The Plant Cell, Vol. 1, 623–631, June 1989 © 1989 American Society of Plant Physiologists

# Nuclear Factors Interact with a Soybean β-Conglycinin Enhancer

**Randy D. Allen, François Bernier, Philip A. Lessard, and Roger N. Beachy[1]**

Department of Biology, Washington University, Saint Louis, Missouri 63130

**Upstream sequences of the gene encoding the α′ subunit of β-conglycinin were analyzed for interactions with nuclear proteins from immature soybean seeds. Two factors were identified that interact with specific sequence elements within 257 base pairs 5′ of the transcription start site. One factor, SEF 3, binds exclusively to a region composed of two elements located at −183 to −169 base pairs and −153 to −134 base pairs relative to the start of transcription. Each of these sites includes the hexanucleotide sequence AACCCA, which may serve as a primary recognition sequence. During seed development, SEF 3 binding activity was found to increase in soybean embryos during the time of β-conglycinin synthesis and to decrease as seeds neared maturity. The position of the SEF 3 binding sequence corresponds with a previously reported seed-specific enhancer region, and it seems likely that this factor may act as a positive regulator of transcription of the β-conglycinin, α′ subunit gene in developing soybean seeds. The second factor, SEF 4, also binds within the −257 to −77 region but also interacts with sites located further upstream.**

## INTRODUCTION

Many plant genes are expressed in a developmentally regulated or organ-specific manner. In several cases, 5′-flanking sequences that are involved in regulating the expression of these genes have been characterized (see Kuhlemeier, Green, and Chua, 1987; Schell, 1987 for reviews). Nuclear proteins have been identified that bind to specific upstream sequences of several regulated genes (Jofuku, Okamuro, and Goldberg, 1987; Maier et al., 1987) and, in a few cases, protein binding sites have been shown to correspond with sequences that have regulatory function (Green, Kay, and Chua, 1987; Jensen et al., 1988; Green et al., 1989).

β-Conglycinin is a soybean seed storage protein that consists of three primary subunits: α, α′, and β. These subunits are encoded by separate genes that are expressed exclusively in maturing seeds. Significant differences in the level and timing of expression of the α′ and β subunit genes have been reported (Hill and Breidenbach, 1974; Meinke, Chen, and Beachy, 1981; Walling, Drews, and Goldberg, 1986). The α′ subunit gene is expressed earlier and at higher levels than the β subunit gene. Similar expression patterns were observed when genes encoding the α′ and β subunits were transferred to tobacco and petunia plants (Beachy et al., 1985; Bray et al., 1987; Naito, Dube, and Beachy, 1988). Important regulatory sequences of the α′ subunit gene have been localized between −257 bp and −69 bp relative to the transcription initiation site by deletion analyses of the 5′-flanking sequences (Chen, Schuler, and Beachy, 1986). When this region was inserted upstream of the cauliflower mosaic virus 35S promoter, it specifically enhanced expression of a marker gene in maturing seeds. This enhanced expression occurred independently of the orientation of the fragment (Chen, Pan, and Beachy, 1988). This indicates that DNA sequences between −257 and −77 bp of the α′ subunit gene can function as a seed-specific transcriptional enhancer.

To analyze the role of *cis*-acting elements in the 5′-flanking region of the α′ subunit gene in gene expression, we have tested these sequences for interactions with proteins in nuclear extracts from developing soybean seeds. Two factors were found to bind specific domains within this important region. One of these factors apparently interacts with two closely spaced upstream sequence elements that correspond with enhancer activity.

