The Plant Cell, Vol. 9, 171–184, February 1997 © 1997 American Society of Plant Physiologists

The Wheat Transcriptional Activator SPA: A Seed-Specific bZIP Protein That Recognizes the GCN4-like Motif in the Bifactorial Endosperm Box of Prolamin Genes

Diego Albani, Michael C. U. Hammond-Kosack, Caroline Smith, Steven Conlan, Vincent Colot,[1] Michael Holdsworth,[2] and Michael W. Bevan[3]

Molecular Genetics Department, John Innes Centre, Colney Lane, Norwich NR4 7UJ, United Kingdom

The conserved bifactorial endosperm box found in the promoter of wheat storage protein genes comprises two different *cis* elements that are thought to be involved in regulating endosperm-specific gene expression. Endosperm nuclear extracts contain binding activities. One is called ESBF-I, which binds to the endosperm motif (EM), and the other is called ESBF-II, which binds to the GCN4-like motif (GLM). Here, we present a functional analysis of the endosperm box of a low-molecular-weight glutenin gene found on the 1D1 chromosome of hexaploid wheat (*LMWG-1D1*) in transgenic tobacco plants. Our analysis demonstrates the necessity of the EM and GLM for endosperm-specific gene expression and suggests the presence in tobacco of functional counterparts of wheat ESBF-I and ESBF-II. Furthermore, we describe the isolation and characterization of cDNA clones encoding SPA, a seed-specific basic leucine zipper protein from wheat that can activate transcription from the GLMs of the −326-bp *LMWG-1D1* promoter in both maize and tobacco leaf protoplasts. This activation is also partially dependent on the presence of functional EMs, suggesting interactions between SPA with ESBF-I–like activities.

INTRODUCTION

Developing seeds of flowering plants accumulate and store carbohydrates, lipids, and proteins that subsequently are used by the germinating seedlings as a source of energy, carbon, and nitrogen. Seed storage proteins are synthesized in the endosperm of monocotyledonous seeds and in the embryonic axis and cotyledons of dicotyledonous seeds. These seed storage proteins are not synthesized in other tissues. The major seed storage proteins in wheat (the most important and widely grown crop plant) are the prolamins, which account for >90% of the total protein content of the grain. Wheat prolamins consist of two major groups of proteins, the gliadins and the glutenins, which differ in their ability to form polymers. The gliadins are a complex mixture of monomeric proteins subdivided into α-, β-, γ-, and ω-gliadins on the basis of their electrophoretic mobilities. The glutenins are composed of high-molecular-weight and low-molecular-weight (LMW) subunits that assemble into high-mass polymers stabilized by intermolecular disulfide bonds (Colot, 1990). Comparisons of the promoter sequences of various cereal prolamin genes have identified a conserved region located ~300 bp upstream of the transcriptional start; this region has been named the endosperm box (Kreis et al., 1985). In members of the grass tribe Poideae, which includes wheat, barley, and rye, the conserved region contains two different conserved motifs, called the endosperm motif (EM) and the GCN4-like motif (GLM), which are very similar to the binding sites of the yeast GCN4 and mammalian AP1 factors (Müller and Knudsen, 1993). The sequence of these elements is shown in Figure 1A.

The development of efficient transformation protocols for cereals has recently permitted the first analysis of a rice seed storage protein glutenin promoter in a stable homologous transgenic system (Zheng et al., 1993), but detailed promoter analyses of cereal storage protein genes have only been performed in model dicot species such as tobacco (Colot et al., 1987; Marris et al., 1988; Schernthaner et al., 1988; Robert et al., 1989; Thomas and Flavell, 1990; Takaiwa et al., 1991). Expression in the heterologous dicot host was restricted to the endosperm, suggesting a conservation of the *cis* elements responsible for cell-type specificity in monocots and dicots. Furthermore, deletion analysis of a wheat LMW glutenin (*LMWG-1D1*) promoter demonstrated that a fragment located from positions −326 to −160 and containing the bifactorial endosperm box was essential for expression in tobacco endosperm (Colot et al., 1987). In vivo dimethyl sulfate footprinting of the same LMW glutenin

[1] Current address: Institut de Genetique et Microbiologie, Université Paris Sud, 91405 Orsay Cedex, France.
[2] Current address: Institute of Arable Crop Research, Long Ashton, Bristol BS18 9AF, UK.
[3] To whom correspondence should be addressed.

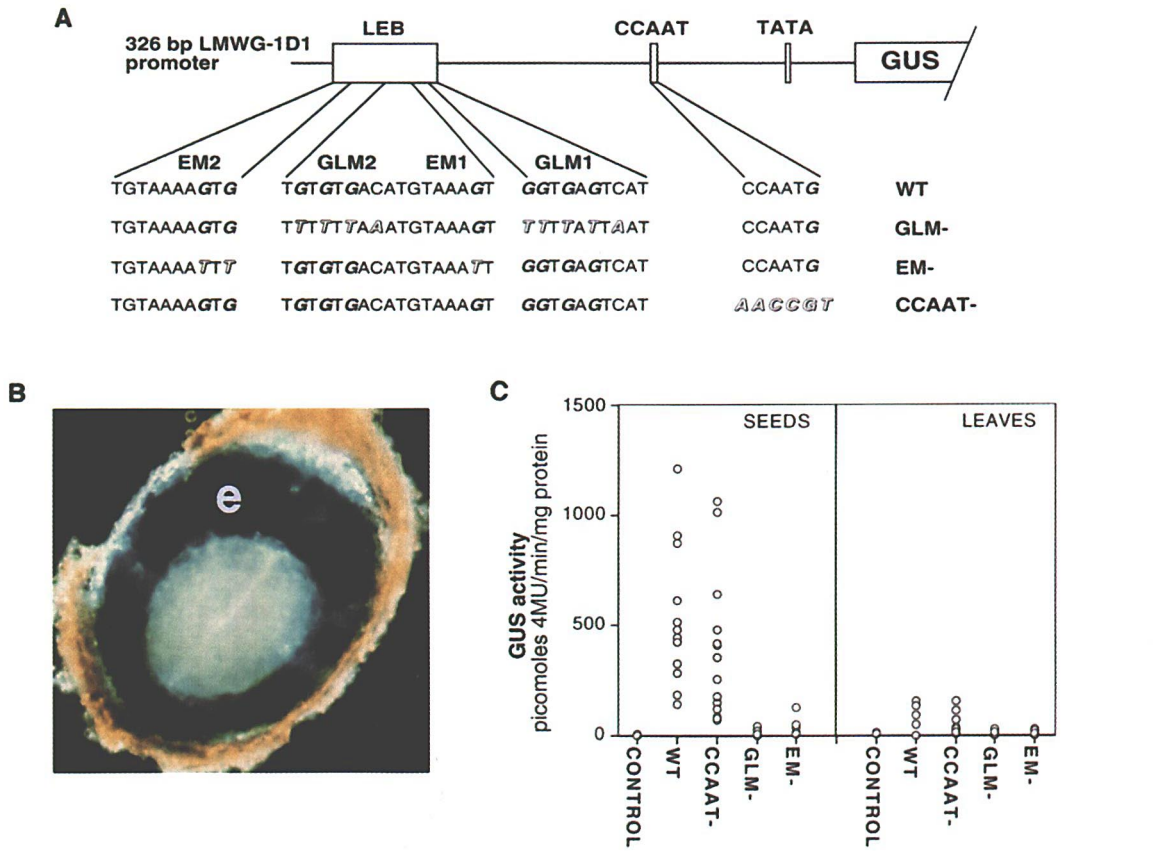

Figure 1. Functional Analysis of the −326-bp *LMWG-1D1* Promoter in Transgenic Tobacco Plants.

(A) Structure of the promoter fused to the *GUS* marker gene and sequences of the putative *cis* elements in the wild-type and mutated promoter constructs. The G residues previously footprinted in vivo (Hammond-Kosack et al., 1993) are shown in boldface italics, whereas the mutated residues are outlined. WT, wild type; GLM−, mutated GLM; EM−, mutated EM; CCAAT−, mutated CCAAT motif.

(B) Histochemical analysis of GUS activity in a cross-section of a tobacco seed transformed with the wild-type −326-bp *LMWG-1D1* promoter construct. e, endosperm.

(C) GUS activity in extracts from seeds and leaves of multiple independent tobacco transformants containing various promoter constructs. GUS activities in untransformed tobacco plants are shown for reference. The analysis was performed on seeds collected 15 DAA. 4MU, 4-methylumbelliferone.

promoter in developing wheat endosperm detected the sequential binding of putative *trans*-acting factors to the EM and GLM (Hammond-Kosack et al., 1993). This work identified an additional footprinted copy of each motif in an adjacent upstream region of the promoter that, together with the downstream elements, forms a so-called long endosperm box (LEB). Electrophoretic mobility shift assays (EMSAs) using wheat endosperm nuclear protein extracts showed the specific binding of two different factors, ESBF-I and ESBF-II, to the EM and GLM, respectively (Hammond-Kosack et al.,1993).

