Regulation of $\beta$-Glucuronidase Expression in Transgenic Tobacco Plants by an A/T-Rich, cis-Acting Sequence Found Upstream of a French Bean $\beta$-Phaseolin Gene

Mauricio M. Bustos, Mark J. Guiltinan, Juan Jordano, Dilara Begum, Fatma A. Kalkan, and Timothy C. Hall

Department of Biology, Texas A&M University, College Station, Texas 77843-3258

A 0.8-kilobase fragment from the 5'-flanking region of a French bean $\beta$-phaseolin gene yielded strong, temporally regulated, and embryo-specific expression of $\beta$-glucuronidase (GUS) in transgenic tobacco plants, paralleling that found for the seed protein phaseolin [Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., Hall, T.C., and Kemp, J.D. (1985) Proc. Natl. Acad. Sci. USA 82, 3320–3324]. Gel retardation and footprinting assays using nuclear extracts from immature bean cotyledons revealed strong binding of nuclear proteins to an upstream region (–628 to –682) that contains two inverted A/T-rich motifs. Fusion of a 103-base pair fragment or a 55-base pair synthetic oligonucleotide containing these motifs to a minimal 35S promoter/GUS cassette yielded strong GUS expression in several tissues. A different pattern of GUS expression was obtained in immature embryos and germinating seedlings from the nominally constitutive, full-length, 35S promoter. Whereas GUS expression under the control of the 0.8-kilobase $\beta$-phaseolin regulatory region is limited to immature embryos, expression from constructs containing the A/T-rich motifs is strongest in roots. These data, combined with S1 mapping, provide direct evidence that a plant upstream A/T-rich sequence that binds nuclear proteins can activate transcription in vivo. They also indicate that additional regulatory elements in the $\beta$-phaseolin 5'-flanking region are required for embryo-specific gene expression.

INTRODUCTION

Seed storage protein gene expression is temporally regulated during embryogenesis and restricted to seed tissues such as the cotyledons or the endosperm (Meinke, Chen, and Beachy, 1981; Higgins, 1984; Larkins et al., 1984; Sengupta-Gopalan et al., 1985; Walling, Drews, and Goldberg, 1986; Schernthaner, Matzke, and Matzke, 1988). Globulin storage proteins found in dicot embryos provide excellent model systems for the study of plant gene regulatory mechanisms. Their corresponding mRNAs accumulate to high levels during the phase of maturation which, in legumes, follows a period of rapid cell division known as the cotyledon stage and precedes seed desiccation and dormancy (Goldberg, Barker, and Perez-Grau, 1989). Phaseolin mRNAs account for a large proportion of the total pool of polyadenylated mRNAs in maturing French beans (Sun, Buchbinder, and Hall, 1975; Hall et al., 1978) and are under both transcriptional and post-transcriptional regulation (Chappel and Chrispeels, 1986). Transcriptional regulation of eukaryotic gene expression involves the interaction of 5'-flanking DNA sequences which, in the case of genes transcribed by RNA polymerase II, include promoters and upstream regulatory elements, with trans-acting protein factors that bind to these regulatory DNA sequences (Dyan and Tjian, 1985; Maniatis, Goodbourn, and Fischer, 1987; Ptashne, 1988). An upstream DNA sequence from positions –78 to –257 of the $\alpha$-subunit of soybean $\beta$-conglycinin has been shown to enhance the expression of a cauliflower mosaic virus (CaMV) 35S promoter in tobacco seeds (Chen, Pan, and Beachy, 1988). Similarly, the –326 to –160 region of a wheat low molecular weight glutenin gene (Chen, Pan, and Beachy, 1988). An upstream DNA sequence from positions –78 to –257 of the $\alpha$-subunit of soybean $\beta$-conglycinin has been shown to enhance the expression of a cauliflower mosaic virus (CaMV) 35S promoter in tobacco seeds (Chen, Pan, and Beachy, 1988). Similarly, the –326 to –160 region of a wheat low molecular weight glutenin gene (Chen, Pan, and Beachy, 1988) is involved in endosperm-specific expression. cis-Acting elements have also been found in other highly regulated plant genes, such as the photoregulated ss3.6 gene (Timko et al., 1985), pea rbcS-3A (Fluh et al., 1986; Green, Kay, and Chua, 1987; Kuhlmeier et al., 1987) and pea lhcp (Simpson et al., 1986) genes, soybean leghemoglobin lbc (Stougard et al., 1987), chalcone synthase (Kaulen, Schell, and Kreuzaler, 1986), soybean heat-shock
supplementary text: (Baumann et al., 1987), and maize adh-1 (Feri and Nick, 1987). Here, we show that sequences upstream of a β-phaseolin gene can regulate the expression of an Escherichia coli β-glucuronidase (GUS) gene in a way that closely mimics the regulation of the phaseolin polypeptide in transgenic tobacco. In addition, we have established that a 55-bp sequence found between -628 bp and -682 bp of the phaseolin transcription initiation site binds nuclear protein(s) from immature bean cotyledons and strongly enhances GUS transcription from a minimal CaMV 35S promoter in transgenic tobacco plants.

RESULTS

The β-Phaseolin 5'-Flanking Region Confers Correct Tissue Specificity and Temporal Regulation to a GUS Reporter Gene in Tobacco

Phaseolin expression is restricted to the embryonic tissues of French beans (Phaseolus vulgaris). Similar tissue specificity was observed for the phaseolin polypeptide in tobacco plants transformed with a β-phaseolin 3.8-kb genomic clone (Sengupta-Gopalan et al., 1985), indicating a high degree of conservation in the mechanisms responsible for the regulation of seed storage protein genes between those two species. To assess the relative contribution of transcriptional control to the overall pattern of regulation in vivo, we fused the putative β-phaseolin promoter and upstream sequences to the GUS coding sequence. The high stability of this enzyme in the tobacco cellular environment and the availability of both a quantitative fluorimetric assay and a histochemical localization technique make this a suitable reporter of promoter activity in plants (Jefferson, Kavanagh, and Bevan, 1987). Plasmid pp+20/GUS, shown in Figure 1A, consists of an 802-bp BglII-ScaI restriction fragment from genomic clone A177.4 (Slightom, Sun, and Hall, 1983) that extends from position -782 to +20 of the β-phaseolin gene, inserted at the Smal site of plasmid pBl110.3 (Jefferson, 1987). This hybrid gene construction was introduced into the genome of tobacco via Ti plasmid-mediated transformation (see Methods), and the expression of the reporter gene was monitored in developing seeds and young seedlings of transgenic plants by measuring β-glucuronidase specific activity. Approximately 17 DAF, after which time it remained relatively constant through 25 DAF to 30 DAF. The variability observed at each developmental stage, indicated by the magnitude of the standard deviation, reflects differences in the rate of seed maturation between individual ovaries, as well as allelic segregation and possibly position effects. Histochemical analysis of GUS activity using the color-yielding substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) in mid-maturation to late-maturation embryos, shown in Figure 9, panels B and E, revealed a gradient of GUS expression along the embryos. GUS activity was consistently highest in the cotyledons, lower in the hypo-
cotyl and middle region of the embryonic axis, and almost completely absent in the radicle pole.