## RESULTS

### DNA Binding Proteins Interact with β-Conglycinin 5′ Sequences

Sequences located between −257 and −77 bp of the transcription initiation site of the α′ subunit gene of β-

[1] To whom correspondence should be addressed.

conglycinin have previously been implicated in seed-specific enhancement of transcription (Chen, Schuler, and Beachy, 1986; Chen, Pan, and Beachy, 1988). DNA fragments within this region were assayed, as shown in Figure 1, for specific DNA sequences that interact with protein factors extracted from nuclei of immature soybean seeds. The presence of bands that have reduced electrophoretic mobility indicates interaction of the probe with nuclear proteins. The DNA binding proteins that induce these shifted bands were termed soybean embryo factors (SEF) 2, 3, and 4, as indicated in Figure 1B. The specificity of the interaction of SEF 2, 3, and 4 with DNA sequences within the −257 to −77 region was investigated by adding unlabeled competitor DNA to binding reactions. These competitor DNAs were derived by digestion of the 5′-flanking sequence of the $\alpha'$ gene (BamHI-SphI fragment, −830 to −77) with RsaI to give the following fragments: $\alpha$1, −830 to −608; $\alpha$2, −607 to −501; $\alpha$3, −500 to −228; $\alpha$4, −227 to −77 (Figure 1A). Addition of 50-fold molar excess of any of these upstream competitor fragments was found to efficiently eliminate binding of SEF 2 to the $\alpha$4 probe (Figures 1A and 1B), and addition of 100-fold excess of MspI-digested *Escherichia coli* DNA also competed for SEF 2 binding activity (data not shown). Interaction of SEF 3 with the $\alpha$4 probe is eliminated by the presence of 50-fold molar excess of unlabeled $\alpha$4 competitor DNA, and $\alpha$3 DNA also competes somewhat for SEF 3 binding activity. The SEF 3/DNA complex is not affected by 50-fold excess of the other competitor fragments, nor is it affected by *E. coli* DNA (data not shown). Although $\alpha$3 DNA fragments (−500 to −227) do compete weakly for SEF 3 binding (Figure 1B), competition with this fragment is much less effective than with $\alpha$4 fragment (see below). Binding of SEF 4 is competed by $\alpha$1, $\alpha$3, and $\alpha$4 DNA, but is not competed by $\alpha$2 DNA (Figure 1B), nor by *E. coli* DNA (data not shown). These analyses indicate that two activities (SEF 3 and 4) bind to specific regions of this DNA fragment and induce shifted bands in our assay. The DNA binding activity of these factors was found to be sensitive to heat treatment (65°C for 5 min) and proteinase K digestion, which indicates that these factors are proteinaceous in nature (data not shown). An additional DNA binding activity (SEF 2) also interacts with this fragment; however, since SEF 2 binding is eliminated by competition with 20-fold to 50-fold molar excess of all of the upstream fragments, *E. coli* DNA, or pUC19 DNA (Figure 1B and data not shown), its interaction appears not to be sequence-specific. The SEF 2 band is also not seen if poly dA:T is used in the binding reactions rather than poly dI:dC (data not shown).