The endosperm box of the prolamin genes of members of the grass tribe Andropogonoideae, which includes maize, sorghum, and Coix, contains a conserved EM. In vitro footprinting experiments have shown that the EM of a zein promoter is bound by a nuclear protein from maize endosperm nuclear extracts (Maier et al., 1987). This binding activity may correspond to the maize counterpart of wheat ESBF-I. Maize *opaque2 (o2)* mutants show an overall reduction of α-zein content, with a particularly strong effect on the expression of the 22-kD α-zeins (Kodrzycki et al., 1989). O2 is an endosperm-specific member of the basic leucine zipper (bZIP) family of transcription factors that has been shown to bind specifically the 22-kD zein promoter and activate transcription in vitro (Schmidt et al., 1992). O2 has also been shown to activate endosperm-specific transcription of the b-32 gene, a type I ribosome-inactivating protein, via a distinct *cis* sequence (Lohmer et al., 1991). Several lines of evidence suggest that O2 may be functionally related to the

wheat ESBF-II factor. First, O2 can bind to the GLM of the LMW glutenin endosperm box and activate transcription in yeast and plants through this element (Holdsworth et al., 1995). Second, O2 expressed in yeast complements a *gcn4* mutant by binding and activating the *HIS3* gene at its GCN4 binding site (Mauri et al., 1993), indicating that the GCN4 motif, the GLM sequence, and the O2 binding sites in maize genes are functionally related. Finally, the recombinant O2 protein has been shown to bind in vitro to a GLM sequence of a pea lectin gene that is necessary and sufficient for seed-specific expression in transgenic tobacco plants (de Pater et al., 1993, 1994).

In this study, we describe a functional analysis of the bifactorial endosperm box of the wheat *LMWG-1D1* promoter in transgenic tobacco plants that defines the importance of both the EM and GLM for endosperm-specific expression and the LEB as a functional unit. In addition, we describe the isolation and characterization of cDNA clones encoding SPA (for storage protein activator), a novel seed-specific transcriptional activator from wheat that shows partial sequence similarity to the maize O2 transcription factor. SPA can bind to both GLMs of the LEB and is able to transactivate a reporter gene linked to a minimal promoter containing the LEB sequence. SPA requires an intact EM motif for maximal transcriptional activation. In addition, SPA can activate transcription of a maize 22-kD zein gene, and in the transient assay systems employed, SPA and O2 activated transcription to similar levels.

RESULTS

Functional Analysis of the *LMWG-1D1* Endosperm Box in Transgenic Tobacco Plants

A deletion analysis of the *LMWG-1D1* promoter in transgenic tobacco previously indicated that sequences located between 326 and 160 bp upstream of the transcriptional start site were essential and sufficient for endosperm-specific expression (Colot et al., 1987). The presence of a conserved endosperm box in this region indicated a possible involvement of this sequence in tobacco endosperm-specific expression. Mutations corresponding to in vivo–footprinted G residues in the endosperm box (Hammond-Kosack et al., 1993) were introduced in the EMs or GLMs of the LEB in a −326-bp *LMWG-1D1* promoter–β-glucuronidase (*GUS*) gene fusion. These mutated promoter constructs are shown in Figure 1A. In an additional construct, mutations were introduced in a putative CCAAT box sequence containing a footprinted G residue of the *LMWG-1D1* promoter (Hammond-Kosack et al., 1993). The various mutated promoter constructs were cloned in the binary vector pBin19 (Bevan, 1984) and introduced into tobacco by Agrobacterium-mediated transformation. The promoter sequences of the various constructs in trangenic plants were verified from a few randomly selected transgenic plants by polymerase chain reaction (PCR) amplification and direct sequencing.

Quantitative and histochemical analyses of GUS activity in seeds of transgenic plants containing the wild-type promoter construct were performed. As shown in Figure 1B, the *LMWG-1D1–GUS* fusion was specifically expressed in the endosperm of transgenic seeds. GUS activities were determined in extracts from leaves and seeds of approximately nine to 15 independent transformants for each construct at 15 days after anthesis (DAA) and are summarized in Figure 1C. Similar relatively high levels of GUS activity were observed in the seeds of plants transformed with the wild type and CCAAT box mutated constructs, whereas plants transformed with constructs containing mutations in the EMs or the GLMs showed only background levels of expression.

It appears that the mutations introduced at the CCAAT box did not significantly affect the activity of the −326-bp *LMWG-1D1* promoter in transgenic tobacco seeds. Very low levels of *GUS* expression were detected in the leaves of some of the plants transformed with the wild type and the CCAAT promoter constructs. Therefore, it is possible that the activity of the −326-bp *LMWG-1D1* promoter in transgenic tobacco is not as strictly seed specific as are longer promoter constructs (Colot et al., 1987). It is clear that the two conserved motifs contained in the bifactorial endosperm box are essential for expression in tobacco endosperm and are probably recognized by tobacco homologs of ESBF-I and ESBF-II, because the introduced promoter mutations, based on in vivo footprinting data, specifically abolished in vitro binding of the ESBF-I and ESBF-II activities (Hammond-Kosack et al., 1993).

Isolation of a cDNA Encoding a bZIP Protein from Wheat Endosperm

Because of the potential functional similarity between maize O2 and the wheat ESBF-II endosperm nuclear factor, finding transcripts encoding an O2-like protein in wheat seeds might eventually lead to the isolation of cDNAs encoding the ESBF-II activity. But RNA gel blot analyses performed at low stringency with wheat seed poly(A)$^+$ and an *O2* cDNA probe failed to show cross-hybridizing bands. Furthermore, an initial attempt to isolate cDNAs encoding a wheat O2-like bZIP protein by performing a low-stringency screening of a wheat endosperm cDNA library was unsuccessful. These results did not rule out the presence of a wheat O2-like bZIP protein showing only very limited homology to the maize factor or very low expression in endosperm; therefore, a PCR approach to isolating a cDNA encoding a putative O2-like protein from wheat endosperm was adopted. Two different degenerate primers, derived from DNA sequence encoding the amino acid sequence NRE(S/A)A, which is highly conserved in the basic region of various plant bZIP proteins

were used in 3′ RACE (for rapid amplification of cDNA ends) reactions performed with wheat seed poly(A)$^+$ mRNA. As a positive control, similar reactions were performed using maize seed poly(A)$^+$ mRNA. The resulting PCR products were resolved electrophoretically and hybridized with an *O2* probe. The wheat 3′ RACE sample obtained with one of the degenerate primers showed a faint but clear hybridization of a band ~750 bp, which demonstrated the presence of an O2-like bZIP in wheat seeds (data not shown). Hybrid selection of the PCR products by using an immobilized *O2* probe, followed by PCR amplification, enriched the *O2*-specific fragment sufficiently for purification and cloning.

Ten individual clones were sequenced, and all corresponded to nearly identical RACE products encoding the C terminus of a bZIP protein (data not shown). This partial protein showed only limited sequence similarity to O2 (46% identity) but had a much more striking homology to the maize OHP1 and OHP2 proteins (84% identity). OHP1 and OHP2 are not specifically expressed in endosperm, but they show considerable sequence similarity to O2 in their bZIP region and have been shown to heterodimerize with O2 (Pysh et al., 1993). The partial-length cDNA clone was tentatively called wOHP because of its close sequence similarity to the maize OHP clones.

To isolate full-length wOHP clones, we screened a wheat endosperm cDNA library in λgt10 with the partial wOHP cDNA insert. The screening was performed at a slightly reduced stringency to avoid discriminating against possible homologous classes of bZIP proteins. Two longer cDNA clones of different size that, according to partial sequencing data, had an identical overlapping nucleotide sequence, were isolated. Surprisingly, they did not correspond to longer wOHP cDNA clones but encoded a potential bZIP protein different from wOHP, which subsequently was designated SPA. The complete nucleotide sequence of the longest cDNA was determined and is shown in Figure 2, together with the deduced amino acid sequence of the SPA protein. The shorter cDNA starts seven nucleotides downstream of the 5′ end of the longer one, and its smaller size is largely due to the different position of its poly(A) tail, which is located 85 nucleotides 5′ of the other. The longest SPA cDNA is 1647 bp long, excluding the poly(A) tail, and contains an open reading frame encoding a potential protein of 409 amino acids with a predicted molecular mass of 44.4 kD. The ATG start codon of this open reading frame is preceded by a 174-bp-long 5′ untranslated sequence containing an in-frame TAA stop triplet. An ATG triplet, followed shortly by an in-frame stop codon, is found at the beginning of the 5′ untranslated sequence.

