The Plant Cell, Vol. 5, 877–886, August 1993 © 1993 American Society of Plant Physiologists

A 22-bp Fragment of the Pea Lectin Promoter Containing Essential TGAC-like Motifs Confers Seed-Specific Gene Expression

Sylvia de Pater,[a,b,1] Khanh Pham,[a,b] Nam-Hai Chua,[c] Johan Memelink,[b] and Jan Kijne [a,b]

[a] Center for Phytotechnology, Leiden University, Nonnensteeg 3, 2311 VJ Leiden, The Netherlands
[b] Institute of Molecular Plant Sciences, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands
[c] Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021

To elucidate the molecular mechanisms responsible for seed-specific gene expression in plants, the promoter of the pea lectin (*psl*) gene, encoding an abundant seed protein, was used as a model. Leaf and seed nuclear proteins bound to a region in the *psl* promoter containing three overlapping TGAC-like motifs, which have been shown to be a binding site for basic/leucine zipper proteins, including TGA1a. A trimer of a 22-bp region of the *psl* promoter, containing the TGAC-like motifs, coupled to a heterologous minimal promoter conferred low reporter gene expression in root, stem, and leaf and high expression in seed of transgenic tobacco. Expression increased during the midmaturation stage of seed development and was observed in the endosperm as well as in the embryo, where it strongly decreased within a few days after germination. This expression pattern is qualitatively identical to the expression pattern conferred by a 2000-bp fragment of the *psl* promoter. Nucleotides within the TGAC-like motifs important for in vitro binding are also essential for in vivo transcription activation in vegetative tissue as well as in seed. The electrophoretic mobility of a DNA–protein complex containing seed nuclear protein was different from that formed with leaf nuclear protein. Furthermore, the TGA1a steady state mRNA level in immature seed was relatively low. These results suggest that a seed-specific factor different from TGA1a, but with similar binding specificity, is responsible for gene activation in seed. We conclude that the 22-bp region contains all the information, including an essential TGAGTCATCA sequence, necessary for seed-specific expression and very likely plays an essential role in the seed-specific expression pattern of the *psl* gene.

INTRODUCTION

Development of different organs and tissues of higher eukaryotes involves temporal and spatial regulation of specific sets of genes. Embryogenesis and seed development are examples of highly orchestrated developmental processes in higher plants (reviewed by Goldberg et al., 1989). A variety of proteins accumulate and decay during this stage of development and provide an excellent system for investigating different aspects of gene regulation. We studied promoter sequences and *trans*-acting factors involved in seed-specific gene expression using the pea lectin (*psl*) gene as a model. Pea has one functional lectin gene (Gatehouse et al., 1987; Kaminski et al., 1987). This gene is highly expressed in seeds with the highest steady state mRNA level during the midmaturation stage (Buffard et al., 1988). The protein has been located in the cotyledons and the axis of the embryo (Van Driessche et al., 1981). Evidence is accumulating that lectins in seeds of leguminous plants play a role in defense against herbivorous insects (Boulter et al., 1990). In roots of 21-day-old plants, the mRNA is ~4000 times less abundant than in embryos, whereas in leaf no mRNA is detected (Buffard et al., 1988). Small amounts of PSL protein are located on the tips of pea root hairs (Diaz et al., 1986), where it was shown to have an important function in the determination of the *Rhizobium* host range for nitrogen-fixing root nodule formation (Diaz et al., 1989).

The rate of PSL protein synthesis during pea seed development is directly related to the steady state mRNA level (Higgins et al., 1983), which suggests that lectin abundance in seed is transcriptionally regulated. Identification of promoter sequences required for seed-specific expression is an important step toward the elucidation of the precise mechanism responsible for controlled transcription. A few sequence elements have been shown to confer seed-specific expression in a gain-of-function approach. Lam and Chua (1991) described seed-specific expression conferred by a tetramer of a 21-bp synthetic element. This element was also responsive to the plant growth factor abscisic acid, which plays a critical role in embryogenesis (reviewed by Skriver and Mundy, 1990). Although this synthetic element functions as a seed-specific activator sequence in plants, it is not known whether this sequence occurs

[1] To whom correspondence should be addressed at the Center for Phytotechnology.

in any plant promoter, and neither have transcription factors been identified that bind to this sequence.

Seed-specific expression has also been observed with a reporter gene construct containing a G-box–related motif derived from abscisic acid–inducible rice promoters (Salinas et al., 1992). The *trans*-acting factor that is responsible for seed-specific expression conferred by this element is unknown. The cloned transcription activator factor TAF-1 that weakly binds to this sequence is thought not to be involved in seed-specific expression mediated by this G-box, because it is predominantly expressed in roots.

In another approach to dissect the mechanisms for seed-specific gene expression, maize mutants with aberrant endosperm were characterized. Several *trans*-acting factors were identified with regulatory functions in storage protein synthesis and pigmentation. One of these regulatory proteins, Opaque-2 (O2), was found to transactivate promoters of the *b-32* albumin gene (Lohmer et al., 1991) and zein genes (Schmidt et al., 1992). Binding sites for the O2 protein in the *b-32* promoter contain the core sequence TGAPyPu (Lohmer et al., 1991). Other transcription factors binding to this core sequence have been identified (Katagiri et al., 1989; Singh et al., 1990; Fromm et al., 1991; Tabata et al., 1991; Schindler et al., 1992), and they all belong to the basic/leucine zipper (bZIP) class of DNA binding proteins. One of these is the tobacco transcriptional activator TGA1a (Katagiri et al., 1989, 1990; Yamazaki et al., 1990).

Here we show that nuclear proteins from leaf and seed bind to a short fragment of the *psI* promoter, displaying similar nucleotide requirements as the bZIP protein TGA1a. In a gain-of-function approach, we demonstrated that a 22-bp fragment including this binding site contains a seed-specific enhancer element. We present evidence suggesting that seed-specific factors different from TGA1a, but with similar binding specificity, are involved in the seed-specific expression conferred by this element.

RESULTS

Binding Specificity of Leaf and Seed Nuclear Proteins to Wild-Type and Mutant *psI* Promoter Oligonucleotides

The promoter of the *psI* gene contains a binding site for bZIP proteins (S. de Pater, F. Katagiri, J. Kijne, and N.-H. Chua, manuscript in preparation). This binding site contains three overlapping TGAC-like motifs. To test whether proteins can be detected in nuclear extracts that bind to this binding site, gel shift assays were performed with leaf and seed nuclear extracts, as shown in Figure 1. Nucleotide requirements of nuclear proteins were compared with those of the bZIP protein TGA1a. Synthetic oligonucleotides were made containing wild-type or mutant *psI* promoter sequences from −56 to −35 (Figure 1A).