The high level of GUS activity found in seeds contrasted with the rapid decay seen during the early stages of seedling germination. Figure 1C shows a plot of GUS specific activity in total tobacco seedlings as a function of days after imbibition. The time indicated as zero corresponds to seeds that had been surface-sterilized and allowed to imbibe for 2 hr in nutrient medium before being assayed. A 50% reduction in GUS activity took place between the latest stages of seed development analyzed and the beginning of the period of germination. The remaining GUS activity decayed exponentially, reaching background levels (<0.1 pmol of 4-methylumbelliferyl/min/µg of protein) by 12 days. A protein half-life of 48 hr was calculated from the data in Figure 1C. The amount of total protein present in the extracts (indicated with a dashed line in Figure 1C) increased only twofold between 0 days and 6 days after imbibition, indicating that this value may represent a slight underestimate of the real half-life. A half-life of 50 hr has been previously estimated for GUS in tobacco (Jefferson, Kavanagh, and Bevan, 1987). Only background GUS activity was found in the leaves and roots of mature pβ+20/GUS plants. We concluded from these experiments that sequences between nucleotide positions -782 and +20 of the β-phaseolin gene were sufficient to confer temporal regulation and seed-specific expression upon the reporter GUS gene in tobacco. This also indicated that tobacco was a suitable host for functional analysis of cis-acting sequences flanking the β-phaseolin gene.

Identification of an Upstream Region Flanking the β-Phaseolin Gene That Binds Cotyledon Nuclear Proteins

We employed two complementary strategies to define cis-acting DNA elements within the phaseolin regulatory region: mutagenesis and expression in transgenic hosts, and protein-DNA binding experiments in vitro. A series of mutations covering the entire phaseolin promoter and upstream sequences have been introduced into transgenic tobacco plants (to be reported elsewhere). In addition, a search for specific sequences capable of binding nuclear proteins in vitro was undertaken. As summarized in Figure 2A, a series of 32P-labeled overlapping restriction fragments were prepared, spanning the region from -782 to -111 of the phaseolin gene. Crude nuclear extracts from immature bean cotyledons, 12 mm to 15 mm in length (corresponding to the period of 14 DAF to 16 DAF), were incubated with each DNA fragment, and the resulting complexes were separated on 5% polyacrylamide (mobility shift) gels (Fried and Crothers, 1981; Garner and Revzin, 1981). Fragments capable of forming complexes under the experimental conditions utilized here (see Methods) are

![Figure 2. Restriction Map of the French Bean β-Phaseolin Gene 5' Region and Location of Fragments Assayed for Binding Activity to Bean Nuclear Extracts.](image-url)

(A) Restriction enzyme map indicating the sites for enzymes used to generate DNA probes. Scale on top indicates approximate distances in base pairs relative to the transcription start site (+1). Sequences 5' are shown as negative numbers and sequences 3' as positive numbers. Abbreviations: R1, EcoRI; D, Dral; B, BglII; S, SspI; N, Ncol; H, Hhal; H2, HindII; R, RsaI. The N terminus of the coding sequence is indicated by a hatched box. Fragments used for gel retardation experiments are listed by size (in base pairs), restriction enzyme sites, and position relative to the cap site. Presence and absence of binding activity are indicated by the full and open bars, respectively. Fragments designated 103SD and oligo III are included here to show their relative locations. Both fragments bind nuclear proteins as shown in Figures 3, 4, and 5.

(B) Mobility-shift analysis of fragments marked with an asterisk on (A) above. The fragment 402 BH was labeled at either the 5' end (lanes 1 and 2) or the 3' end (lanes 9 and 10). The electrophoretic migration of the different labeled fragments was determined after incubation in the absence (odd-numbered lanes) or presence (even-numbered lanes) of bean cotyledon nuclear extract protein. See Methods for conditions for binding and electrophoresis. Fragment lengths (in base pairs) exclude sequences from the pUC19 polylinker: 51 bp for the 402 BH fragment, 34 bp for the 5' labeled fragments, and 17 bp for the 3' labeled fragments.
indicated with full bars. All fragments were tested; autoradiograms of representatives marked with an asterisk in Figure 2A are shown in Figure 2B. Protein-DNA complexes were initially identified on the autoradiograms as broad bands, the presence of which was dependent on the addition of crude nuclear extract and which displayed a lower electrophoretic mobility than the corresponding free probes. The identification of such bands as representing protein-DNA complexes was further supported by the fact that they were completely abolished by either proteinase K or heat treatment (65°C, 10 min) of the crude nuclear extracts (data not shown). In some cases (see lanes 5, 6, 11, and 13), minor bands could be observed associated with some of the free probes in the absence of added nuclear protein, which were attributed to partial denaturation or secondary structure of the DNA probes. DNA fragments capable of forming putative complexes (360 BN, 402 BH, 190 BD, and 279 SH) shared all or part of a 103-bp region contained between positions −695 and −593. This region was not present in restriction fragments that failed to form complexes under the same experimental conditions, three of which, 191 DH, 42 NH, and 87 BS, are also shown in Figure 2B. An Sspl-Dral restriction fragment corresponding to that region, and designated 103SD (Figure 2A), was subcloned into pUC18 and used in subsequent mobility shift and footprinting experiments to localize further the site(s) of protein binding.