The SPA protein shows limited but significant sequence similarity to O2, OHP1, and OHP2 from maize and to the O2 homologs Sbopa2 and Cljop2 from sorghum and Coix, plants closely related to maize (Pirovano et al., 1994). SPA also shows limited homology to the rice seed–specific bZIP protein Osbzippa (GenBank accession number D78609). As shown in Figure 3, the most conserved region of these cereal bZIP proteins is centered around the basic and leucine zipper domains. The overall amino acid sequence identity between SPA and these other cereal bZIP proteins ranges from 44% with OHP2, 41% with Osbzippa, and 31% with O2, whereas the similarity of their bZIP regions (amino acid 214 to 287 of SPA) ranges from 61 to 69% amino acid iden-

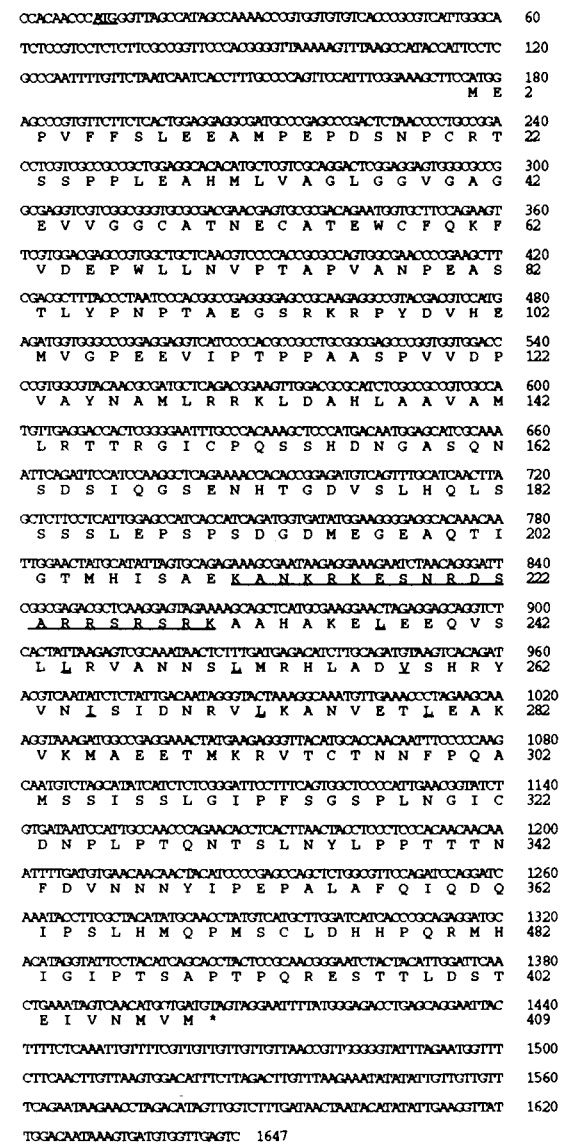

Figure 2. Nucleotide Sequence and Deduced Amino Acid Sequence of SPA cDNA.

The basic region and the periodic leucine, valine, and isoleucine residues in the putative leucine zipper of the SPA protein as well as the ATG start codon in the 5′ untranslated leader sequence are underlined. The asterisk indicates the stop codon.

tity. In addition, limited sequence similarity (38% overall amino acid identity and 63% amino acid identity in their bZIP regions) was found between SPA and common plant regulatory factor 2 (CPRF-2), a parsley bZIP protein that binds to the G-box of the chalcone synthase promoter (Weisshaar et al., 1991). SPA has lower sequence similarity to other wheat bZIP proteins such as HBP-1 (23%; Tabata et al., 1989) and EmBP-1a (22%; Nui and Guiltinan, 1994) than to O2 and O2-like proteins. In addition, Figure 3 shows that SPA and the other cereal bZIP proteins share some small conserved domains outside their bZIP regions, which extend predominantly toward the N terminus and which are not conserved in the parsley protein. These small conserved regions are relatively rich in negatively charged amino acids, a feature associated with acidic activation domains of transcription factors.

SPA Transcripts Are Specifically Expressed in Wheat Seeds

The temporal and spatial pattern of expression of a transcription factor can indicate the tissue specificity and the timing of expression of the genes that it regulates. Therefore, the tissue-specific expression of SPA and potential target genes, such as *LMWG-1D1*, were compared using an RNA gel blot of poly(A)$^+$ RNA isolated from wheat seeds harvested at five different developmental stages, according to the Zadoks scale (Zadoks et al., 1974), ranging from 10 to 28 DAA, and poly(A)$^+$ RNA from wheat leaves, young roots, and coleoptiles. After hybridization with the SPA probe, the same blot was successively hybridized with a LMW glutenin probe and a ubiquitin probe as a control for sample uniformity. The control hybridization with the ubiquitin probe showed a progressive reduction of the abundance of ubiquitin transcripts in the five seed poly(A)$^+$ samples (Figure 4). A similar result was obtained with total RNA samples, and the reduction most likely reflects a reduction in the expression of these transcripts during seed development rather than a reduction in the amount of poly(A)$^+$ RNA loaded on the gel. Figure 4 shows that both the LMW glutenin and SPA mRNAs can only be detected in the seed poly(A)$^+$ samples. Even after extended exposure of the blot, SPA expression could not be seen in any of the vegetative tissues analyzed (data not shown). The SPA mRNA was first detected 10 DAA and increased moderately in abundance in the following stages, peaking at 18 DAA (stage 3) before declining slightly as seeds approached maturity. The temporal pattern of expression of SPA is different from that of the LMW glutenin genes, which were expressed at low levels 10 DAA and progressively increased in the following stages. The observation that SPA expression was restricted to the seeds and preceded transcription of the LMW glutenin genes indicates a potential involvement of this bZIP protein in activation of the glutenin gene.

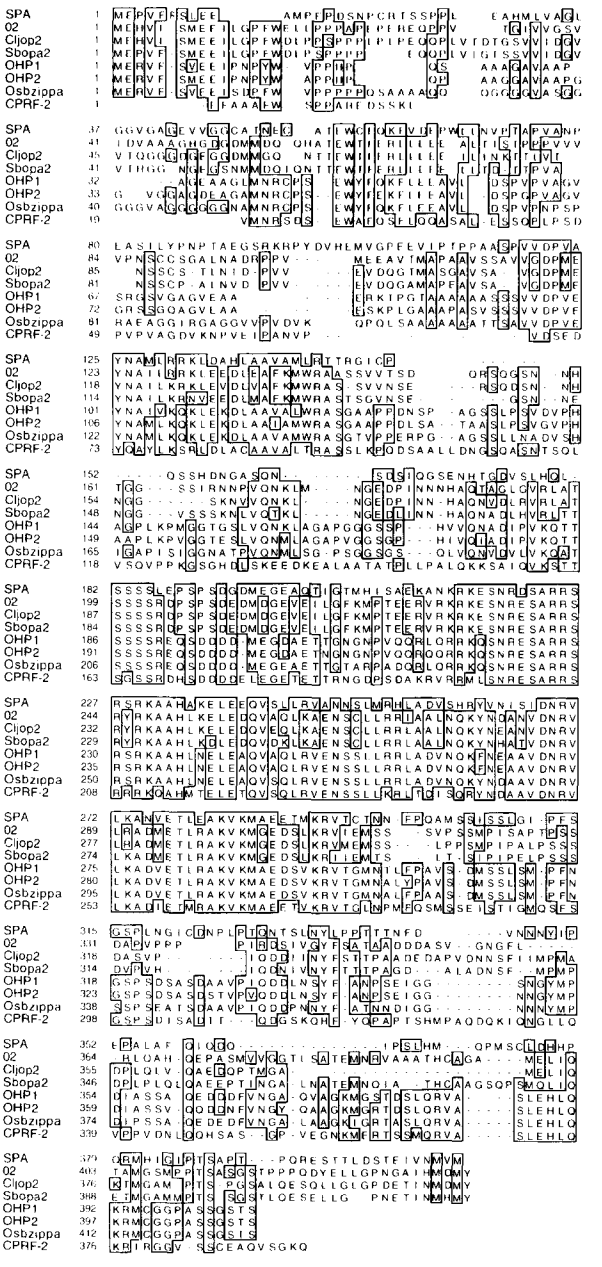

Figure 3. Amino Acid Sequence Alignment of bZIP Proteins, Some of Which Regulate Cereal Seed-Specific Expression.

SPA, O2, the O2-like proteins Cljop2 of Coix, Sbopa2 of sorghum, the parsley CPRF-2 protein, the rice Osbzippa protein, and the maize OHP1 and OHP2 proteins are shown. The residues conserved in all of the eight proteins are contained within the boxes. Dashes were used to optimize alignment.

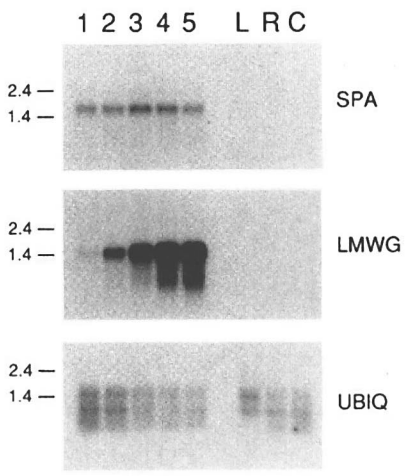

Figure 4. Developmental and Spatial Pattern of Expression of SPA mRNA.

RNA gel blot analysis was performed with poly(A)$^+$ RNA (500 ng) isolated from developing wheat seeds staged according to the Zadoks scale (Zadoks et al., 1974) and on poly(A)$^+$ RNA from leaves (L), roots (R), and coleoptiles (C). After hybridization with the SPA probe, the same blot was subsequently hybridized with an LMW glutenin (LMWG) probe and a ubiquitin probe from Antirrhinum (UBIQ). The positions of the 2.4- and 1.4-kb RNA markers are shown at left.