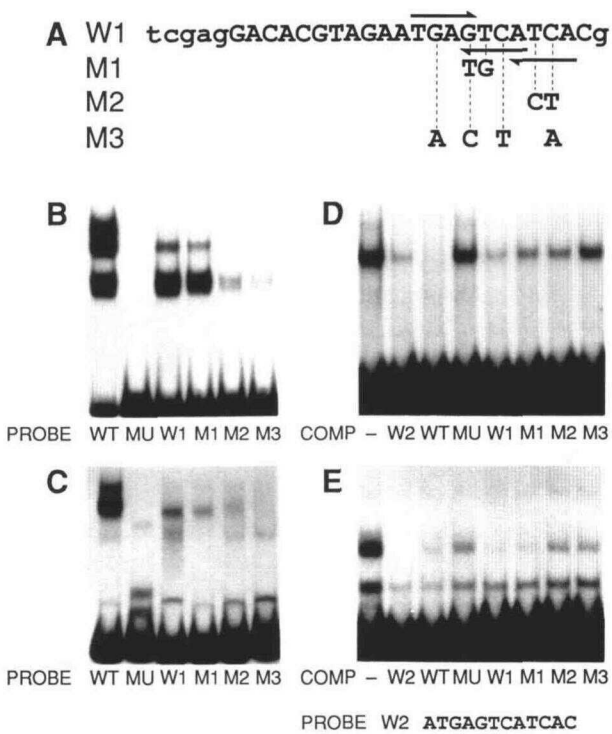

Figure 1. Binding of TGA1a and Nuclear Proteins to Wild-Type or Mutant Oligonucleotide Binding Sites.

(A) Oligonucleotides containing 22 bp of the *psI* promoter, including the binding site for TGA1a (S. de Pater, F. Katagiri, J. Kijne, and N.-H. Chua, manuscript in preparation). W1 contains three copies of the −56 to −35 sequence of the *psI* promoter and 6 extra bp (lowercase letters) to create restriction sites used for cloning. The arrows indicate the TGAC-like motifs. M1 to M3 contain mutations in these TGAC-like motifs as indicated.

(B) Gel shift assays with purified TGA1a and tetramerized wild-type *as-1* (4A1; WT), tetramerized mutant *as-1* (4A3; MU), and trimers of the oligonucleotides listed in **(A)**.

(C) Gel shifts with tobacco leaf nuclear extract and the binding sites used in **(B)**.

(D) Competitive gel shifts with tobacco leaf nuclear extract. A probe (20 fmol) containing a tetramer of the binding site shown (W2) was incubated with leaf nuclear extract. The following competitors (Comp) were used in 50-fold molar excess: no competitor (–), W2, 4A1 (WT), 4A3 (MU), W1, M1 to M3.

(E) Competitive gel shifts with tobacco seed nuclear extract. Ten fentomoles of probe (W2) and 200-fold molar excess of competitors (Comp) were used.

The position of the W1 oligonucleotide in the *psI* promoter is shown in Figure 2. Trimers of these oligonucleotides were used as probes in a gel shift assay with purified TGA1a (Figure 1B). Besides W1 and M1 to M3, oligonucleotides 4A1 and 4A3 were used as probes for comparison.

Oligonucleotide 4A1 is a tetramerized activation sequence 1 (*as-1*) element. This binding site for TGA1a is derived from the cauliflower mosaic virus (CaMV) 35S promoter, where it

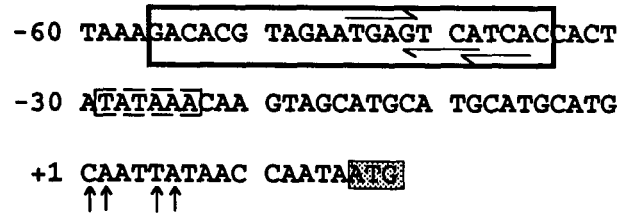

Figure 2. Context of the W1 Sequence in the *psI* Promoter.

The 5′ ends of the mRNA are indicated by arrows and the ATG start codon is shaded. The putative TATA box is boxed with broken lines. The W1 oligonucleotide is boxed with solid lines. TGAC-like motifs are overlined with arrows when present in the coding strand or underlined when present in the noncoding strand.

is located between positions -82 to -62. The nuclear binding activity that binds to this sequence is termed activation sequence factor 1 (ASF-1). Two base pairs of each binding site present in 4A1 are mutated in 4A3, causing a severe reduction of binding (Lam et al., 1989). Otherwise, the sequences of 4A1 and 4A3 are identical.

Oligonucleotide 4A1 was clearly shifted by TGA1a. Slower migrating complexes represent multiple proteins bound to one DNA molecule. No complex was observed after mutation of the TGAC motifs in *as-1* (4A3). The wild-type *psI* binding site for TGA1a (W1) was also bound; however, the amount of complex was less when compared to the amount obtained with 4A1. Mutations in the left half of the binding site (M1) slightly decreased the amount of complex formed. Mutations in the right half of the binding site (M2) clearly reduced the amount of complex. Binding to M3, in which four G residues are mutated, was hardly detectable. These four G residues were also shown to be important for binding in methylation-interference assays (S. de Pater, F. Katagiri, J. Kijne, and N.-H. Chua, manuscript in preparation).

Subsequently, binding of tobacco leaf nuclear proteins to these oligonucleotides was tested (Figure 1C). Besides some nonspecific bands, the results were similar to those obtained with purified TGA1a. The amount of complex was highest for 4A1. Using the oligotrimer W1, a smaller percentage of the probe was shifted when compared to 4A1. Mutations present in M1 and M2 reduced the amount of complex, and retardation of M3 was below the level of detection. The relative binding affinities of TGA1a and leaf nuclear protein for M1 were slightly different. The affinity of TGA1a for M1 is only slightly less than for W1, whereas the affinity of leaf nuclear protein for M1 is comparable to the affinity for M2. A minor complex was observed using 4A3 as probe, but this complex migrates slightly differently from the other complexes and is therefore probably not ASF-1 related.