Footprinting of Protein-DNA Complexes within the 103SD Fragment

Two different techniques of in vitro footprinting, DNase I (Galas and Schmitz, 1978) and ortho-phenanthroline-copper ion (Op-Cu) (Kuwabara and Sigman, 1987), were used to define more accurately the site(s) of protein-DNA interaction within the 103SD region. Each strand of the cloned 103SD fragment was radiolabeled by end-filling at adjacent EcoRI or BamHI sites and incubated with nuclear extracts from bean cotyledons. Following treatment with DNase I, the products of enzymatic digestion were separated on urea/polyacrylamide gels together with Maxam and Gilbert ladders (prepared from the same fragments). As shown in Figure 3, three different sequences (−676 to −645, −640 to −632, −630 to −617) that exhibited varying degrees of protection from DNase I digestion were resolved on the coding strand. Addition of nuclear extracts increased the sensitivity to cleavage by DNase I at nucleotide positions −643 and −642 (indicated by arrows), situated approximately in the middle of the protected region. The significance of the lack of DNase I footprinting on the noncoding strand is still unclear. Similar results have been reported in the cases of a mouse immunoglobulin heavy chain gene (Augerau and Chambon, 1986) and of a whey acidic protein gene (Lubon and Hennighausen, 1987). To extend and verify our results with the DNase I technique, the protein-DNA binding reaction mixtures were separated on 5% polyacrylamide mobility shift gels. Bands corresponding to the complex as well as unbound probe were treated with Op-Cu directly in the gel matrix (Kuwabara and Sigman,
Figure 4. In Situ ortho-Phenanthroline-Copper Ion Footprinting of Cotyledon Nuclear Proteins Bound to Sequences within the 103SD Fragment.

Cloned fragment 103SD labeled on the coding or noncoding strand was incubated with different amounts (7 µg or 20 µg of protein) of cotyledon nuclear extract. After electrophoretic separation of the protein-DNA complexes, the gels were treated with ortho-phenanthroline-copper and autoradiographed. Op-Cu-cleaved DMA from the complex (C) and free-probe (F) bands was purified and analyzed in 6% polyacrylamide/urea gels alongside sequencing ladders prepared from the corresponding fragments (lanes A, A+G; lanes T, C+T). Arrows indicate the 3' to 5' direction in each case. Boxes indicate the sequences protected on each strand.

1987). DNA purified from each band was analyzed on urea/polyacrylamide gels as described for the DNase I experiments. In contrast with our results using DNase I, Op-Cu footprinting within the gel matrix revealed binding to both DNA strands. The sequence protected on the coding strand correlated very closely with the one seen using the DNase I technique. A shorter sequence was protected on the noncoding strand, between positions -670 and -625. Thus, both DNase I (Figure 3) and Op-Cu (Figure 4) footprinting approaches detected protein binding sites at similar locations within the 103SD fragment.

Sequence Specificity of Binding Was Demonstrated by Competition Experiments

The results of the preceding sections led to the conclusion that protein(s) from crude bean nuclear extracts bind to a minimal region of approximately 45 bp centered at nucleotide position -642 of the β-phaseolin upstream regulatory region. A 55-bp double-stranded oligonucleotide, designated as oligo III in Figure 5A, was synthesized corre-

Figure 5. Binding Competition Experiments with Mutated Synthetic Oligonucleotides.

(A) Nucleotide sequence of the oligonucleotides used for binding and competition experiments. Oligo III extends from -682 to -628 and includes the minimal region of protection (underlined) on both DNA strands as deduced from DNase I and Op-Cu footprinting experiments (Figures 3 and 4). Oligo II contains phaseolin sequences between positions -692 and -671. Oligos IV to VI were derived from oligo III by the indicated nucleotide substitutions.

(B) Binding competition by wild-type and mutant sequences. Binding reactions contained 1 ng of labeled oligo III in the presence (lanes 2 to 10) or absence (lane 1) of cotyledon nuclear extract (5 µg of protein). Unlabeled competitors were added in lane 3 (25 ng of double-stranded oligo III), lane 4 (25 ng of double-stranded oligo II), lane 5 (1 µg of single-stranded oligo III), lane 6 (25 ng of double-stranded oligo IV), lane 7 (25 ng of double-stranded oligo V), lane 8 (25 ng of double-stranded oligo VI), lane 9 (1 µg of supercoiled pUC19), lane 10 (1 µg of p402 BH). Arrows point to the complex of oligo III with nuclear proteins.
sponding to phaseolin sequences between positions -682 and -628, which contain the minimal region of binding to both strands. As shown in Figure 5B (lane 2), this sequence formed a complex with bean nuclear protein(s). The binding specificity was investigated by competition with synthetic DNA sequences and cloned DNA fragments using the gel mobility shift assay. Four additional synthetic oligonucleotides (Figure 5A) that contained various sequence alterations were utilized in these experiments. Oligo II spans positions -692 through -671, extending only four nucleotides into the 5' border of the footprinted region. Oligo IV contained 14 changes to complementary bases (13 of those represent A→T exchanges). In addition, 14 changes to noncomplementary bases were introduced in oligo V at the same locations. Oligo VI had four nucleotide changes within a symmetrical A/T-rich motif (T₄A₃T₄, -652 to -643). Figure 5B shows a summary of competition experiments using a moderate molar excess of cold double-stranded oligo III itself (lane 3) or only its coding strand (lane 5), the four double-stranded synthetic oligos (oligo II, lane 4; oligo IV, lane 6; oligo V, lane 7; oligo VI, lane 8), and either pUC19 or pUC19 containing a 402-bp restriction fragment (fragment 402 BH, Figure 2A), designated p402BH. Of these, only the double-stranded oligos III and IV and plasmid p402BH competed for complex formation with the oligo III probe. The fact that oligo IV but not oligo V competed effectively suggested that the interaction of this region with nuclear proteins could tolerate a considerable amount of sequence variation as long as a high A/T content was maintained. The failure of oligo VI to compete also indicated that T→G transversions involving at least some of the four nucleotide positions -651, -650, -644, and -645 can abolish binding, implying that they play a role in complex formation or maintenance.