SPA Protein Can Bind in Vitro to the GCN4-like Motifs of the *LMWG-1D1* Endosperm Box

To verify the potential involvement of SPA in the regulation of glutenin genes, we assessed the binding of SPA to potential target sites in the *LMWG-1D1* promoter. A recombinant SPA protein containing an N-terminal fusion of six histidine residues was expressed in *Escherichia coli* by using a single NcoI restriction site at the putative translational start to clone the SPA coding region in the expression vector pRSET-A (Invitrogen, San Diego, CA). As shown in Figure 5A, isopropyl β-D-thiogalactoside treatment induced expression of a protein of ~55 to 60 kD that was not present in uninduced cell extracts. The apparent size of the recombinant protein determined by SDS-PAGE was slightly larger than the expected 49-kD molecular mass of the His-tagged SPA fusion protein. A similar apparent increase in size on SDS gels is not uncommon for recombinant proteins containing His tags because it has been previously reported for another bacterially expressed recombinant bZIP protein (Niu and Guiltinan, 1994).

The recombinant His–SPA protein was purified to near homogeneity by nickel ion chelate chromatography and used directly for DNA binding assays. Previous studies using another His-tagged bZIP protein (Niu and Guiltinan, 1994) indicated that the His tag at the N terminus of SPA might not affect its DNA binding specificity.

The binding of His–SPA protein to the GLMs of the *LMWG-1D1* endosperm box was determined by EMSAs. The probe used in EMSAs corresponded to the LEB region extending from positions −302 to −231 in the *LMWG-1D1* promoter (wild type) containing two intact GLMs and a probe (GLM−) containing mutations in both GLMs corresponding to those shown in Figure 1A. These probes had been used previously in gel shift assays with endosperm nuclear extracts to demonstrate binding of the ESBF-II activity to the GLMs (Hammond-Kosack et al., 1993). Figure 5B shows that the His–SPA fusion protein binds to the wild-type probe but not to the GLM− sequence. The two protein–DNA complexes formed with the wild-type probe could possibly be derived from the binding of SPA homodimers at only one of the two GLMs in the probe, whereas the other complex could represent binding at both motifs. Previously, Hammond-Kosack et al. (1993) proposed the binding of the same activity, ESBF-II, to both GLMs, but no direct evidence for this was available. To determine whether SPA could bind to both motifs, we used oligonucleotide probes corresponding to the individual GLMs in their wild type or mutated forms in EMSAs. Figure 5C shows that both wild-type probes GLM1 and GLM2 give rise to a similar retarded complex that was not observed with the mutant probes. The GLM1 probe showed an additional, faster migrating complex that was specific to the GLM1 wild-type probe, but the significance of this retarded complex is unclear. The similar mobilities of the His–SPA complexes with the wild-type LEB probe (Figure 5B) and with the individual GLM1 and GLM2 probes indicate that the two bound complexes formed on the LEB probe probably correspond to His–SPA binding to both GLMs in the LEB probe.

Transient Expression of the SPA Protein Transactivates from the GCN4-like Motifs of the *LMWG-1D1* Promoter

Having determined the in vitro binding of SPA homodimers to the GLMs of the *LMWG-1D1* LEB, protoplast transient expression assays were performed to determine whether the SPA protein could activate expression from these target sites in vivo. The expression assays were conducted with protoplasts from maize black Mexican sweet (BMS) cell suspension cultures and from tobacco mesophyll cells. Figure 6A shows the reporter constructs used in this analysis. The wild type and versions of the *LMWG-1D1* LEB containing mutations of the EM sequences (EM−) and mutations of the GLM1 and GLM2 sequences separately or together were fused in the native orientation 5′ of a −67-bp cauliflower mosaic virus (CaMV) 35S promoter–*GUS* reporter gene. In transactivation experiments, the effector plasmid consisted of the entire SPA coding region placed under the control of the double CaMV 35S promoter in the expression plasmid pJIT60. All experiments were performed in replicates of five, using the same batch of protoplasts, and each experiment

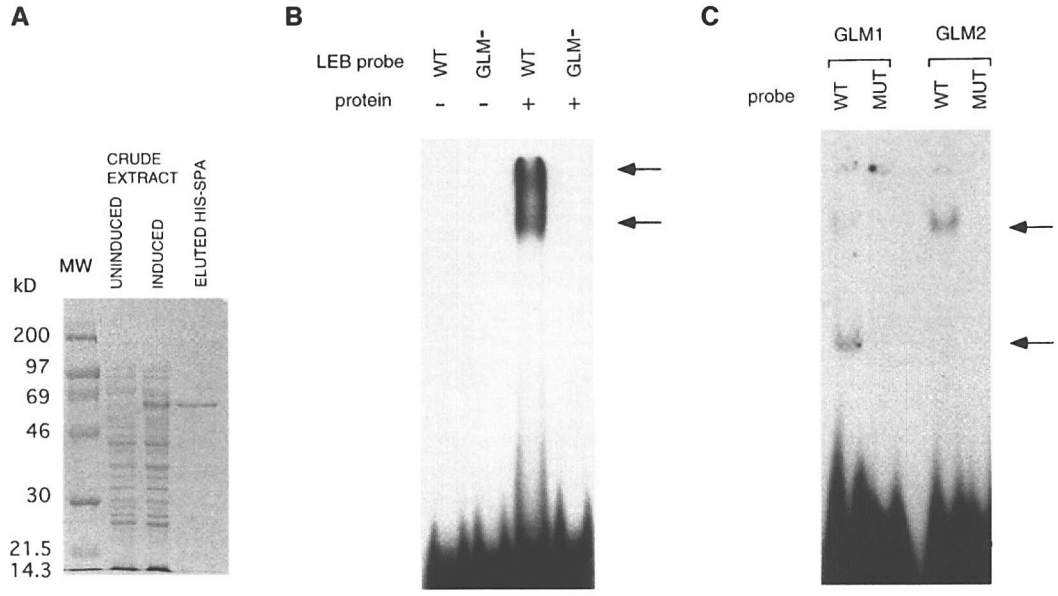

Figure 5. Binding of Recombinant SPA to the GLMs in the LEB of the *LMWG-1D1* Promoter.

(A) Expression and purification of recombinant HIS–SPA protein. Crude extracts from uninduced and induced bacteria harboring the pHIS-SPA expression vector were resolved on an SDS–polyacrylamide gel along with an aliquot (1 μg) of the eluted protein. The molecular mass markers are indicated at left in kilodaltons.

(B) EMSA showing the binding of the HIS–SPA protein to the wild-type (WT) LEB probe. Binding was abolished by the mutations introduced into the two GLMs (GLM−), as described in Figure 1A. The arrows show the positions of the two LEB–SPA complexes. (−) indicates absence and (+) indicates presence of the LEB–SPA protein.

(C) Binding of the HIS–SPA protein to the individual wild-type (WT) GLMs as shown by EMSA. Neither of the mutated GLM probes (MUT) is recognized by the recombinant protein. The single GLM2–SPA complex and the two GLM1–SPA complexes are indicated by arrows. The sequences of the individual probes are given in Methods.

was repeated between three and six times with consistent results. Differences in GUS enzyme levels were significant at $P < 0.01$.

Figure 6B shows that SPA was able to transactivate the wild-type LEB–CaMV 35S construct (4.5-fold; section i) in tobacco mesophyll cells. In BMS cells, SPA transactivated the wild-type LEB–CaMV 35S construct 11-fold (Figure 6C, section i). Transactivation in tobacco cells and in BMS protoplasts depended on the presence of both copies of the GLMs in the LEB, as shown in Figures 6B and 6C, sections ii, iv, and v. In addition, mutation of the EMs reduced transactivation by SPA of the LEB–CaMV–*GUS* fusion (from 4.5- to twofold; Figure 6B, section iii) in tobacco mesophyll cells and also reduced SPA-mediated activation of the LEB–CaMV 35S construct in BMS cells (from 11-fold to sixfold; Figure 6C, section iii).

As shown in Figures 5B and 5C, SPA can bind to the individual GLM probes lacking the EM sequences but cannot bind to the LEB sequence in which both GLMs are mutated. As determined in EMSA analyses conducted with wheat endosperm nuclear extracts, the ESBF-II binding activity specifically recognizes the GLMs and is not affected by the mutations in the EMs (Hammond-Kosack et al., 1993). Similarly, the binding of SPA to the GLMs was not expected to be influenced by the mutations in the EMs. To verify this, additional EMSAs were performed using the same LEB probes as described before and using a LEB probe mutated in the two EMs. Figure 7 shows protein–DNA complex formation with both the wild-type probe and the probe mutated in the EMs (EM−), demonstrating that SPA binding at the GLMs is independent of the presence of functional EMs.

This finding suggests that factors present in tobacco and maize leaves that recognize the EMs are necessary for full activation by SPA from the GLMs. Nevertheless, the transactivation of the chimeric LEB–CaMV 35S promoter in both maize and tobacco protoplasts clearly demonstrates that SPA is a transcriptional activator that can activate via the conserved GLMs contained in the bifactorial endosperm box of several prolamin genes.