Competition assays were performed to test whether the same plant nuclear proteins bind to the TGA1a binding site of the 35S promoter (4A1) and to the TGA1a binding site of the *psI* promoter (W1). Because nuclear proteins resulted in many nonspecific DNA–protein complexes with the oligotrimer W1, a smaller oligonucleotide, termed W2, was made that contained four copies of the TGA1a binding site of the *psI* promoter without flanking sequences (Figure 1). Using the W2 oligotetramer, a complex was obtained with tobacco leaf extract (Figure 1D). The addition of unlabeled W2 drastically decreased complex formation. The addition of 4A1 as a competitor eliminated binding to W2 completely, whereas the presence of 4A3 had no effect. Thus, the leaf nuclear protein that binds to W2 is ASF-1. Competition with the oligonucleotides listed in Figure 1A showed that the longer oligotrimer W1 reduced complex formation to the same extent as W2. M1 and M2 partially competed for binding, whereas the addition of M3 had no effect at all. Very similar results were obtained with tobacco seed nuclear proteins (Figure 1E). Minor differences are the slightly lower binding affinities of the seed protein for 4A1 and M2 when compared to the binding affinities of the leaf protein. Thus, both leaf and seed nuclear protein from tobacco bind to the W2 oligonucleotide and exhibit binding specificities similar to those of TGA1a for the different mutant binding sites tested.

The 22-bp *psI* Promoter Fragment Containing the TGAC-like Motifs Drives Seed-Specific Expression in Transgenic Tobacco

To investigate the in vivo expression pattern conferred by the ASF-1 binding site in the *psI* promoter, constructs containing trimers of the 22-bp oligonucleotides shown in Figure 1A were coupled to −46GUS (Benfey et al., 1990). This reporter gene construct contains the TATA box of the CaMV 35S promoter (−46 to +8), the *Escherichia coli* β-glucuronidase (*gusA*) coding region, and the polyadenylation sequence of a pea gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase (*rbcS-3C*). These constructs were introduced into tobacco by Agrobacterium-mediated leaf disc transformation. Protein extracts of several independently transformed plants for each construct were tested in a fluorometric assay for GUS activity. Control plants containing -46GUS were tested to determine background levels. For leaf, stem, and root protein extracts, in vitro–grown plants were used, whereas for seed protein extracts, plants were grown in a greenhouse.

Figure 3 shows that in leaf tissue of plants transformed with the construct containing the wild-type oligotrimer W1, GUS activity was 2.5 times higher on average than in leaf tissue of −46GUS-transformed plants. The two mutant sequences, M1 and M2, which still bind ASF-1, although to a lesser extent, also increased GUS activity when compared to −46GUS. However M3, to which ASF-1 did not bind, did not confer increased expression compared to −46GUS. On average, expression of this construct was even lower than that of −46GUS. Similar results were obtained with stem and root extracts, except that all values were on the average three- and 10-fold higher, respectively. This is probably due to endogenous plant glucuronidase

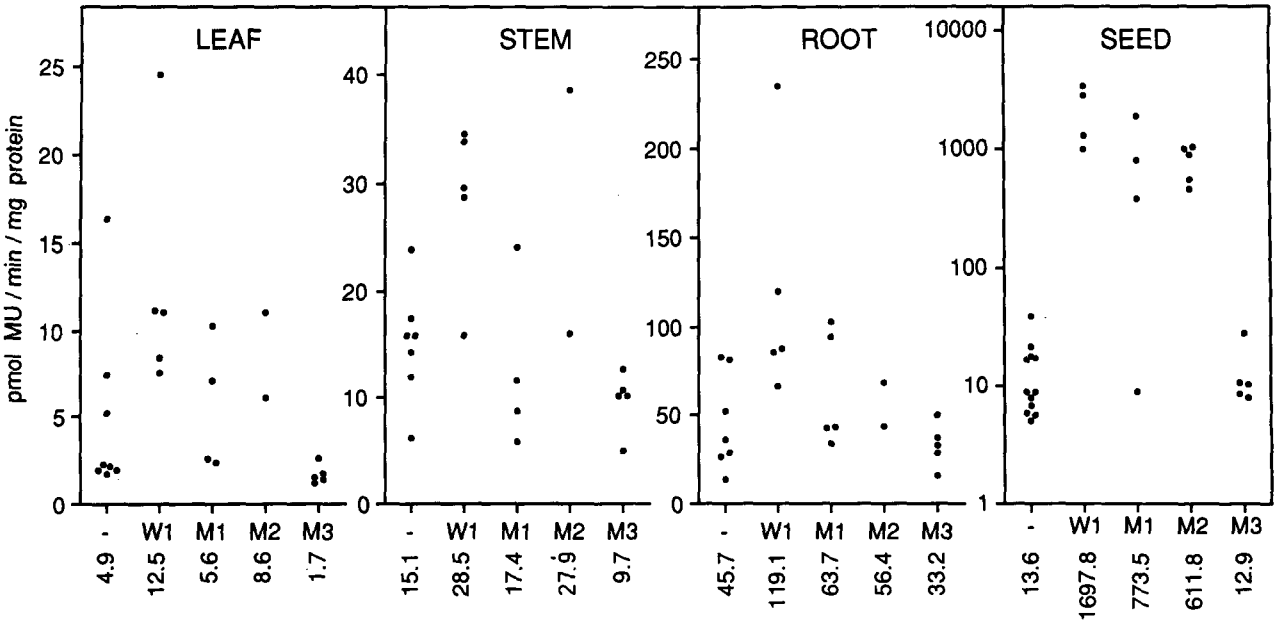

Figure 3. GUS Activity in Transgenic Tobacco Conferred by the W1 and M1 to M3 Oligonucleotides.

GUS activity was determined in leaf, stem, root, and mature seed of tobacco plants transformed with constructs containing trimers of the oligonucleotides listed in Figure 1A (W1 and M1 to M3) coupled to −46GUS. For comparison, plants transformed with −46GUS were analyzed. Each dot represents the GUS activity in an independently transformed plant. The mean values of the GUS activities per construct are given at the bottom of each graph. Note the logarithmic scale of the seed panel.

activity, because higher values were also observed in stem and root of untransformed plants (results not shown). A similar observation was previously reported by Lam and Chua (1990).

Analysis of extracts of mature seeds of the different transgenic plants revealed that GUS activity in seeds of plants containing the W1-GUS construct was more than 100 times higher than in seeds of plants containing −46GUS. GUS activity in seeds of M1-GUS and M2-GUS plants was ~50 times higher, whereas presence of the M3 oligonucleotide did not affect GUS activity when compared to −46GUS.

