M, Söll D (1985) Conservation and variability of wheat α/β-gliadin genes. Nucleic Acids Res 13:3905–3916

Taylor JRN, Von Benecke R, Carlsson FHH (1989) Distribution, purification and N-terminal amino acid sequence of sorghum reduced-soluble protein. J Cereal Sci 9:169–177

Thompson GA, Boston RS, Lzynik LA, Hodges TK, Larkins BA (1990) Analysis of promoter activity from an α-zein gene 5′ flanking sequence in transient expression assays. Plant Mol Biol 15:755–764

Töpfer R, Matzeit V, Gronenborn B, Schell J, Steinbiss H-H (1987) A set of plant expression vectors for transcriptional and translational fusions. Nucleic Acids Res 15:5890

Wang S-Z, Esen A (1986) Primary structure of a prolamin-rich zein and its cDNA. Plant Physiol 81:70–74

Wu L, Ueda T, Messing (1993) 3′-end processing of the maize 27 kDa zein mRNA. Plant J 4:535–544

Ye G-N, Daniell H, Sanford JC (1990) Optimization of delivery of foreign DNA into higher plant chloroplasts. Plant Mol Biol 15:809–819