Comparison of SPA- and O2-Activated Transcription Levels

The similarities between SPA and O2 led us to compare the relative levels of expression directed by these two transcription

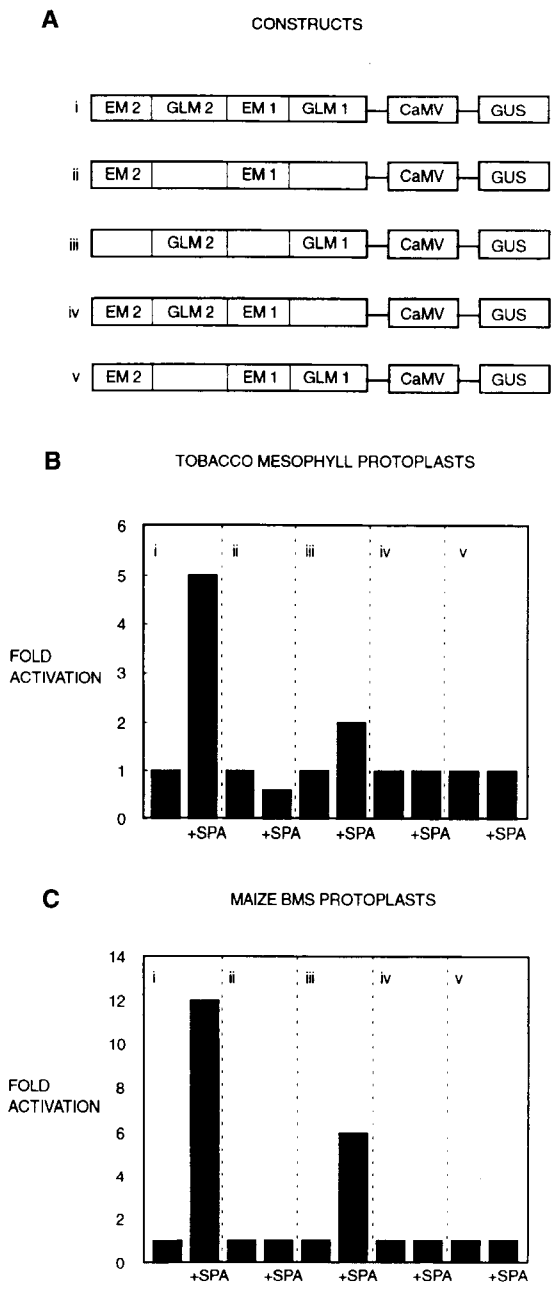

Figure 6. Activation of Transcription from the GLM Sequences by SPA.

(A) Structure of the LEB reporter constructs used in this study. Section i, wild-type LEB–CaMV–*GUS* fusion; section ii, both GLMs mutated in the LEB fusion; section iii, both EMs mutated in the LEB fusion; section iv, GLM1 motif mutated in the LEB; section v, GLM2 motif mutated in the LEB. The sequences of wild-type and mutant LEBs are shown in Figure 1A.

(B) Transient expression of constructs in tobacco mesophyll protoplasts without or with (+SPA) coexpression of a CaMV 35S:SPA plasmid is shown in sections i to v.

factors from the LEB–CaMV 35S construct and from the endosperm box region of a maize 22-kD α-zein promoter fused to a minimal −60 CaMV 35S promoter (22Z4). The 22-kD zein gene promoter is a well-characterized target gene of O2 (Schmidt et al., 1992). The activation potentials of SPA and O2 were assessed in transient assays using BMS cells, and the results are shown in Figures 8A and 8B. Consistent results were obtained in replicated experiments with different batches of protoplasts, with differences in GUS enzyme levels significant at $P < 0.01$.

O2 and SPA activated expression from the wild-type *LMWG-1D1* LEB–CaMV 35S construct ninefold and 11-fold, respectively (Figure 8A, EM^+GLM^+). SPA and O2 activation was dependent on the GLMs (Figure 8A, EM^+GLM^-). Interestingly, as also shown previously in Figure 6C, section iii, SPA-mediated activation was reduced to nearly half by mutation of the EM sequence (Figure 8A, EM^-GLM^+), but this mutation had no signficant effect on O2-mediated activation (nine- and eightfold, respectively). This observation suggested that SPA and O2 may have different modes of action and that only SPA relies on the integrity of the EM for full activation. SPA and O2 activated transcription from the 22Z4 O2 target promoter by 13- and 10-fold, respectively (Figure 8A, 22Z4). These levels compare with previous transient assay results of Ueda et al. (1992), who used a comparable 22-kD zein promoter (seven- to eightfold in BMS cells and 11-fold in endosperm cell lines). Currently, it is not clear whether SPA as well as O2 activate the zein promoter only from the O2 *cis* element, because a putative GLM is also present in this promoter (Figure 8B). In this respect, it is interesting that O2 can still activate the 22Z4 construct with a mutated O2 site to 50% of the level of the wild-type promoter (data not shown).

DISCUSSION

The information obtained about the function of GLM and EM *cis* sequences in the *LMWG-1D1* promoter and the isolation of a seed-specific transcription factor that specifically activates transcription from the GLM has led us to the following conclusions. First, the GLM sequences in the context of the −326-bp *LMWG-1D1* promoter are necessary for seed-

(C) Transient expression of constructs in maize BMS protoplasts without or with (+SPA) coexpression of a CaMV 35S:SPA plasmid is shown in sections i to v.

GUS activities were determined in duplicates of each sample, which was replicated five times in each experiment. Each experiment was repeated at least three times with similar levels of expression ($P < 0.01$).

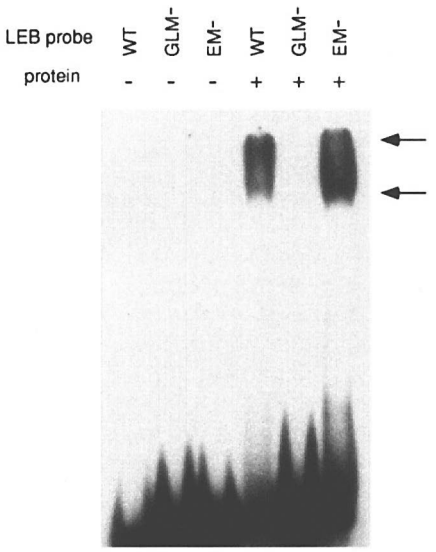

Figure 7. Binding of Recombinant SPA to the GLMs of the LEB Is Not Affected by Mutations in the EMs.

The radiolabeled wild-type (WT) LEB probe, a LEB probe with mutated endosperm motifs (EM−), and a LEB probe with mutated GCN4-like motifs (GLM-) were used for EMSA experiments. (−) or (+) indicates whether the HIS–SPA fusion protein (300 ng) was added. The sequences of the probes are shown in Figure 1A. The arrows indicate the LEB–SPA complexes.

specific expression. Second, the seed-specific SPA transcription factor binds to and activates transcription from the GLM sequences. Third, this activation is enhanced by the adjacent EM sequences, which are not required for SPA binding.

Isolation of a cDNA Encoding SPA

The adoption of a PCR approach for the amplification of bZIP-encoding genes from wheat endosperm, coupled to a round of hybrid selection, resulted in the isolation of two different wheat bZIP cDNA clones encoding proteins showing partial amino acid sequence similarity to the maize transcriptional activator O2. wOHP, the partial cDNA clone initially obtained by PCR, is most similar to the maize bZIP proteins OHP1 and OHP2, and the similar expression patterns observed for both wheat and maize genes indicate that wOHP may be a possible ortholog of OHP1 and OHP2. In maize, OHP1 is expressed at a much lower level than is O2 in developing endosperm, and it is not able to transactivate the O2 target site in transient expression assays. It is thought that OHP1 and OHP2 act as negative modulators of O2 transactivation (Aukerman and Schmidt, 1994). The other wheat bZIP clone, SPA, subsequently was isolated by screening an endosperm cDNA library with the wOHP probe. Our inability to amplify the SPA clone in initial PCR reactions could have resulted from the change of the highly conserved amino acid sequence NRE(S/A)A, from which the sequence of the degenerate primers was derived, to NRDSA in the SPA protein. In addition, SPA is expressed at a level much higher than that of wOHP in developing wheat seeds (data not shown); this fact explains the relative ease of finding SPA in the endosperm cDNA library.

The SPA protein shows a higher overall sequence similarity to OHP1/OHP2 than to O2, but unlike OHP1 and OHP2, SPA is a strong transcriptional activator. The isolation of O2-like clones from other cereals has been reported recently in the case of Coix and sorghum, which belong to the same grass tribe as maize. The close phylogenetic relationship of

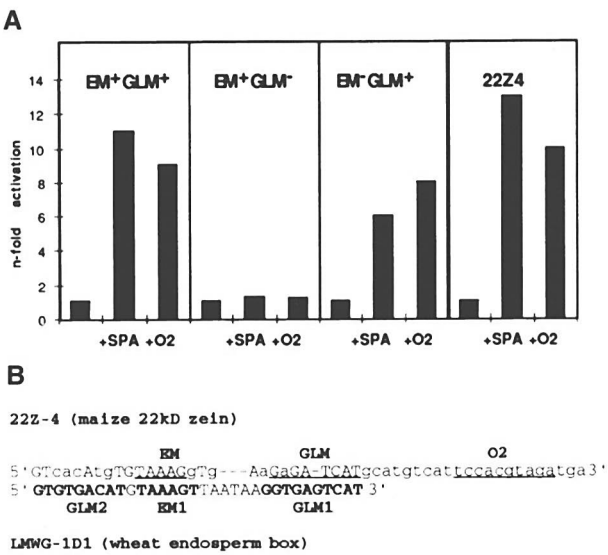

Figure 8. Transactivation by SPA and O2.