To further investigate tissue specificity of the expression driven by the short *psI* promoter fragment, histochemical analyses of transgenic tobacco seeds were performed, as shown in Figure 4. Fresh sections were made of W1-GUS and M3-GUS mature seeds, and they were subsequently incubated with a GUS substrate for histochemical staining. Seeds transformed with the W1-GUS construct were clearly stained, whereas seeds transformed with the M3-GUS construct did not show any staining (Figure 4A). Young W1-GUS embryos were uniformly stained, indicating expression in the cotyledons and in the embryonic axis. At a later stage, the staining was restricted to the cotyledons (Figure 4B). A light blue coloration was also observed in the endosperm. In fruits of W1-GUS–transformed plants, GUS expression was restricted to the seeds and was not observed in the vegetative tissue. For comparison, sections of mature seeds and an immature fruit from a transformed plant containing the *gusA* reporter gene under control of a 2000-bp fragment of the *psI* promoter were stained (Figure 4A). This plant exhibited an expression pattern that is very similar to W1-GUS–transformed plants: in fruit, expression is restricted to the seeds, which were stained in the embryo as well as in the endosperm. The expression in leaf and root of the *psI*-GUS construct is very low (results not shown).

GUS Activity Increases during Seed Development and Decreases upon Germination

Construct W1-GUS is highly expressed in the embryo but very poorly expressed in vegetative tissue of mature plants (Figure 3). This suggests that W1-GUS expression drastically increases during seed development and decreases during germination. To determine expression at different developmental stages, seeds and seedlings were stained for GUS activity 5 to 14 days after flowering (DAF) and 2 to 6 days after germination (Figure 4B). No staining was found in seeds at 5 and 7 DAF. GUS expression was first observed 10 DAF. Ripening of the seeds resulted in a decrease of staining, caused by thickening of the seed coat. Two days after germination, seeds containing the W1-GUS construct were strongly stained, especially at the site where the seed coat opens. Three days after germination, most embryos were completely blue. However, after 4 days of

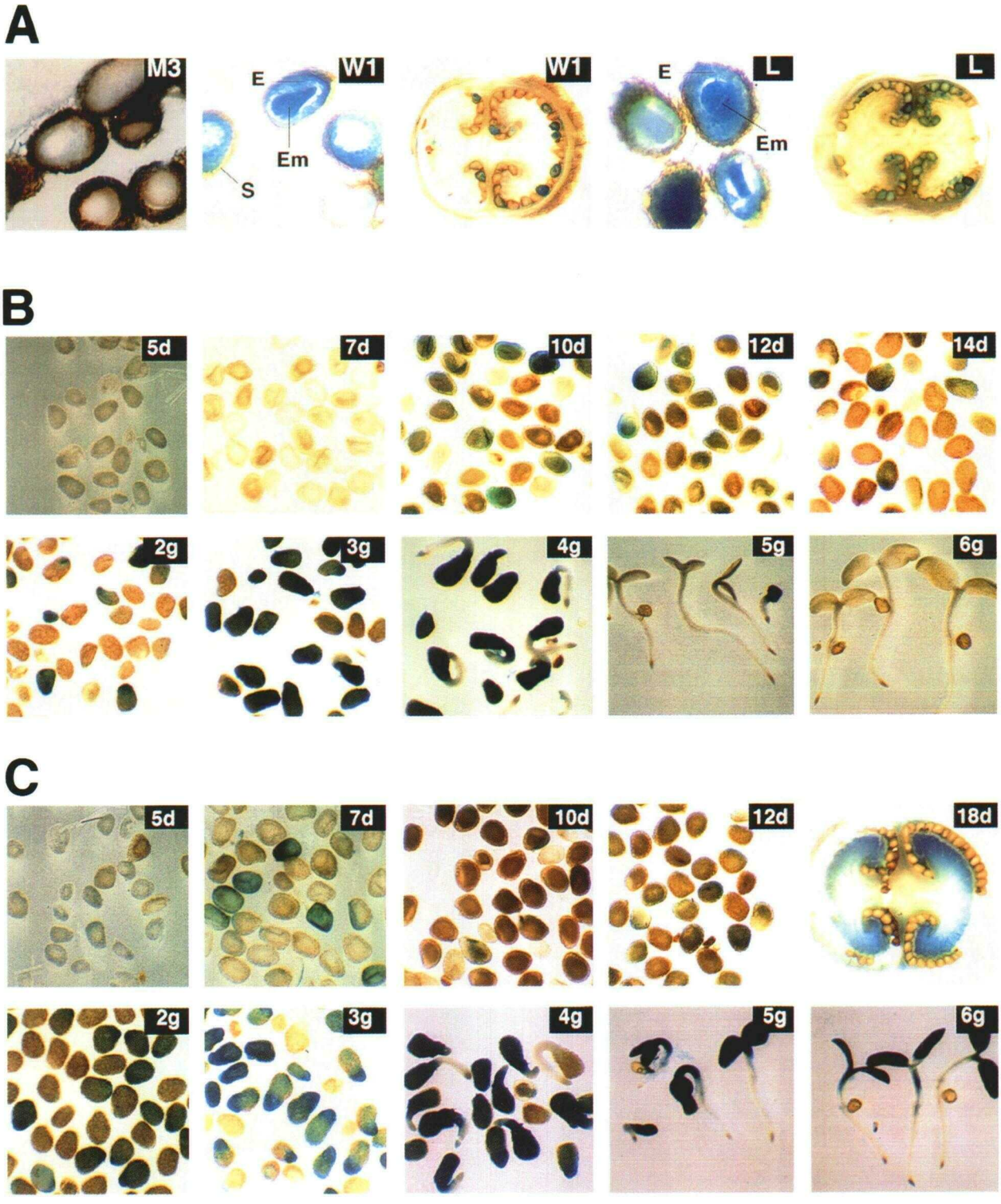

Figure 4. Histochemical Staining of Transgenic Tobacco Seeds and Seedlings at Different Developmental Stages.

(A) Sections of mature seeds and fruits (18 DAF) of M3-GUS–transformed (M3), W1-GUS–transformed (W1), and *psl*-GUS–transformed (L) tobacco plants were stained for GUS activity using the substrate X-gluc. Em, embryo; E, endosperm; S, seed coat.

(B) Seeds and seedlings of a W1-GUS–transformed plant.

(C) Seeds and seedlings of a 35S-GUS–transformed plant.

In **(B)** and **(C)**, seeds and seedlings were stained for GUS activity over a period of 5 to 14 DAF and 2 to 6 days after germination. In addition, a fruit (18 DAF) of a 35S-GUS–transformed plant is shown. d, DAF; g, days after germination.