The Phaseolin Upstream Element Functions as a Transcriptional Activator in Transgenic Tobacco

The experiments described in preceding sections clearly show that the region from -682 to -628 represented by oligo III interacts with specific protein(s) from immature bean cotyledon nuclear extracts in vitro, therefore fulfilling one of at least two minimum characteristics of cis-acting DNA regulatory elements. The ability of this sequence to stimulate the transcriptional activity of a minimal plant gene promoter was tested by fusing both the 103SD restriction fragment and the synthetic oligo III upstream of a minimal CaMV 35S promoter/GUS gene cassette (pBI120.x, R. Jefferson, 1988, personal communication). The corresponding gene constructions are shown diagrammatically in Figure 6A. Blunt-ended DNA fragments were introduced in two orientations at the Smal site present in the polylinker region, and the resulting recombinant plasmids were sequenced to verify that no point mutations or deletions had occurred during cloning. All five constructs were transplanted into tobacco, and GUS activity was measured in seeds and seedlings from each plant.

Figure 6. Expression of GUS Activity Is Stimulated by the 103SD and Oligo III DNA Sequences in Transgenic Tobacco Tissues.

(A) Schematic of the gene constructions used for analysis of GUS expression driven by phaseolin nuclear protein binding sequences in transgenic tobacco (see Methods).

(B) GUS specific activity in tobacco seeds. Enzyme activity and protein content were measured in triplicate, in groups of 10 seeds from each plant.

(C) GUS specific activity in separate organs of 12-day-old tobacco seedlings grown under sterile conditions on nutrient medium, in the light. In both (B) and (C), bars represent the average (± s.d.) over all plants with each type of gene construct. Ntx, nontransformed tobacco; pBI120.x, four plants; 103SD, four plants; 103DS, two plants; 55-4, four plants; 55-3, four plants; 35S/GUS, two plants.
ferred into the tobacco genome via Ti plasmid-mediated transformation, and the temporal regulation and tissue specificity of GUS expression were investigated in multiple transformants of each type. The different stages of seed development analyzed were chosen to match key developmental events previously recorded in connection with a hybrid gene construction containing the β-phaseolin regulatory region fused to GUS (construction pβ+20/GUS, Figure 1). In that case, GUS activity was zero at 11 DAF, the rate of accumulation was highest at 16 DAF, and it had reached a maximum level by 21 DAF, decreasing only after the completion of seed development (Figures 1B and 1C). Each bar shown in Figures 6B and 6C symbolizes the average GUS specific activity in four different plants transformed with each gene construct. The calculated variances for each data set are indicated by the error bars. Both types of insert strongly stimulated the activity of the minimal heterologous promoter, which was found to be very low (0.2 pmol to 0.5 pmol 4-methylumbelliferyl/min/μg of protein) at all stages of seed development. Inspection of Figure 6B reveals that, in all cases, the reporter activity followed a similar pattern, being highest at 16 DAF. The data also indicated that the 55-bp synthetic oligo III conferred a level of enhancement similar to that seen with the 103-bp SspI-DraI restriction fragment. However, it is possible that the latter may have been more active during the early stages of seed development, as judged from the approximately twofold higher GUS activity found at 11 DAF in 103SD and 103DS relative to 55-4 and 55-3 seeds. In addition, both DNA fragments appeared to be equally active in either orientation. Two major differences existed between the patterns of GUS regulation during seed development by the A/T-rich upstream element (Figure 6B) and by the 0.8-kb phaseolin upstream region (Figure 1B): (1) the A/T-rich sequence activated transcription from the heterologous promoter at 11 DAF, before the onset of maturation, at a time when the more complete phaseolin regulatory region was still inactive; and (2) whereas GUS activity increased in pβ+20/GUS seeds between 16 DAF and 21 DAF in pβ+20/GUS seeds, it decreased in all four constructs containing the A/T-rich sequence over the same period of development.

The tissue specificity of GUS expression was further investigated in cotyledonary leaves, leaves, hypocotyls, and roots of tobacco seedlings at a time (12 days after imbibition) when GUS activity was no longer detectable with the phaseolin regulatory region construct pβ+20/GUS (Figure 1C). Figure 6C shows that all the constructs containing either the 103SD or oligo III inserts stimulated the expression of GUS specific activity over that seen either in nontransformed or in pBI120.x plants in the four vegetative tissues analyzed. A comparison with values for GUS activity in seedlings containing a complete CaMV 35S promoter fused upstream of the GUS gene indicated that the phaseolin sequence was more active in roots than in the other three vegetative organs. Table 1 lists values for net relative enhancement of GUS expression over pBI120.x (arbitrarily assigned a value of 1) calculated from data presented in Figure 6B. The data in Table 1 revealed that, in the reverse orientation, both DNA fragments conferred a higher degree of enhancement (threefold to fourfold on average) than in the forward orientation in leaves, cotyledons, hypocotyls, and roots. Table 1 also shows that in tobacco root, constructs 103SD, 103DS, and 55-4 conferred a degree of enhancement similar to that of sequences found upstream of position −90 in the CaMV 35S promoter complex. In contrast, relative enhancement values in leaves, cotyledons, and hypocotyls were considerably lower (12-fold to 13-fold).

Correct Transcription Initiation by the Chimeric 55-4 Gene Construction in Tobacco

To support our results on the tissue specificity of the different promoters based upon direct determinations of GUS specific activities, gel blots of total RNA from the shoots of axenically grown seedlings or from immature seeds were hybridized with a nick-translated BamHI-HindIII restriction fragment from plasmid pBI120.x to determine relative levels of GUS mRNA in plants containing constructs pBI120.x, pβ+20/GUS, 55-4, and 35S/GUS. The result of one such experiment using RNA from the shoots of 35-day-old seedlings is shown in Figure 7. Seedlings containing the 35S/GUS construct had the highest level of GUS mRNA. By comparing the hybridization signal from different amounts of 35S/GUS RNA with 55-4 RNA in a different experiment (data not shown), we estimated that the complete CaMV 35S promoter resulted in a 10-fold higher level of mRNA than the chimeric 55-4 promoter, in agreement with the differences observed for GUS specific activity in leaves, cotyledonary leaves, and hypocotyls (Figure 6). No GUS mRNA could be detected in seedlings containing either pβ+20/GUS or the control pBI120.x. Similar blots of immature seed (20 DAF to 21 DAF) RNAs indicated that, as expected from the GUS activity data, the