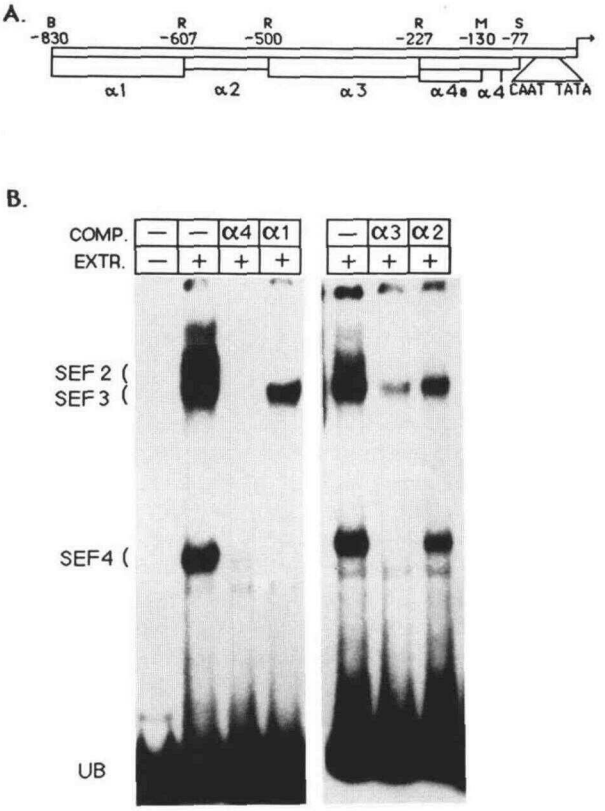


**Figure 1.** Interaction of Nuclear Factors with Upstream Sequences of the $\beta$-Conglycinin, $\alpha'$ Subunit Gene.

**(A)** Map of 5′-flanking sequence of the $\alpha'$ subunit gene. Subcloned fragments used as competitors in DNA:protein binding reactions are indicated. Restriction sites are B, BamHI; M, MboII; R, RsaI; S, SphI.

**(B)** Electrophoretic band shift assay. Binding reactions that included 2.5 $\mu$L (except where noted) of nuclear extract (EXTR) derived from immature soybean seeds 32 days after pollination were incubated with $^{32}$P-labeled $\alpha$4 DNA fragment (−227 to −77). Shifted bands resulting from binding of factors are indicated as SEF 2, SEF 3, and SEF 4. UB, unbound probe. Competition reactions with 50-fold molar excess of unlabeled DNA fragments are indicated (COMP).

The relative affinity of SEF 3 for upstream fragments was determined by titration of SEF 3 binding from a specific probe ($\alpha$4a) with $\alpha$4, $\alpha$3, and $\alpha$1 competitor fragments, as seen in Figure 2. Addition of fivefold molar excess of $\alpha$4 fragment competes for SEF 3 as effectively as a 100-fold excess of $\alpha$3 fragment. This analysis indicates that $\alpha$4 has approximately 20-fold greater affinity for SEF 3 than does the $\alpha$3 fragment. No detectable competition was observed with $\alpha$1 fragments, even at 100-fold excess over labeled probe. Since SEF 3 does appear to bind specifically to sequences with known enhancer activity and does not bind detectably to other regions of the $\alpha'$ subunit gene upstream sequences, we have concentrated our efforts on the characterization of this factor.

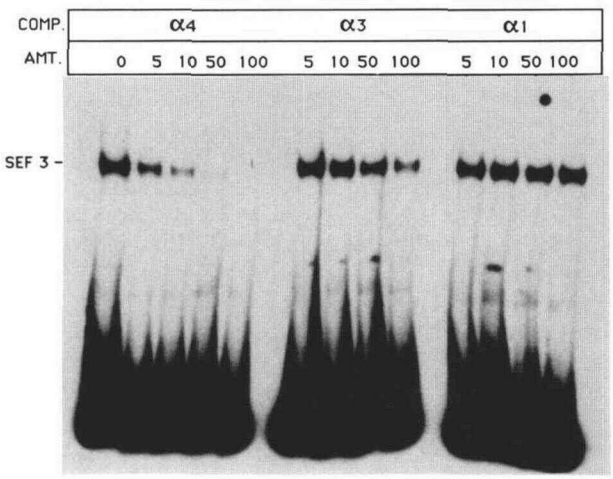


**Figure 2.** Titration of SEF 3 Binding to α4a Probe by Competition with α4, α3, and α1 Fragments in Band Shift Assays.

Fragments were added in amounts equal to fivefold, 10-fold, 50-fold, and 100-fold molar excess (AMT).

### Localization of Binding Sites

The binding site for SEF 3 was initially localized by band shift analysis to a 60-bp BbvI-MboII restriction fragment (−193 to −133). We have also determined the approximate locations of three SEF 4 binding sites by band shift analysis of restriction fragments to DNA sequences between −140 to −77, −350 to −257, and −830 to −608 (data not shown). Although one of the SEF 4 binding sites is located within the region previously shown to have enhancer activity (Chen, Schuler, and Beachy, 1986), the other sites are outside of the region. Therefore, we have chosen not to include an extensive analysis of this factor in this study. The possible regulatory roles of this DNA binding factor are currently under investigation and its characteristics will be reported later.