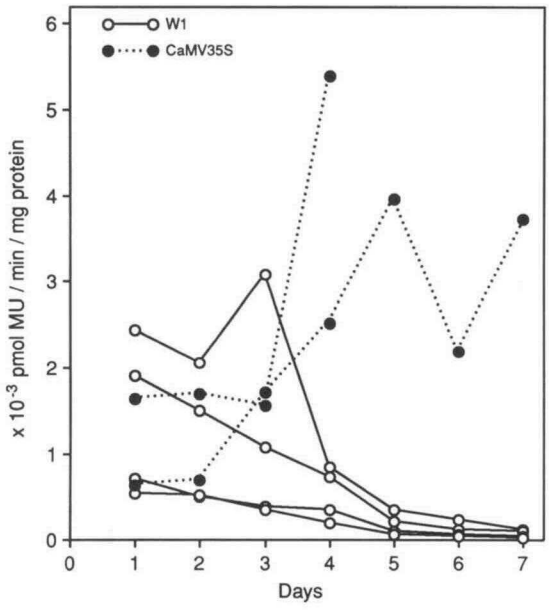

Figure 5. Time Course of GUS Enzyme Activity in Transgenic Tobacco Seeds and Seedlings upon Germination.

GUS activity was determined in protein extracts from seeds and seedlings of four independently transformed W1-GUS plants (solid lines) and two independently transformed 35S-GUS plants (stippled lines) 1 to 7 days after germination.

germination, staining decreased; 1 day later, staining could only be observed in those seedlings that were retarded in their development. Six days after germination, no staining was detectable.

For comparison, seeds from a plant transformed by a *gusA* coding region under control of the CaMV 35S promoter (35S-GUS) were stained (Figure 4C). During seed development, seeds stained weakly at all stages, including very young seeds (5 and 7 DAF) in which no W1-GUS expression was observed. The vegetative tissue of the fruit was also stained; thus, expression is not restricted to the seeds. Seeds and seedlings stained strongly at each stage of germination tested. In seedlings, staining was most apparent in the cotyledons and in the region just below the apical shoot meristem.

To more precisely quantify the expression level during seedling development, GUS enzyme activity was determined fluorometrically. Figure 5 shows the analysis of four independently transformed W1-GUS plants and two independently transformed 35S-GUS plants. GUS activities in W1-GUS seeds 1 day after germination were similar to those in mature seeds (Figure 3). During germination, GUS activities in W1-GUS seedlings decreased in accordance with the histochemical staining pattern. After 7 days, expression in W1-GUS seedlings was 20- to 29-fold less when compared to the level at 1 day after germination, with values in the range of the expression levels in vegetative tissue of mature plants. In contrast, the activity of the CaMV 35S promoter increased dramatically upon germination, consistent with the results of Lam and Chua (1991). The results presented in Figures 3 to 5 clearly demonstrate that the W1 oligotrimer confers seed-specific expression, whereas the CaMV 35S promoter exhibits expression throughout plant development.

Leaf and Seed Nuclear Proteins Form Different Complexes with the W2 Oligotetramer

Binding to the TGAC-like motifs in the W2 oligotetramer detected with nuclear extracts from both leaf and seed could be due to the same protein species expressed in both organs, or, alternatively, distinct protein species could be present in each organ with similar binding specificities. Close inspection of the competitive gel shifts with W2 (Figure 1) shows that the binding specificities of leaf and seed nuclear protein are slightly different: seed nuclear protein has a somewhat higher affinity for W2 when compared to 4A1, whereas the reverse is true for leaf nuclear protein. This suggests that the proteins are different. Another indication that leaf and seed factors are different is the electrophoretic migration rates of the complexes formed following the incubation of W2 with leaf or seed nuclear extracts. As shown in Figure 6A, the migration rates of leaf and seed complexes are different, indicating that either seed and leaf proteins are different or that the same protein is differently modified in leaf and seed.

To investigate whether the seed nuclear protein that binds to W2 could be a modified form of TGA1a, the TGA1a steady state mRNA level in seed was determined by RNA gel blot analysis. Equal amounts of total RNA from green fruit, flower, leaf, stem, and root of tobacco were hybridized with a TGA1a cDNA

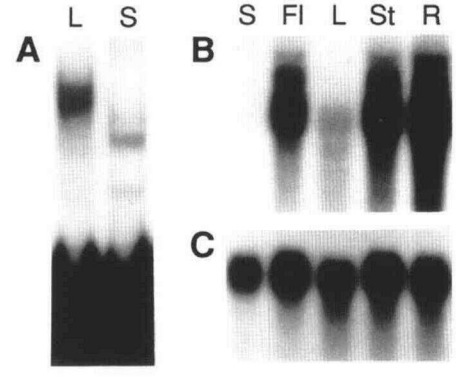

Figure 6. Seed Nuclear Protein Binding to the Oligotetramer W2 Differs from Leaf Nuclear Protein.

(A) Gel shifts with W2 and tobacco leaf (L) and seed (S) nuclear protein. **(B)** and **(C)** TGA1a and pCNT6, respectively, expressed in various tobacco tissues. Gel blots contain RNA from green fruit (S), flower (Fl), leaf (L), stem (St), and root (R) of tobacco. The RNA was hybridized with ^{32}P-labeled cDNA inserts. Each lane contains 10 μg of total RNA.

clone. Figure 6B shows that the highest amount was found in root, which is in agreement with the results of Katagiri et al. (1989). In flower and stem, the amount was somewhat lower, whereas in leaf it was just above the level of detection. No hybridization could be detected in RNA from green fruit. Green fruit contains immature seed, which is the tissue that expresses the W1-GUS construct. As a control for RNA loading, the blot was reprobed with pCNT6, a tobacco cDNA clone that corresponds to an mRNA that is abundant in all tissue types (Memelink et al., 1987). Figure 6C shows a similar signal obtained with pCNT6 in all lanes, indicating the presence of equal amounts of mRNA. Thus, the amount of TGA1a mRNA in green fruit is very low and much less than in leaf. Therefore, it is unlikely that the seed protein that binds to W2 is a modified form of TGA1a.

DISCUSSION

In Vitro Binding of Nuclear Proteins to TGAC-like Motifs in the *psI* Promoter

Nuclear proteins from tobacco leaf and seed bind to the TGAGTCATCA sequence in the *psI* promoter with similar DNA binding site preferences as cloned TGA1a. Complexes formed with the leaf nuclear protein migrate identically in retardation gels as do TGA1a complexes (results not shown). This possibly reflects the fact that leaf ASF-1 is composed of TGA1a and/or other related proteins. Several findings suggest, however, that the seed protein, although displaying similar binding specificity, is different from the leaf protein and TGA1a. The complex obtained with seed nuclear extract and the W2 oligotetramer migrates differently from similar complexes with leaf nuclear extract or TGA1a. This strongly suggests that the proteins in the complexes are different or differently modified. It is unlikely that the seed nuclear protein is a modified form of TGA1a, because the TGA1a steady state mRNA level in immature seed is extremely low. It cannot be excluded that the seed nuclear protein is a TGA1a-related protein encoded by another member of the ASF-1 gene family that does not hybridize to TGA1a under the conditions used for RNA gel blot analysis.