<table>
<thead>
<tr>
<th>Construct</th>
<th>Seeds</th>
<th>Leaves</th>
<th>Cotyledons</th>
<th>Hypocotyls</th>
<th>Roots</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBI120.x</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>35S/GUS</td>
<td>98</td>
<td>1437</td>
<td>530</td>
<td>408</td>
<td>1160</td>
</tr>
<tr>
<td>103SD</td>
<td>127</td>
<td>37</td>
<td>60</td>
<td>86</td>
<td>1680</td>
</tr>
<tr>
<td>103DS</td>
<td>220</td>
<td>37</td>
<td>33</td>
<td>46</td>
<td>710</td>
</tr>
<tr>
<td>55-4</td>
<td>107</td>
<td>62</td>
<td>113</td>
<td>117</td>
<td>1390</td>
</tr>
<tr>
<td>55-3</td>
<td>80</td>
<td>25</td>
<td>13</td>
<td>29</td>
<td>200</td>
</tr>
</tbody>
</table>

Enhancement values were calculated from the data on GUS specific activity presented in Figures 6B and 6C. GUS levels in control pBI120.x plants have been assigned an arbitrary value of 1. Construct names are the same as in Figure 6A. Seed values are from the 15-DAF stage.
35S in tobacco (Odell, Nagy, and Chua, 1985) revealed that the site of transcription initiation in construct 55-4 was identical to the site of initiation by the entire CaMV 35S promoter complex.

**Histochemical Localization of GUS Activity in Tobacco Embryos: Comparison between p\(\beta+20\)/GUS, 35S/GUS, and 55-4 Embryos**

The chromogenic substrate 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc), which yields a readily visible blue

![Figure 7. Steady-State Level of GUS mRNA in Tobacco Seedlings.](image)

A gel blot of total RNA extracted from the shoots of axenically grown tobacco seedlings was hybridized with a \(^{32}\)P-labeled GUS DNA probe. RNA loadings were: lanes 1 and 5, 35S/GUS (two plants), 4 \(\mu\)g and 15 \(\mu\)g, respectively; lane 2, pBI120.x (two plants), 15 \(\mu\)g; lane 3, p\(\beta+20\)/GUS RNA (four plants), 15 \(\mu\)g; lane 4, 55-4 (four plants), 15 \(\mu\)g. DNA molecular weight standards (MW) are indicated. Numbers on the left indicate size in kilobases. Arrow indicates the position of the GUS transcript.

The amount of GUS mRNA in p\(\beta+20\)/GUS seeds was higher than in seeds from either p35S/GUS or p55-4/GUS plants (data not shown).

The 5' end of the GUS transcript expressed in 55-4 seedlings was mapped by the S1 nuclease method using a \(^{32}\)P-labeled, single-stranded DNA probe containing 100 bases to 150 bases on either side of the ATG codon. After treatment with S1 nuclease (see Methods), the size of the fragment protected from nuclease cleavage by formation of a DNA:RNA hybrid was analyzed by electrophoresis on a urea/polyacrylamide gel. As shown in Figure 8, the first base of the mature 55-4 transcript is a C complementary to the G residue indicated with an asterisk. Comparison with experiments on transcription initiation from the CaMV

![Figure 8. S1 Mapping of GUS Transcription Initiation in 55-4 Transgenic Tobacco.](image)

A GUS single-stranded DNA probe (see Methods) was hybridized to 50 \(\mu\)g of 55-4 RNA, treated with 75 units of S1 nuclease, and subsequently analyzed on a polyacrylamide/urea gel with a di-deoxy sequencing ladder prepared from the same probe. (1), 55-4 S1-treated mixture; P, full-length probe; *, protected fragment. G residue marked with an asterisk on the right corresponds to the 5'-end of the protected fragment.
color in the presence of β-glucuronidase activity, was used to investigate the distribution of β-glucuronidase activity in developing tobacco embryos. Figure 9 shows whole mounted and cryo-sectioned mid-maturation to late-maturation embryos representative of a large number (40 to 50) of similar specimens examined from each type of transformant. Panel A illustrates the absence of detectable β-glucuronidase reaction product in embryos from control pBI120.x plants. An identical result was obtained with embryos from nontransformed plants (data not shown).

Panels B and E show a whole mount and a longitudinal section of 21-DAF pβ+20/GUS embryos. As mentioned above, the β-phaseolin promoter and upstream regulatory sequences yielded very strong expression of the reporter gene in the cotyledons and most of the root-hypocotyl axis. Notably, GUS activity was much lower or absent in the root meristem and cap cells. Panels C and F exemplify the distribution of GUS activity seen in tobacco embryos transformed with the 35S/GUS construct, revealing that the CaMV 35S promoter was preferentially active in the

Figure 9. Histochemical Localization of GUS Activity in Immature Tobacco Embryos.
Eighteen-DAF to 20-DAF tobacco embryos were dissected and incubated in the presence of X-gluc for 4 hr.
(A) to (D) Glycerol clarified whole mounted embryos.
(E) to (G) Cryostat sections (20 μm to 30 μm thickness) made from freshly stained embryos.
(A) pBI120.x.
(B) and (E) pβ+20/GUS.
(C) and (F) 35S/GGUS.
(D) and (G) 55-4. The approximate locations of cotyledon (C), hypocotyl (H), and radicle (R) are indicated in (G).
The distribution in 55-4 embryos was drastically different from those seen in the other two cases, being most prominent in the hypocotyl apex of the embryonic axis (panels D and G). From that point, the GUS activity decreased in both directions toward the base of the cotyledons and the radicle pole of the embryonic axis. In all three cases, the epidermis and provascular cylinder appeared to contain a higher amount of \( \beta \)-glucuronidase activity than the storage parenchyma. This should be interpreted with caution, however, since apparent differences in the concentration of colored reaction product may reflect differences in size and degree of vacuolation among the various cell types. From the information presented in the last three sections, we conclude that the upstream A/T-rich region functions as a transcriptional activator in tobacco.