(A) Comparison of the activation potential of SPA (+SPA) and O2 (+O2) from the wheat LEB–60CaMV–*GUS* construct and a maize 22-kD zein 22Z4–60CaMV–*GUS* construct by transient expression in maize BMS cells. EM^+GLM^+ is the wild-type LEB, EM^+GLM^- has the two GLM sequences mutated, EM^-GLM^+ has the two endosperm motifs mutated, and 22Z4 is the maize zein reporter construct. Each experiment was repeated three times with five replicates of each GUS assay. Differences in GUS enzyme levels were significant at $P < 0.01$.

(B) Sequence comparison of the maize 22Z4 and wheat *LMWG-1D1* promoters in their endosperm box regions. The wheat GLM and EM sequences are shown in boldface, and the corresponding elements in the maize promoter as well as the canonical O2 binding site are underlined.

these three plant species is reflected in the extensive sequence conservation, in both coding and regulatory regions, that exists among the 22-kD α-zein genes of maize, an α-kafirin gene from sorghum, and an α-coixin gene from Coix (Ottoboni et al., 1993).

In view of the conservation of putative *cis* elements in the promoters of these 22-kD–like α-prolamin genes, the presence of O2-related proteins in Coix and sorghum is not surprising. These proteins show 70% (Cljop2) and 63% (Sbopa2) amino acid identity to O2, but their pattern of expression has not been reported (Pirovano et al., 1994; Yunes et al., 1994). Nevertheless, their high sequence similarity to O2 and the ability of the sorghum clone to transactivate the maize b-32 promoter to a similar extent as O2 in a transient expression assay strongly suggest that they are likely to be orthologs of O2. A potential functional relationship between O2 and SPA may be reflected by the level of sequence conservation within the DNA binding domain of SPA and O2. However, this degree of sequence similarity is also shared by the parsley CPRF-2 protein (Weisshaar et al., 1991), the rice seed protein Osbzippa, and the rice transcriptional activator RITA-1 (Izawa et al., 1994). Little information is available about the tissue specificity and transcription targets of Osbzippa. RITA-1 is predominantly but not specifically expressed in rice aleurone and endosperm, whereas CPRF-2 is expressed in cell cultures. Therefore, it is unlikely that SPA is an ortholog of O2, based on sequence comparisons and expression patterns. This conclusion is supported by other functional information, such as the differences in the requirement of SPA and O2 for an intact EM for transcriptional activation. It is possible that SPA may have an ortholog in maize that may activate transcription from the GLM motifs present in several classes of maize storage protein genes, in contrast to the specific role of O2 in activating the 22-kD class of zein genes.

The spatial and temporal pattern of SPA expression is consistent with its involvement in the regulation of seed-specific gene expression, and its ability to bind and activate transcription from the GLMs of the *LMWG-1D1* promoter indicates that SPA may correspond to the DNA binding factor ESBF-II previously detected in wheat endosperm nuclear extracts (Hammond-Kosack et al., 1993). This hypothesis needs to be proven by independent tests using antibodies in "supershift EMSA" experiments, for example. SPA may be involved in the activation of many wheat prolamin genes, because most of them contain conserved GLMs in the proximal regions of their promoters that are possible targets for SPA.

The presence of an ATG codon in the 5′ untranslated region of the SPA and O2 transcripts (Lohmer et al., 1993) raises the possibility of translational regulation of these genes. In view of the proposed nitrogen response of the GLM in barley (Müller and Knudsen, 1993), SPA expression or translation could be differentially regulated by nitrogen regimes. Further experiments will test these hypotheses and will verify whether SPA is involved in a more general control of storage protein genes and whether it can also regulate other endosperm-specific genes.

EM and GLM Sequences of the Endosperm Box Are Necessary for Seed-Specific Expression

Analysis of the −326-bp *LMWG-1D1* promoter in transgenic tobacco plants revealed that mutations of either the EMs or the GLMs abolished expression in seeds. This result suggests the existence in tobacco endosperm of counterparts to the DNA binding activities ESBF-I and ESBF-II of wheat (Hammond-Kosack et al., 1993). The binding sites of both putative *trans*-acting factors are needed to activate gene expression in tobacco, suggesting a positive interaction between the activities binding to them. Analysis of a barley C-hordein promoter in an endosperm transient bombardment assay also suggested an interaction between the EM and GLM (Müller and Knudsen, 1993). In this study, both the EM and GLM were shown to be required for high-level activation, and a deletion removing the EM but leaving the GLMs intact dramatically lowered expression. These findings are consistent with the important role of the EM described here.

SPA Activates Transcription via the GLM

A construct containing the LEB region fused to a −67-bp minimal CaMV 35S promoter was efficiently transactivated by SPA in both maize BMS and tobacco mesophyll protoplasts. Both GLM sequences in the LEB were required for SPA-mediated activation. Interactions between SPA proteins bound at both GLM sites with the ESBF-I factor bound at the EMs may be necessary for expression, or cooperative binding of SPA to each GLM may be necessary for activation. The comparison of SPA and O2 activities on the LEB–CaMV 35S construct showed that SPA could activate transcription to high levels similar to those produced by O2 in the transient assay system used. The levels of activation imparted by O2 were consistent with previous studies using transient expression in maize endosperm cells of native regions of the 22-kD zein promoter (Ueda et al., 1992). SPA and O2 also activated the 22-kD zein promoter to similar levels (13- and 10-fold, respectively). The site on the 22-kD zein promoter from which SPA activates transcription may be the O2 site, but there is a well-conserved GLM motif adjacent to the EM (see Figure 8B), which is also found in a similar position in the promoters of the 19- and 27-kD zein genes. Assessing the relative contributions of the O2 binding sites and putative GLM sites on these promoters to SPA- and O2-mediated activation would define any functional similarities between SPA and O2.

A functional analysis of a rice storage protein promoter in a homologous and stable transgenic system (Zheng et al., 1993) demonstrated that different *cis* elements direct the endosperm-specific expression of a rice glutelin promoter in rice and tobacco. Several synergistic regulatory elements important for expression in rice were identified, of which some showed sequence similarity to the GLMs. It is possible, therefore, that the regulatory mechanisms directing seed-specific expression in monocots and dicots are partially overlapping and that the lack of effect of mutating the footprinted CAAT box in transgenic tobacco may reflect these differences. Transformation of these promoter deletions into wheat plants would address this issue. The demonstration of a GLM in a pea lectin promoter fragment that is able to confer seed-specific expression in transgenic tobacco plants (de Pater et al., 1993) suggests that the GLMs are a major determinant of seed-specific expression in tobacco plants and probably exert a similar function in monocot plants.

Interestingly, mutations in the EMs of the LEB considerably lowered the transactivating potential of SPA in the transient expression assays. This finding is consistent with the important role of the EM and its cognate binding activity, ESBF-I, in tobacco seed expression. EMSA analyses demonstrated that the binding of SPA to the GLMs in the context of the LEB is not affected by mutation of the EMs and that SPA can specifically bind to the individual GLM1 and GLM2 sequences. These results reflect the binding activities of the ESBF-II factor in wheat nuclear extracts, and mutations of the EMs are known to abolish the binding of the wheat endosperm nuclear factor ESBF-I (Hammond-Kosack et al., 1993). Therefore, the effect of these mutations in the transactivation assays could result from the abolition of binding of a similar factor present in tobacco and maize leaves. The requirement for both GLMs for SPA-mediated activation, together with the requirement for the EM sequences, indicates that complex interactions between SPA bound at both GLMs and with the activity bound at the EMs (ESBF-I) may be required for gene activation. Testing this hypothesis requires the isolation and characterization of the ESBF-I binding activity.

The observed similarities between SPA and O2 do not extend to their interactions with the EM. SPA-mediated activation from the LEB–CaMV 35S construct was enhanced twofold by an intact EM, whereas O2-mediated activation was not affected by the EM. Because both factors activated transcription from the same GLMs on the LEB–CaMV 35S construct (Holdsworth et al., 1995; this study), there appears to be a fundamental difference in their mode of action upon the LEB, and this may also extend to their activities on other promoters. This aspect of SPA and O2 function will be better understood when the protein(s) corresponding to the EM binding activity ESBF-I has been characterized. In vivo footprinting of the endosperm box showed that the EM became occupied before the GLM during endosperm development (Hammond-Kosack et al., 1993), and both motifs were occupied during maximal expression of the LMW glutenin promoter. These observations and the data presented here agree with a model in which ESBF-I binds to the EM and promotes subsequent binding of SPA to the GLM to achieve high-level expression.