Our approach of identifying a binding site for a cloned bZIP transcription factor has revealed a sequence that confers seed-specific expression, mediated by another factor with similar nucleotide requirements. This approach was successful probably because different classes of bZIP factors, mediating expression in different tissues, bind to the same core sequence with binding being dependent on the flanking nucleotides (Schindler et al., 1992; Williams et al., 1992; Izawa et al., 1993). Identification of the factor responsible for seed-specific expression mediated by W1 will therefore necessarily include determination of binding affinities for W1 of putative candidates. Furthermore, they should be expressed in the correct developmental stages. The possibility of bZIP proteins forming heterodimers will complicate the identification of these factors, because the heterodimeric complex between two bZIP proteins may have a significantly different affinity for a binding site than either protein as a homodimeric complex.

In Vivo Function of the TGA1a Binding Site of the *psI* Promoter

The small 22-bp *psI* promoter fragment confers high expression in seed and low expression in vegetative tissue in transgenic tobacco. This expression pattern parallels the expression pattern of the *psI* promoter in pea (Buffard et al., 1988) and tobacco (Figure 4 and results not shown). GUS activity in tobacco seeds transformed with either the W1-GUS or the *psI*-GUS construct was observed mainly in the embryo and to a lesser extent in the endosperm. Previously, PSL was immunologically detected only in cotyledons and the embryonic axis in pea seeds, and, therefore, PSL is known as an embryo-specific protein. The tissue specificity of the *psI* promoter and the W1 oligotrimer in tobacco seems to be somewhat different from the localization pattern of the *psI* gene product in pea. A similar result was found with β-conglycinin, which is an embryo-specific soybean storage protein. After introduction into tobacco, the promoter of β-conglycinin also showed some activity in endosperm (Chamberland et al., 1992). The endosperm of mature seeds of legumes is very small because it is absorbed by the embryo during seed development, whereas in tobacco, the endosperm is a prominent part of the mature seed. This means that GUS expression was observed in a tobacco tissue type that is almost undetectable in pea seeds at a comparable developmental stage. Most likely the apparent difference in tissue specificity reflects a difference in seed morphology between tobacco and pea rather than differences in gene regulation.

The in vivo expression conferred by the short *psI* promoter fragment is dependent on the presence of the same nucleotides that are required for in vitro binding of ASF-1 or related proteins. Most ASF-1 binding sites studied so far are from promoters of bacterial or viral origin (Fromm et al., 1989; Lam et al., 1989, 1990; Singh et al., 1990). Here, we clearly demonstrate that these binding sites are also involved in the expression of endogenous plant genes.

In this report, seed specificity has been attributed to a very small fragment, thereby reducing the complexity of binding sites and *trans*-acting factors, which are involved in the establishment of this expression pattern. To date, very few other tissue-specific promoters have been dissected to such an extent in plants. Most sequences important for high-level tissue-specific expression were found by loss-of-function deletion analysis and could well be general enhancers. Sequences identified by gain-of-function approaches were rather large and very likely contain multiple transcription factor binding sites. Several relatively small promoter elements, such as binding sites for ASF-2 (Lam and Chua, 1989) and GT-1 (Lam and Chua, 1990), have been shown to confer tissue-specific expression. The activity of these promoter elements required the

presence of −90 to +8 of the CaMV 35S promoter. This promoter (−90) contains an ASF-1 binding site and was shown to act synergistically with other regulatory elements (Benfey et al., 1990) to give expression in tissue types that do not express reporter genes driven by these elements or −90 alone. In this work, we used −46 to +8 of the CaMV 35S promoter (−46) as a minimal promoter, which is commonly used by many groups and was never found to contain transcriptionally active sequences besides the TATA box. Therefore, the expression patterns of the chimeric constructs reflect the activity of the regulatory sequences alone.

A striking correlation was found between in vitro binding of nuclear proteins and in vivo transcription activation. A similar correlation was found for O2 (Lohmer et al., 1991; Schmidt et al., 1992). The binding site for this transcription factor is essential for transactivation of the *b-32* albumin gene and zein genes in maize seed. Like the ASF-1 binding site of the *psI* promoter, binding sites in the *b-32* promoter and *O2* promoter contain a TGAC-like sequence. The presence of O2 binding sites in a reporter gene construct enhances transient expression in endosperm protoplasts approximately twofold, whereas no enhancement was found in protoplasts of a leaf tissue–derived suspension culture (Ueda et al., 1992). This finding suggests that O2 binding sites confer endosperm-specific expression, but conclusions about the tissue specificity conferred by the O2 binding site must await the analysis of transgenic plants. Furthermore, the construct also contained an ASF-1 binding site as part of the heterologous promoter used (−90). Therefore, it remains to be determined whether O2 binding sites contain all of the essential information to confer endosperm-specific expression.

The TGAGTCATCA sequence binds tobacco leaf and seed nuclear factors in vitro with similar affinity, whereas in vivo the expression level conferred by a 22-bp fragment including this sequence is much higher in seed than in leaf. A similar discrepancy was observed for Drosophila homeodomain proteins. Different homeodomain proteins show cross-competition in vitro but do not necessarily activate transcription from the same binding site in vivo (reviewed by Hayashi and Scott, 1990). There are several possible ways in which the seed-specific expression mediated by the W1 oligotrimer could be achieved. The first possibility is that in vivo seed factors bind to this element with higher affinity when compared to leaf factors. Slight differences in affinity or protein concentration near a transcriptional regulation threshold level may make a big difference in the transcriptional effect on a target promoter. The second possibility is that seed factors are stronger activators than leaf factors. Strong *trans*-activation or high affinity for the target site could also depend on seed-specific cofactors that selectively modify the activity of seed DNA binding proteins through protein–protein interactions or covalent modifications. A third possibility is that there is a synergistic interaction between the TGAC half sites and an unknown accessory site on the W1 DNA element, binding a seed-specific factor. Tissue-specific expression patterns in Drosophila conferred by the NP element, which is a binding site for several Drosophila homeodomain proteins, probably also resulted from combined action at the NP element and other cryptic sequences in the heterologous promoter (Vincent et al., 1990).

By demonstrating seed-specific gene activation mediated by the short *psI* promoter fragment and identification of nucleotides that are essential for in vitro binding as well as in vivo transcription activation, important tools are now available to answer the questions raised above and other questions concerning the mechanisms responsible for seed-specific gene expression. The results of such research may give new insights into factors that control tissue-specific gene expression in general.