**DISCUSSION**

**Transcriptional Control by the \( \beta \)-Phaseolin 5′ Regulatory Region in Tobacco**

Fusion of \( \beta \)-phaseolin gene 5′-flanking sequences to a bacterial glucuronidase reporter gene provided a clear indication of the importance of transcriptional control in the overall pattern of regulation of this gene. This artificially constructed gene mimicked most, if not all, of the features of phaseolin gene regulation, such as seed specificity and temporal regulation during seed development. The timing of GUS accumulation in tobacco embryos correlated well with that reported previously (Sengupta-Gopalan et al., 1985) for the \( \beta \)-phaseolin polypeptide under the control of the same 5′ regulatory region. In that case, and most likely due to the lower sensitivity of the immunodetection technique, the \( \beta \)-phaseolin protein was first detected 15 days after flowering, that is, 3 days later than the earliest time at which we could detect significant GUS activity. A rapid decay of GUS activity took place upon germination and growth of \( \beta \beta +20 \)/GUS tobacco seedlings, strongly suggesting that the chimeric gene was no longer transcriptionally active. Although our results clearly argue in favor of transcriptional control as a major regulatory mechanism in phaseolin gene expression, they do not completely rule out an involvement for post-transcriptional regulation of phaseolin mRNA levels in French beans or tobacco. Histochromatic localization of \( \beta \)-glucuronidase revealed high reporter activity in the epidermis, storage parenchyma and provascular cells of the cotyledons, and the middle and upper axes of \( \beta \beta +20 \)/GUS embryos. GUS activity was greatly reduced or absent in all cell types of the lower axis. In situ hybridization has been used to demonstrate a similar pattern of regional distribution of mRNA coding for the homologous soybean \( \beta \)-conglycinin protein in transgenic tobacco embryos (Barker, Harada, and Goldberg, 1988). In that case, however, \( \beta \)-conglycinin mRNA was not expressed in the provascular cylinder of either the cotyledons or the embryonic axis. There is no indication that transport of either the reporter enzyme or the reaction product between different regions of the embryos (i.e., between the cotyledons and the axis) is taking place in our experiments since the GUS activity remains largely confined to distinct areas in each case (compare panels B, C, and D in Figure 9). We conclude that the differences observed between our results and those obtained with \( \beta \)-conglycinin are real, and possibly attributable either to different cell type specificities of the two promoters or differential stability of the \( \beta \)-conglycinin and GUS mRNAs in provascular cells. In situ hybridizations to sections of tobacco embryos expressing either an authentic \( \beta \)-phaseolin mRNA or \( \beta \)-glucuronidase mRNA are presently being conducted to resolve this question.

**An A/T-Rich Sequence from the Phaseolin Upstream Region Binds Proteins Present in Bean Nuclear Extracts**

By combining mobility-shift assays with DNase I and in situ Op-Cu footprinting experiments, we have shown that the region between positions −628 and −682 upstream from the transcriptional start site of the \( \beta \)-phaseolin gene binds nuclear proteins present in crude nuclear extracts from French bean immature cotyledons. Footprinting of the protein-DNA complexes indicated that the major region of binding is located between nucleotides −670 and −626 from the cap site. The size of the protected region (which spans at least five full turns of the DNA double helix) suggests that more than one protein molecule is present in the complex. Hypersensitivity to DNase I cleavage was observed at positions −643 and −642, roughly in the middle of the region of protection. Additional sites of increased sensitivity were observed in sequences flanking the region of protection (see Figure 3) downstream of position −617 in the coding strand and upstream of position −676 in the noncoding strand. This region coincides with one of the sites reported to be hypersensitive to S1 nuclease (centered at −670) in supercoiled plasmid DNA containing sequences of the phaseolin gene promoter (Murray and Kennard, 1984). Interestingly, the DNase I hypersensitivity effect extends even to flanking vector sequences, suggesting that it may be the consequence of a protein-induced conformational change. The observed differences in protection detected by both footprinting techniques could be explained by stabilization of the protein-DNA complexes inside the gels, i.e., by the “caging” mechanism proposed by Fried and Crothers (1981), combined with the different sequence specificities of Op-Cu (Veal and Rill, 1988) and DNase I (Drew, 1984; Ward et al., 1988). In situ Op-Cu footprinting has been used suc-
cessfully to localize positions of binding sites for trans-
acting factors in procaryotes (Kuwabara and Sigman,
1987) and animal cells (Kakkis and Calame, 1987). Our
results demonstrate the applicability of this technique to
the study of protein-DNA interactions in plant genes.

The nucleotide sequence of the protected region is very
A/T-rich, as was noted for DNA elements involved in the
binding of plant trans-acting factors to 5' regions of soy-
bean lectin (Jofuku, Okamuro, and Goldberg, 1987) and
leghemoglobin (Jensen et al., 1988) genes. Results ob-
tained by mobility-shift experiments with oligo I indicate
that the nucleotides at the 5' border of the protected
region are not sufficient to confer binding. In contrast, the
results with mutant oligo IV (Figure 6) may imply that
structural features of A/T-rich DNA sequences in general,
and of the phaseolin sequence in particular, are as, or
more, important than is a particular nucleotide sequence
for protein binding. Such loose sequence binding require-
ments have been reported for yeast activator proteins
(Pfeifer, Prezant, and Guarente, 1987) and for a mamma-
lian high mobility group protein (Solomon, Strauss, and
Varshavsky, 1986). Alternatively, the same results could
mean that the nucleotides changed in oligo IV are not
absolutely essential, and mutations to complementary (but
not to noncomplementary bases: oligo V) are compatible
with efficient binding. Within the upstream protected re-
gion, two short motifs, "TTAATT TTAA" (−650 to −640)
and in reverse orientation "AATATTTTAA" (−657 to
−667), can be discerned that delimit an imperfect palin-
dromic sequence. These two motifs are homologous to a
consensus "AATATTTTATT" NBF binding motif found up-
stream of light-regulated genes (Datta and Cashmore,
1989), to a sequence from a soybean leghemoglobin gene,
"AATATTTTA" (Jensen et al., 1988), and to a
"TTTATTTTGT" motif present in a region found upstream
of a sunflower helianthinin gene that competes with the
phaseolin 103SD fragment for binding of both bean and
sunflower cotyledon nuclear proteins (Jordano, Almo-
guera, and Thomas, 1989). The significance of these se-
quence similarities remains uncertain, although a possible
function in protein binding for at least the −650 to −640
sequence is suggested by the fact that three of the four
T→G transversions that abolish the binding activity of oligo
VI reside within it. Methylation interference experiments are
being performed to determine the nucleotides con-
tacted by the factor(s) that bind the A/T-rich region.