METHODS

Plasmid Constructions and Transformation

Mutations in the −326-bp *LMWG-1D1* promoter fragment were introduced by polymerase chain reaction (PCR) (modified from Ho et al., 1989). The resulting PCR fragments were cloned into the BamHI site located upstream of the β-glucuronidase (*GUS*) gene in the plasmid pKGT (Jefferson, 1987) and sequenced. For *Agrobacterium tumefaciens*–mediated plant transformation, the *GUS* constructs were excised as HindIII-EcoRI fragments and introduced into the binary vector pBIN19, and *Nicotiana tabacum* cv Samsun was transformed, as previously reported (Bevan, 1984). Histochemical analysis of GUS activity in developing seeds and replicated fluorometric analyses of GUS activity in seeds and leaves of at least 10 independent transformants were performed as described by Jefferson (1987). The chimeric long endosperm box (LEB)–cauliflower mosaic virus (CaMV) 35S promoter constructs were obtained by cloning the 37-bp LEB fragments into the HindIII-BamHI sites upstream of the −67 35S CaMV–*GUS* gene fusion in the plasmid pBI221.9. The 22Z4 zein reporter construct was obtained by cloning the double-stranded oligonucleotide 5′-agcttAGGTTGTCACATGTGTAAAGGTGAAGAGATCATGTCATCCACGTAGATGAAAAGAAT-TCg-3′ in the same vector. The nucleotides in lowercase are cloning sites. For the construction of the effector plasmid, the same SalI-EcoRI fragment used for the construction of the bacterial expression vector was cloned into the corresponding restriction sites in the polylinker of the CaMV 35S expression vector pJIT60. Transformation of tobacco mesophyll and maize black Mexican sweet (BMS) protoplasts was performed using polyethylene glycol–$CaCl_2$, as described by Negrutiu et al. (1987). Fluorometric detection of GUS activity was determined in replicate samples, according to Jefferson (1987).

RNA Isolation and RNA Gel Blot Analysis

Seeds of wheat plants (cv Chinese Spring) were collected at five stages of development, according to the Zadoks scale (Zadoks et al., 1974), ranging from 10 to 28 days after anthesis (DAA), and leaves were collected at 10 DAA. The earliest stage from which seed components could be recognized on the ear was 10 DAA. Roots and coleoptiles were obtained from 10-day-old germinated seeds. Total RNA was isolated by the hot-phenol method (Shirzadegan et al., 1991). Poly(A)$^+$ RNA was isolated using Dynabeads oligo(dT)$_{25}$ (Dynal, Oslo, Norway). Poly(A)$^+$ mRNA samples were resolved in formaldehyde gels and transferred to Hybond-N membranes (Amersham, Bucks, UK). Prehybridization, hybridization, and washing were performed under standard conditions (Sambrook et al., 1989). The storage protein activator (SPA), low-molecular-weight (LMW) glutenin (Colot et al., 1987), and ubiquitin probes were labeled by the random primer method (Pharmacia, Uppsala, Sweden).

PCR Isolation of a Partial wOHP cDNA

To amplify wheat basic leucine zipper (bZIP) cDNAs expressed in endosperm, two degenerate primers were devised based on the highly conserved amino acid sequence NRE(A/S)A found in the basic domain of most plant bZIP proteins. The first degenerate primer (DA1) contains the sequence 5′-*AAGGGATCC*AAYMGIGARKCIGC-3′ (where I is inosine, Y is A/G, M is A/C, K is G/T, R is C/T, and italicized letters represent the bases of the cloning site), whereas the second primer (DA2) contains the sequence 5′-*AAGGGATCC*AAYMGIGARAGYGC-3′. These two primers were used in 3′ rapid amplification of cDNA ends (RACE) reactions conducted using poly(A)$^+$ RNA from stage 3 wheat seed (18 DAA). As a positive control, similar reactions were conducted on poly(A)$^+$ RNA from maize seed to verify the amplification of *opaque2* (*o2*) cDNA 3′ends. Reverse transcriptions were performed for 1 hr at 42°C in a volume of 20 μL, using 500 ng of poly(A)$^+$ RNA, 100 ng of the poly-(dT)$_{17}$ adapter (Frohman et al., 1988), and 200 units of Superscript reverse transcriptase (Gibco BRL, Paisley, UK). PCR reactions were performed in a volume of 50 μL, using 5 μL of each reversed-transcribed sample diluted 10-fold with H_2O, 5 μL of 10 × PCR buffer (Perkin-Elmer, Warrington, UK), 5 μL of 25 mM $MgCl_2$, 5 μL of 2 mM deoxynucleotide triphosphates (dNTPs), 75 ng of adapter primer (Frohman et al., 1988), and 500 ng of degenerate primer. After incubation at 94°C for 5 min, 2.5 units of Amplitaq (Perkin-Elmer, Warrington, UK) was added, and PCR was performed with three cycles of 1 min at 94°C, 1 min at 40°C, and 2 min at 72°C followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, and 2 min at 72°C.

An aliquot of the PCR products was resolved on an agarose gel, blotted, and hybridized at low stringency (50°C, with 5 × SSPE [20 × SSPE is 3 M NaCl, 0.2 M NaH_2PO_4, pH 7.4, 0.02 M EDTA]) with a cDNA encoding O2 (Lohmer et al., 1991). Weak hybridization of an ~750-bp band was detected. Hybrid selection was then performed to isolate the 750-bp products hybridizing with the *O2* clone. Approximately 2 μg of pJIT60 plasmid containing *O2* cDNA was diluted to 1 μg/μL in 10 × SSC (20 × SSC is 5 M NaCl, 0.3 M trisodium citrate, pH 7.0) and spotted on 0.5 cm^2 of a prewetted Hybond-N membrane. After denaturation for 5 min and neutralization for 1 min in standard buffers (Sambrook et al., 1989), the DNA was UV cross-linked, and the membrane was prehybridized for 5 hr at 50°C in 1 mL of 0.5 M NaH_2PO_4, pH 7.0, 7% SDS, and 5 mM EDTA. Hybridization was performed for 20 hr at 50°C in 0.5 mL of the same solution as above but containing 20 μL of denatured wheat seed PCR product (~2 μg of DNA). After hybridization, the membrane was washed twice for 1 hr at 50°C in 1.5 mL of 2 × SSC and 0.1% SDS and twice for 2 hr at 50°C in 1.5 mL of 0.5 × SSC and 0.1% SDS. After a brief rinse in 0.5 × SSC, the hybridized DNA was released by incubating the membrane for 5 min at 100°C in 100 μL of H_2O. Five microliters of the selected products was used in a PCR reaction, as given above, performing 30 cycles of 1 min at 94°C, 1 min at 55°C, and 2 min at 72°C. After separation by electrophoresis, the major PCR fragment, corresponding to the partial wOHP cDNA, was eluted, digested with BamHI, and cloned into the plasmid pBluescriptII KS+ (Stratagene, La Jolla, CA).

Isolation of SPA cDNAs

Approximately 60,000 plaque-forming units from a wheat oligo(dT) primed endosperm cDNA library constructed in λZAPII (Stratagene) were screened with the partial wOHP cDNA clone. Hybridization was performed for 20 hr at 60°C in standard hybridization solution (Sambrook et al., 1989). The filters were then washed twice for 30 min at 60°C in 2 × SSC and 0.1% SDS and twice for 30 min at 60°C in 0.5 × SSC and 0.1% SDS. Two weakly hybridizing plaques were purified. The plasmids containing the SPA cDNAs were excised in vivo according to the manufacturer's protocol, and sequencing was performed on both strands of the longest cDNA clone. The sequence has an EMBL accession number of Y09013.

Expression and Purification of Recombinant SPA

A subclone of the longest SPA cDNA was used for expression in *Escherichia coli*. The 5′ untranslated region containing an additional ATG codon was removed by ligating the blunt-ended NcoI site located at the putative start codon into the blunt-ended ClaI site of the upstream polylinker. The entire SPA coding region was then isolated as a SalI-EcoRI fragment and cloned into the XhoI and EcoRI sites in the polylinker of the His-tagged fusion vector pRSET-A (Invitrogen, San Diego, CA). The resulting plasmid (pHIS–SPA) was introduced into *E. coli* BL21(DE3)pLysS for protein expression. The recombinant product consists of the N-terminal peptide with the sequence 5′-MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKRWGSELD-3′ fused to the entire SPA protein. For expression of the HIS–SPA protein, cells were grown in 50 mL of Luria-Bertani medium with ampicillin (100 μg/mL) to an OD_{600} of ~0.6 and induced with 1 mM isopropyl β-D-thiogalactopyranoside for 1 to 3 hr before harvesting. Pelleted cells were resuspended in 800 μL of lysis buffer (50 mM $Na[PO_4]$, pH 8.0, and 300 mM NaCl) containing 4 mM imidazole and 0.5 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) and then frozen in liquid nitrogen. The thawed suspension was sonicated three times for 5 to 10 sec and then centrifuged for 20 min at 12,000*g* at 4°C. The resulting supernatant (600 μL) was incubated for 90 min with constant movement at 4°C with 200 μL of a 1:1 slurry of NiNTA (Qiagen, Dorking, UK) resin preequilibrated in lysis buffer. The resin with the bound protein was then pelleted by brief centrifugation and washed twice at 4°C for 5 min with 1 mL of lysis buffer containing 4 mM imidazole and 0.5 mM AEBSF, twice with 1 mL of lysis buffer containing 20 mM imid-azole and 0.5 mM AEBSF, and once with 200 μL of lysis buffer containing 100 mM imidazole. The recombinant HIS–SPA protein was eluted for 5 min at 4°C with 200 μL of lysis buffer containing 250 mM imidazole, aliquoted, and stored at −70°C.

Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSAs) were conducted as previously described (Hammond-Kosack et al., 1993). The radiolabeled oligonucleotide probes were incubated for 30 min at 25°C with 300 ng of the recombinant HIS–SPA protein, 500 ng of poly(dA-dT), and 500 ng of poly(dI-dC) in 15 μL of 25 mM Hepes, pH 7.5, 100 mM KCl, 5% glycerol, 1 mM EDTA, and 10 mM DTT. The protein–DNA complexes were electrophoresed for 2 hr at room temperature on 4% polyacrylamide gels in 0.5 × TBE. The sequences of the individual GLM probes were 5′-gatcccgTAA*GGTGAGTCA*TATAgccg-3′ for GLM1 and 5′-gatcccgAGT*GTGTGACAT*GTAAgccg-3′ for GLM2. The sequence for the GLM1 mutant is 5′-gatcccgTAATCTTACTCATATAgccg-3′, and the sequence for the GLM2 mutant is 5′-gatccgAGTTTCTTATGGTTAAgccg -3′. (The lowercase letters represent the cloning sites, and the italicized sequences represent the protein binding sites.)

ACKNOWLEDGMENTS

D.A. acknowledges a European Commission Human Capital and Mobility Fellowship (Contract ERBCHBICT930457), M.C.U.H.-K. is supported by the European Community BIOTECH Programme (Project of Technological Priority 1993–1996), M.H. was supported by a Department of Trade and Industry–LINK grant (Contract EG/2Y), and V.C. was supported by a fellowship from Sidney Sussex College (Cambridge, UK).

Received July 19, 1996; accepted December 16, 1996.

REFERENCES

Aukerman, M.J., and Schmidt, R.J. (1994). Regulation of α-zein expression during maize endosperm development. In Plant Promoters and Transcription Factors, L. Nover, ed (Heidelberg, Germany: Springer-Verlag), pp. 209–233.

Bevan, M.W. (1984). Binary *Agrobacterium* vectors for plant transformation. Nucleic Acids Res. **12,** 8711–8721.

Colot, V. (1990). The genes encoding wheat storage proteins: Towards a molecular understanding of breadmaking quality and and its genetic manipulation. In Genetic Engineering, Vol. 12, J.K. Setlow, ed (New York: Plenum Press), pp. 225–242.

Colot, V., Robert, L.R., Kavanagh, T.A., Bevan, M.W., and Thompson, R.D. (1987). Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J. **6,** 3559–3564.

de Pater, S., Pham, K., Chua, N.-H., Memelink, J., and Kijne, J. (1993). A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. Plant Cell **5,** 877–886.

de Pater, S., Katagiri, F., Kijne, J., and Chua, N.-H. (1994). bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. Plant J. **6,** 133–140.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85,** 8998–9002.

Hammond-Kosack, M.C.U., Holdsworth, M.J., and Bevan, M.W. (1993). *In vivo* footprinting of a low molecular weight glutenin gene (*LMWG-1D1*) in wheat endosperm. EMBO J. **12,** 545–554.

Ho, S.N., Hunt, H.D., Horton, R.M., Fuller, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77,** 51–59.

Holdsworth, M.J., Munoz-Blanco, J., Hammond-Kosack, M., Colot, V., Schuch, W., and Bevan, M. (1995). Transcriptional activation by Opaque-2 from a novel binding site in the wheat LMW-glutenin *1D1* gene promoter. Plant Mol. Biol. **29,** 711–720.

Izawa, T., Foster, R., Nakajima, M., Shimamoto, K., and Chua, N.-H. (1994). The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. Plant Cell **6,** 1277-1287.

Jefferson, R.A. (1987). Assaying chimeric genes in plants: The GUS gene fusion system. Plant Mol. Biol. Rep. **5,** 387–405.

Kodrzycki, R., Boston, R.S., and Larkins, B.A. (1989). The *Opaque-2* mutation of maize differentially reduces zein gene transcription. Plant Cell **1,** 105–114.

Kreis, M., Shewry, P.R., Forde, B.G., Forde, J., and Miflin, B.J. (1985). Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. In Oxford Surveys of Plant Molecular and Cell Biology, B. Miflin, ed (Oxford, UK: Oxford University Press), pp. 253–317.

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N., Hartings, H., Salamini, F., and Thompson, R.D. (1991). The maize regulatory locus *Opaque-2* encodes a DNA-binding protein which activates the transcription of the *b-32* gene. EMBO J. **10,** 617–624.

Lohmer, S., Maddaloni, M., Motto, M., Salamini, F., and Thompson, R.D. (1993). Translation of the mRNA of the maize transcriptional activator Opaque-2 is inhibited by upstream open reading frames present in the leader sequence. Plant Cell **5,** 65–73.

Maier, U.-G., Brown, W.S., Toloczyki, C., and Feix, G. (1987). Binding of a nuclear factor to a consensus sequence in the 5′ flanking region of zein genes from maize. EMBO J. **6,** 17–22.

Marris, C., Gallois, P., Copley, J., and Kreis, M. (1988). The 5′ flanking region of a barley B-hordein gene controls tissue and developmental specific CAT expression in tobacco plants. Plant Mol. Biol. **10,** 359–366.

Mauri, I., Maddaloni, M., Lohmer, S., Motto, M., Salamini, F., Thompson, R., and Martegani, E. (1993). Functional expression of the transcriptional activator Opaque-2 of *Zea mays* in transformed yeast. Mol. Gen. Genet. **241,** 319–326.

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

Negrutiu, I., Shillito, R., Potrykus, I., Biasini, G., and Sala, F. (1987). Hybrid genes in the analysis of transformation conditions. I. Setting up a simple method for direct gene transfer into plant protoplasts. Plant Mol. Biol. **8,** 363–373.

Niu, X., and Guiltinan, M.J. (1994). DNA binding specificity of the wheat bZIP protein EmBP-1. Nucleic Acids Res. **22,** 4969–4978.

Ottoboni, L.M.M., Leite, A., Yunes, J.A., Targon, M.L.N.P., Souzo Filho, G.A., and Arruda, P. (1993). Sequence analysis of 22-kD–like α-coixin genes and their comparison with homologous zein and kaffirin genes reveals highly conserved protein structure and regulatory elements. Plant Mol. Biol. **21,** 765–778.

Pirovano, L., Lanzini, S., Hartings, H., Lazzaroni, N., Rossi, V., Joshi, R., Thompson, R.D., Salamini, F., and Motto, M. (1994). Structural and functional analysis of an *Opaque-2*–related gene from sorghum. Plant Mol. Biol. **24,** 515–523.

Pysh, L.D., Aukerman, M.J., and Schmidt, R.J. (1993). OHP1: A maize basic domain/leucine zipper protein that interacts with Opaque2. Plant Cell **5,** 227–236.

Robert, L.S., Thompson, R.D., and Flavell, R.B. (1989). Tissue-specific expression of a wheat high molecular weight glutenin gene in transgenic tobacco. Plant Cell **1,** 569–578.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Schernthaner, J.P., Matzke, M.A., and Matzke, A.J.M. (1988). Endosperm-specific activity of a zein gene promoter in transgenic tobacco plants. EMBO J. **7,** 1249–1255.

Schmidt, R.J., Ketudat, M., Aukerman, M.J., and Hoschek, G. (1992). Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. Plant Cell **4,** 689–700.

Shirzadegan, M., Christie, P., and Seemann, J.R. (1991). An efficient method for isolation of RNA from tissue cultured plant cells. Nucleic Acids Res. **19,** 6055.

Tabata, T., Takase, H., Takayama, S., Mikami, K., Nakatsuka, A., Kawatawa, T., Nakayama, T., and Iwabuchi, M. (1989). A protein that binds to a *cis*-acting element of wheat histone genes has a leucine zipper motif. Science **245,** 965–967.

Takaiwa, F., Oono, K., and Kato, A. (1991). 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.

Thomas, M.S., and Flavell, R.B. (1990). Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell **2,** 1171–1180.

Ueda, T., Waverczak, W., Ward, K., Sher, N., Ketudat, M., Schmidt, R.J., and Messing, J. (1992). Mutations of the 22- and 27-kD zein promoters affect transactivation by the Opaque-2 protein. Plant Cell **4,** 701–709.

Weisshaar, B., Armstrong, G.A., Block, A., da Costa e Silva, O., and Hahlbrock, K. (1991). Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. EMBO J. **10,** 1777–1786.

Yunes, J., Neto, G.C., da Silva, M.J., Leite, A., Ottoboni, L.M.M., and Arruda, P. (1994). The transcriptional activator Opaque2 recognizes two different target sequences in the 22-kD–like α-prolamin genes. Plant Cell **6,** 237–249.

Zadoks, J.C., Chang, T.T., and Konzak, C.F. (1974). A decimal code for the growth stages of cereals. Weed Res. **14,** 415–421.

Zheng, Z., Kawagoe, Y., Xiao, S., Li, Z., Okita, T., Hau, T.L., Lin, A., and Murai, N. (1993). 5′ distal and proximal *cis*-acting regulatory elements are required for developmental control of a rice seed storage protein gene. Plant J. **4,** 357–366.