METHODS

Gel Retardation

The TGA1a protein (kindly provided by F. Katagiri, Harvard Medical School and Massachusetts General Hospital, Boston), encoded by a cDNA clone (Katagiri et al., 1989), was isolated from *Escherichia coli* and purified to homogeneity, as described by Katagiri et al. (1990). Nuclear extracts were prepared from tobacco leaf tissue and developing tobacco seeds (12 days after flowering [DAF]), as described by Green et al. (1989), except that the seeds were powdered in liquid N_2 prior to homogenizing in buffer and the crude nuclear extracts were adjusted to 61% saturation with ammonium sulfate to precipitate the proteins. Synthetic oligonucleotides containing XhoI and SalI sticky ends were multimerized and cloned in pBluescript SKII+. Tetramers of wild-type (4A1, *as-1a*) and mutant (4A3, *as-1c*) *as-1* sites were described by Lam et al. (1989). Plasmids were digested with XhoI and HindIII (tetramers 4A1 and 4A3) or XhoI and SalI (trimers W1 and M1 to M3) and 3′ end labeled using the Klenow fragment of DNA polymerase I and α-^{32}P-dCTP. Labeled fragments were isolated from polyacrylamide gels.

Gel shift experiments were performed essentially as described previously (Green et al., 1987). Binding reactions for gel shifts were done in a final volume of 10 μL with 0.04 ng (1 fmol) of probe and 0.9 μg of purified TGA1a or 10 μg of leaf nuclear extract in the presence of 3 μg of poly[d(IC)]. For competitive gel shifts, 0.35 to 0.7 ng (10 to 20 fmol) of W2 probe and 0.4 to 4 μg of nuclear proteins were used in the presence of 2.5 μg of poly[d(IC)] and 2.5 μg of poly[d(AT)].

Construction of Chimeric β-Glucuronidase Gene Constructs and Transformation of Tobacco

A BglII-BamHI fragment (−46GUS) containing the cauliflower mosaic virus (CaMV) 35S promoter TATA box, the *E. coli* β-glucuronidase (*gusA*) coding region (pRAJ260; Clontech, Palo Alto, CA), and the ribulose-1,5-bisphosphate carboxylase small subunit (*rbcS-3C*) polyadenylation sequence was cloned in the BamHI site of pBluescript SKII+ containing trimers of the 28-bp oligonucleotides (Figure 1A). The reporter gene constructs were transferred as XhoI-BamHI fragments to the binary vector pSDM300 (kindly provided by A. Hoekema, Mogen International, Leiden) digested with SalI and BamHI. The −46GUS BglII-BamHI fragment was cloned in the BamHI site of pSDM300. As a positive control construct, an intron containing the *gusA* gene, which is under control of the CaMV 35S promoter (Vancanneyt et al., 1990) and cloned in the HindIII site of pSDM300, was used. For the *psI* promoter–GUS construct, a 2000-bp fragment of the 5′ noncoding region was fused to the *gusA* coding region (pGUSN358S; Clontech) and the

polyadenylation sequence from the pea *rbcS-3C* gene. This construct was cloned in the binary vector pMOGλRGX (pMOG402 derivative; Mogen International). Binary plasmids were transferred to the disarmed *Agrobacterium tumefaciens* MOG101 (A. Hoekema) in a triparental conjugation with *E. coli* HB101 containing plasmid pRK2013. Tobacco (*Nicotiana tabacum*, cv Petit Havana SR1) was transformed with the MOG101 derivatives by the leaf disc transformation method (Horsch et al., 1985). The presence of the *gusA* gene in the tobacco plants was demonstrated by polymerase chain reaction with genomic DNA using the oligonucleotides 5′-GCAACGTCTGGTATCAGC-3′ and 5′-CACTGACCGGATGCCGAC-3′.

Fluorometric GUS Assay and Histochemical Staining

Transgenic plants were analyzed for GUS activity according to the method of Jefferson (1987) using 4-methylumbelliferyl glucuronide (Research Organics, Inc., Cleveland, OH) as a substrate. The reactions were performed in microtiter well plates for 2 (high activities) to 30 hr (low activities), using 1 to 20 μg of protein. Protein concentrations in extracts were determined by the method of Bradford (1976), using the Bio-Rad protein assay dye reagent. Product formation was quantified several times during the reactions in a fluorometer (model LS50; Perkin-Elmer), and GUS activity was calculated by linear regression analysis of the values obtained. GUS activity is given as picomoles of product (4-methylumbelliferone [MU]) formed in 1 min per milligram of protein. Fresh sections of seeds were made by immobilizing mature seeds on a piece of carrot tissue using cyanoacryl adhesive (Benfey et al., 1989). Sections of 100 to 200 μm were incubated directly in the histochemical substrate solution containing 6 mM 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) (Biosynth, Staad, Switzerland) for 1 to 2 days at 37°C. Seeds were germinated on MS medium (Murashige and Skoog, 1962) without antibiotics, and after 1 to 7 days, seedlings were histochemically stained in X-gluc.

RNA Gel Blotting and Hybridization

Total RNA was isolated by grinding tissue in liquid N_2 and phenol/chloroform extraction and was precipitated with LiCl. RNA was electrophoresed on 1.5% agarose/1% formaldehyde gels, stained with ethidium bromide (to check for intact RNA), and blotted onto GeneScreen membranes (New England Nuclear, Boston, MA). TGA1a and pCNT6 cDNA inserts were labeled with α-^{32}P-dCTP using a random-primed DNA labeling kit (Boehringer Mannheim). Hybridization and washing were performed, as described by Memelink et al. (1987).

Nucleotide Sequence of the Pea Lectin Gene

The EMBL accession number of the lectin gene of *Pisum sativum*, cv Feltham First, is X66368.

ACKNOWLEDGMENTS

We thank Peter Hock and the Biology Photographic Service, Leiden University, for preparation of the figures. This work was supported by the Netherlands Organization for Applied Scientific Research (TNO).

Received May 6, 1993; accepted June 17, 1993.

REFERENCES

Benfey, P.N., Ren, L., and Chua, N.-H. (1989). The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. EMBO J. **8,** 2195–2202.

Benfey, P.N., Ren, L., and Chua, N.-H. (1990). Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. EMBO J. **9,** 1677–1684.

Boulter, D., Edwards, G.A., Gatehouse, A.M.R., Gatehouse, J.A., and Hilder, V.A. (1990). Additive protective effects of different plant-derived insect resistance genes in transgenic tobacco plants. Crop Prot. **9,** 351–354.