The β-Phaseolin Upstream A/T-Rich Sequence
Functions as a General Transcriptional Activator in
Tobacco

The ability of the upstream A/T-rich sequence to activate
transcription from a minimal 35S promoter in transgenic
tobacco plants supports the overall conclusion of this
study, namely, that it represents a cis-acting regulatory
element. Yet, a direct correlation between its ability to bind
nuclear proteins in vitro and its behavior as a transcriptional
activator in tobacco cannot be established from the data
reported here. Such confirmation should derive from mu-
tageneration experiments that are under way. In association
with the −89 to +6 region of the CaMV 35S promoter,
this sequence element yielded a pattern of expression that
differed from those observed with either the entire phase-
olin or the entire 35S promoters (compare Figures 1, 6,
and 9). The CaMV 35S −89 to +6 region contains a TATA
element typically necessary for transcriptional activity and
a −89 to −46 region required for activation by the CaMV
35S enhancer (Fang et al., 1989). This region, however, is
incapable of activating transcription per se. All the con-
structs containing either the 103SD region or the synthetic
oligo III showed GUS activity in developing embryos and
in four vegetative organs of developing seedlings. Nuclear
proteins that bind the 103SD probe can also be demon-
strated in vegetative organs of P. vulgaris (i.e., hypocotyls).
Op-Cu footprinting experiments indicate that the hypocotyl
and immature cotyledon factors bind the same sequences
within the −682 to −628 region of the β-phaseolin gene
(M. Bustos and J. Jordano, unpublished observations).
Therefore, the distribution of trans-acting factors in embryo-
onic and vegetative tissues would appear to correlate with
the observed activity of the A/T-rich enhancer in tobacco.
The temporal regulation and the distribution of GUS
expression in 55-4 embryos were different from those seen
in pβ+20/GUS embryos, indicating that this region is not
likely to respond to the same signals that act upon the
entire phaseolin regulatory region represented by con-
struction pβ+20/GUS (Figure 1). Recently, a comparison
of the patterns of accumulation of 47 different types of
mRNAs during cotton embryogenesis revealed that they
can be grouped into 11 classes of coordinately regulated
genes, each responding to a particular combination of five
temporal abundance components (Hughes and Galau,
1989). β-Phaseolin mRNA, like a number of seed storage
protein genes, belongs to the maturation (reserve accu-
mulation) component class. It would be of interest to
extend such classification to other plant model systems
such as tobacco, and to determine to which, if any, of
those categories the mRNAs from 103SD and oligo III
chimeric constructions belong. Equally important is the
question of the actual function of the phaseolin upstream
transcriptional activator in French beans. It is possible that
the 800-bp portion of the phaseolin gene we have referred
to in this study as the phaseolin regulatory region com-
prises multiple cis-acting regulatory elements that function
either independently or synergistically. Such a scheme has
been proposed for plants (Goldberg, Barker, and Perez-
Grau, 1989) and other eukaryotes (mammalian cells and
yeast, Maniatis, Goodbourn, and Fischer, 1987). In accord-
ance with such a model, more or less general positive
regulatory elements, possibly redundant and located distal
to the core promoter, could enhance the activity of a
proximal "qualitative" region that is primarily responsible for the tissue specificity and possibly the temporal regulation.

METHODS

DNA Fragment Isolation and Labeling

Phaseolin DNA fragments were isolated from subclone p1.6, which contains the upstream region of β-phaseolin gene λ177.4 (Slightom, Sun, and Hall, 1983). Restriction fragments to be used as probes were separated on low melting point agarose gels (FMC) and purified by phenol extraction (Maniatis, Fritsch, and Sambrook, 1982). Fragments were end-labeled with either the Klenow fragment of DNA polymerase I or polynucleotide kinase in the presence of α-32P-ATP or γ-32P-ATP (Du Pont-New England Nuclear), respectively, recovered from acrylamide gels by electroelution into dialysis tubing (Maniatis, Fritsch, and Sambrook, 1982) and purified by phenol extraction and ethanol precipitation. For footprinting analysis, plasmid p103SD labeled at either the EcoRI or BamHI sites was subsequently digested with PvuII to yield probes of suitable length labeled only at one end.

Preparation of Crude Nuclear Extracts

Plant materials were frozen in liquid nitrogen and stored at −80°C until extracted. Nuclear extracts were prepared from immature bean (cv Tendergreen) cotyledons (12 mm to 15 mm long) or from 6-day-old bean seedling hypocotyls, essentially as described (Dignam, Lebovitz, and Roeder, 1983; Jensen et al., 1988). Homogenates were not centrifuged through a Percoll-containing buffer; further homogenization was performed with a Dounce homogenizer instead of by sonication, and, after extraction, nuclear proteins were dialyzed against buffer D (Dignam, Lebovitz, and Roeder, 1983). Protein concentrations, measured by the method of Bradford (1976), were approximately 3 mg/mL. The cytoplasmic extracts contained minor amounts of phaseolin and significant levels of nuclear proteins, as indicated by their cross-reactivity with a monoclonal antibody against Physarum B-36 nucleolar proteins (Guiltinan et al., 1988).

Synthetic Oligonucleotides

Oligonucleotides (oligos) were synthesized using a Beckman System 1 Plus DNA synthesizer with reagents and conditions supplied by the manufacturer. After deprotection, oligos were purified on 20% polyacrylamide denaturing gels and Sep-Pak C18 cartridges (Waters Associates, Milford, MA). Oligos were labeled using poly-nucleotide kinase and γ-32P-ATP (Maniatis, Fritsch, and Sambrook, 1982). Equivalent amounts of complementary strands were dissolved in 0.2 M NaCl and annealed by heating to 55°C, followed by slow cooling to 4°C for 2 hr. The single-stranded and double-stranded oligos were then purified by two phenol extractions, precipitated with ethanol, and washed several times with 70% ethanol. Dried oligos were resuspended in 0.2 M NaCl.

DNA Binding Assays

Labeled DNA fragments (0.1 ng, 4000 cpm to 7000 cpm) were incubated with nuclear extracts (2 μg to 6 μg of protein) in the presence of 1.5 μg of poly(dl-dC)-poly(dl-dC) (Pharmacia LKB Biotecnology Inc.). Binding reactions were carried out at 25°C for 15 min in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.4 mM DTT, 5% glycerol, and (only in reactions containing oligos) 0.2 M NaCl. For competition experiments, all DNAs were mixed together prior to the addition of nuclear extract. The reaction products were separated by electrophoresis through 5% polyacrylamide gels containing 0.4 × TBE (for reactions with oligos) or 1 × TBE (Maniatis, Fritsch, and Sambrook, 1982). 0.4 × TBE gels were run with buffer recirculation. Following electrophoresis the gels were dried onto 3MM paper and autoradiographed.

Footprinting Experiments

Protein-DNA binding reactions were essentially as described above and contained approximately 1.5 ng of end-labeled DNA fragments and 7 μg to 40 μg of nuclear protein extract in a total volume of 50 μL. Optimized binding reactions resulting in specific retardation of at least 80% of the probe (as determined by mobility-shift analysis) were used for DNase I footprinting experiments performed as described previously (Jordano and Perucho, 1988). For Op-Cu protection experiments, protein-DNA complexes were first separated on preparative 5% polyacrylamide gels. Footprinting reactions were carried out in the gels for 15 min at 25°C according to Kuwabara and Sigman (1987). Following footprinting, the gel was exposed to x-ray film for 3 hr at room temperature. Bands of interest were cut from the gel, and DNA was eluted overnight at 37°C in 0.5 M ammonium acetate and 1 mM EDTA. The eluted DNA was extracted twice with phenol-chloroform-isooamylic alcohol, and ethanol-precipitated. DNA pellets were washed with 70% ethanol, resuspended in 80% (v/v) formamide, 10 mM NaOH, 1 mM EDTA, 0.1% bromphenol blue, and 0.1% xylene cyanol, and analyzed on 6% polyacrylamide sequencing gels.

Plasmid Constructions and Tobacco Plant Transformation

Plasmid pB120.x was used for expression in transgenic tobacco plants. A derivative of pB121 (Jefferson, Kavanagh and Bevan, 1987), this plasmid contains an inactive CaMV 35S promoter truncated at the EcoRV site (R. Jefferson, personal communication). This truncated 35S promoter retains one half of an inverted repeat and the TATA and CAAT boxes. Constructions 55-4 and 55-3 were obtained by ligating the double-stranded 55-bp oligo III (Figure 5) into the Smal site of plasmid pB120.x. Clones 103SD and 103DS contain a restriction fragment in both orientations (see Figure 6) obtained from the phaseolin upstream region by partial digestion with SspII, followed by complete digestion with Dral, also ligated into the Smal site of pB120.x. The resulting clones were subsequently sequenced to determine the orientation of the inserted sequence relative to the deleted 35S promoter. Plasmids were transferred into Agrobacterium tumefaciens strain LBA4404 (Hoekema et al., 1985) by triparental mating as described by Matzke and Matzke (1986) and used to transform
Nicotiana tabacum cv xanthi by the leaf disc procedure of Horsch et al. (1985). Transformed shoots were selected on 250 μg/mL kanamycin. All data presented here were obtained from self-pollinated plants.

RNA Extraction, Gel Blot Hybridization, and S1 Nuclease Analysis

Seedlings and immature seeds were frozen in liquid nitrogen, and total nucleic acids were extracted by a method described previously (Bustos et al., 1988). RNA was precipitated with 2.6 M sodium acetate and reprecipitated with ethanol. Total RNA aliquots were denatured with glyoxal/Me2SO (Maniatis, Fritsch, and Sambrook, 1982) and separated on 1% agarose, 10 mM phosphate (pH 7.0) gels. After electrophoresis, the RNA was electro-transferred overnight at 10 V onto Nytran membranes. The membranes were subsequently baked in vacuo for 2 hr and hybridized to a GUS-specific DNA probe labeled by the random primed method (Feinberg and Vogelstein, 1983) in the presence of α-32P-dATP (Du Pont-New England Nuclear, 3000 Ci/mmol) in 50% formamide, 4 × SSC, 4 × Denhardt’s solution, 0.5% SDS, and 100 μg/mL denatured salmon sperm DNA. S1 nuclease analysis was done essentially as described (Maniatis, Fritsch, and Sambrook, 1982), using a single-stranded DNA probe prepared by Klenow extension of an end-labeled, GUS-specific synthetic primer followed by strand separation on polyacrylamide gels, and hybridized to 50 μg of total cellular RNA in 80% formamide-containing buffer at 37°C overnight. DNA-RNA hybrids were treated with 75 units of S1 nuclease, precipitated in the presence of tRNA as carrier, and sized on urea-sequencing gels.

GUS Assays

Tissues were ground in lysis buffer containing 50 mM NaPO4, 10 mM EDTA, 0.1% Sarkosyl, 0.1% Triton X-100, and 10 mM β-mercaptoethanol. The samples were centrifuged, and the supernatant was transferred to clean Eppendorf tubes. Protein concentrations were determined by the method of Bradford (1976). Extracts (100 μL) were mixed with an equal volume of 4-

Histochemical in Situ Localization of GUS Activity

Embryos were dissected under a microscope and incubated in Eppendorf tubes or microtiter plates with reaction buffer (1.5 mM 5-bromo-4-chloro-3-indoly-β-D-glucuronide, 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.1 M NaPO4, (pH 7.0), and 10 mM dithiothreitol) at 37°C overnight. After staining, the embryos were rinsed in phosphate buffer (pH 7.0), fixed for 2 hr at room temperature in 5% glutaraldehyde, rinsed again and infiltrated with 50% glycerol on glass slides, and covered. Alternatively, cryo-sections 20 μm to 30 μm thick were made prior to mounting and photographing with Kodak Ektachrome 160 ASA tungsten film.

ACKNOWLEDGMENTS

We thank Dr. Terry L. Thomas and other colleagues for helpful discussions, Jilliane M. Maes for her many excellent contributions, Nasreen Z. Ehtesham, Molly Trinies, and Concepcion Almoguera for technical help, and Helga Sittert-Bhatkar for her cooperation with the histochemical studies. This work was supported by grant DDCB 86-02497 from the National Science Foundation. The use of the Electron Microscopy Center, Texas A & M University, is acknowledged.

Received June 16, 1989; revised July 14, 1989.

REFERENCES


Regulation of beta-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a French bean beta-phaseolin gene.
M M Bustos, M J Guiltinan, J Jordano, D Begum, F A Kalkan and T C Hall
DOI 10.1105/tpc.1.9.839

This information is current as of July 9, 2017