Bradford, M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem. **72,** 248–254.

Buffard, D., Kaminski, P.-A., and Strosberg, A.D. (1988). Lectin-gene expression in pea (*Pisum sativum* L.) roots. Planta **173,** 367–372.

Chamberland, S., Daigle, N., and Bernier, F. (1992). The legumin boxes and the 3′ part of a soybean β-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. Plant Mol. Biol. **19,** 937–949.

Diaz, C.L., van Spronsen, P.C., Bakhuizen, R., Logman, G.J.J., Lugtenberg, B.J.J., and Kijne, J.W. (1986). Correlation between infection by *Rhizobium leguminosarum* and lectin on the surface of *Pisum sativum* L. roots. Planta **168,** 350–359.

Diaz, C.L, Melchers, L.S., Hooykaas, P.J.J., Lugtenberg, B.J.J., and Kijne, J.W. (1989). Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. Nature **338,** 579–581.

Fromm, H., Katagiri, F., and Chua, N.-H. (1991). The tobacco transcription activator TGA1a binds to a sequence in the 5′ upstream region of a gene encoding a TGA1a-related protein. Mol. Gen. Genet. **229,** 181–188.

Gatehouse, J.A., Bown, D., Evans, I.M., Gatehouse, L.N., Jobes, D., Preston, P., and Croy, R.R.D. (1987). Sequence of the seed lectin gene from pea (*Pisum sativum* L.). Nucl. Acids Res. **15,** 7642.

Goldberg, R.B., Barker, S.J., and Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. Cell **56,** 149–160.

Green, P.J., Kay, S.A., and Chua, N.-H. (1987). Sequence-specific interaction of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. EMBO J. **6,** 2543–2549.

Green, P.J., Kay, S.A., Lam, E., and Chua, N.-H. (1989). In vitro DNA footprinting. In Plant Molecular Biology Manual, Supplement II, S. Gelvin and R.A. Schilperoort, eds (Dordrecht: Kluwer Academic Publishers), pp. 1–22.

Hayashi, S., and Scott, M.P. (1990). What determines the specificity of action of Drosophila homeodomain proteins? Cell **63,** 883–894.

Higgins, T.J.V., Chrispeels, M.J., Chandler, P.M., and Spencer, D. (1983). Intracellular sites of synthesis and processing of lectin in developing pea cotyledons. J. Biol. Chem. **258,** 9550–9552.

Horsch, R.B., Fry, J.E., Hoffmann, N.L., Eichholtz, D., Rogers, S.G., and Fraley, R.T. (1985). A simple and general method for transferring genes into plants. Science **227,** 1229–1231.

Izawa, T., Foster, R., and Chua, N.-H. (1993). Plant bZIP protein DNA binding specificty. J. Mol. Biol. **230,** 1131–1144.

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

Kaminski, P.A., Buffard, D., and Strosberg, A.D. (1987). The pea lectin gene family contains only one functional gene. Plant Mol. Biol. **9,** 497–507.

Katagiri, F., Lam, E., and Chua, N.-H. (1989). Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature **340,** 727–730.

Katagiri, F., Yamazaki, K., Horikoshi, M., Roeder, R.G., and Chua, **N.-H.** (1990). A plant DNA-binding protein increases the number of active preinitiation complexes in a human in vitro transcription system. Genes Dev. **4,** 1899–1909.

Lam, E., and Chua, N.-H. (1989). ASF-2: A factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in *Cab* promoters. Plant Cell **1,** 1147–1156.

Lam, E., and Chua, N.-H. (1990). GT-1 binding site confers light responsive expression in transgenic tobacco. Science **248,** 471–474.

Lam, E., and Chua, N.-H. (1991). Tetramer of a 21-base pair synthetic element confers seed expression and transcriptional enhancement in response to water stress and abscisic acid. J. Biol. Chem. **266,** 17131–17135.

Lam, E., Benfey, P.N., Gilmartin, P.M., Fang, R.-X., and Chua, **N.-H.** (1989). Site-specific mutations alter *in vitro* factor binding and change promoter expression pattern in transgenic plants. Proc. Natl. Acad. Sci. USA **86,** 7890–7894.

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.

Memelink, J., Hoge, J.H.C., and Schilperoort, R.A. (1987). Cytokinin stress changes the developmental regulation of several defence-related genes in tobacco. EMBO J. **6,** 3579–3583.

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

Salinas, J., Oeda, K., and Chua, N.-H. (1992). Two G-box-related sequences confer different expression patterns in transgenic tobacco. Plant Cell **4,** 1485–1493.

Schindler, U., Beckmann, H., and Cashmore, A.R. (1992). TGA1 and G-box binding factors: Two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. Plant Cell **4,** 1309–1319.

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.

Singh, K., Dennis, E.S., Ellis, J.G., Llewellyn, D.J., Tokuhisa, J.G., Wahleithner, J.A., and Peacock, W.J. (1990). OCSBF-1, a maize ocs enhancer binding factor: Isolation and expression during development. Plant Cell **2,** 891–903.

Skriver, K., and Mundy, J. (1990). Gene expression in response to abscisic acid and osmotic stress. Plant Cell **2,** 503–512.

Tabata, T., Nakayama, T., Mikami, K., and Iwabuchi, M. (1991). HBP-1a and HBP-1b: Leucine zipper-type transcription factors of wheat. EMBO J. **10,** 1459–1467.

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.

Vancanneyt, G., Schmidt, R., O'Conner-Sanchez, A., Willmitzer, L., and Rocha-Sosa, M. (1990). Construction of an intron-containing marker gene: Splicing of the intron in transgenic plants and its use in monitoring early events in *Agrobacterium*-mediated plant transformation. Mol. Gen. Genet. **220,** 245–250.

Van Driessche, E., Smets, G., Dejaegere, R., and Kanarek, L. (1981). The immuno-histochemical localization of lectin in pea seeds (*Pisum sativum* L.). Planta **153,** 287–296.

Vincent, J.-P., Kassis, J.A., and O'Farrell, P.H. (1990). A synthetic homeodomain binding site acts as a cell type specific, promoter specific enhancer in *Drosophila* embryos. EMBO J. **9,** 2573–2578.

Williams, M.E., Foster, R., and Chua, N.-H. (1992). Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell **4,** 485–496.

Yamazaki, K., Katagiri, F., Imaseki, H., and Chua, N.-H. (1990). TGA1a, a tobacco DNA-binding protein, increases the rate of initiation in a plant *in vitro* transcription system. Proc. Natl. Acad. Sci. USA **87,** 7035–7039.