Analysis of the SEF 3 binding site by DNase I protection indicated a large, poorly defined footprint that spanned approximately 60 bp from −190 to −130 (data not shown). Since DNase I protection analysis did not provide sufficient information about the SEF 3 binding domain to allow accurate localization of its position, exonuclease III protection assays were used to map the 5′ border of the SEF 3 binding region, as shown in Figure 3. This analysis provided a much more detailed representation of the SEF 3 binding site than did DNase I footprinting experiments. Protein-dependent, exonuclease III stop positions, which represent barriers to exonuclease III digestion caused by binding of specific proteins, occur in at least two distinct regions of the α4 probe. A series of bands is seen starting

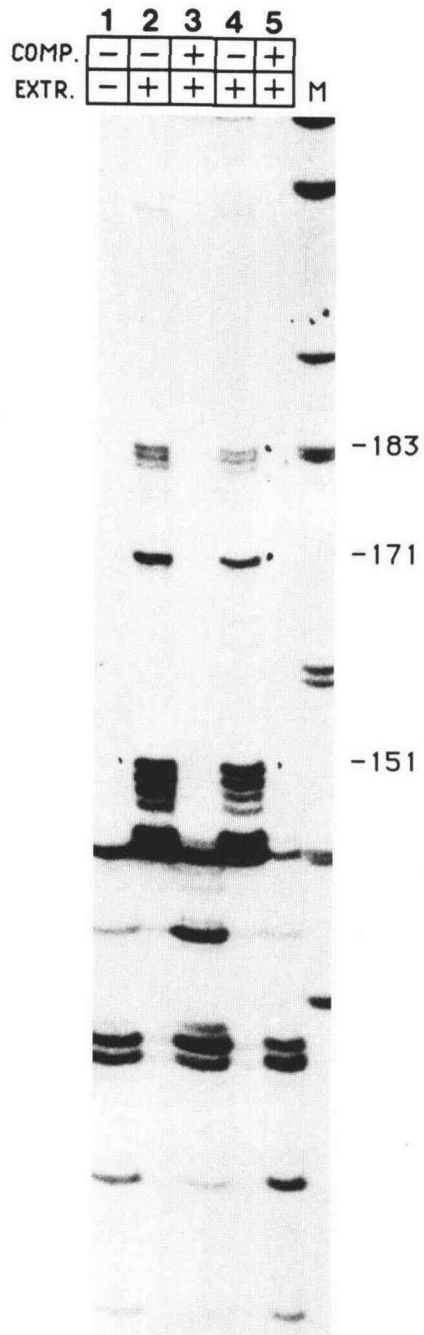


**Figure 3.** Exonuclease III Protection Assays.

Binding reactions included 5 μL of nuclear extract. Probe was α4 DNA fragment (−227 to −77). Competition reactions included 50-fold molar excess of unlabeled α4 DNA. Strong exonuclease-III stops are labeled indicating their position relative to the cap site. EXTR = nuclear extract. Lanes 2 and 3, unprecipitated nuclear extract; lanes 4 and 5, nuclear extract precipitated with 200 mg/mL $(NH_4)_2SO_4$. M, size markers, pBR322 DNA digested with MspI.

at approximately −183 and another set of bands is located further downstream, approximately −153 bp from the transcription initiation site. An additional band is located at about −171. These exonuclease III stops are efficiently eliminated by competition with α4 DNA fragments (−227 to −77) but not with other fragments from the α′ upstream sequence (data not shown), which indicates that these blocks to exonuclease III digestion are due to the presence of sequence-specific DNA binding factor(s). This complex pattern suggests that SEF 3 may contact the DNA at multiple sites.

The possibility that the SEF 3 binding region includes multiple sites of interaction was investigated by extensive band shift analysis using a series of probes that include linker-scan mutations, as well as 5′ and 3′ end deletions within the −257 to −77 region. The results of this analysis are shown in Figure 4. The linker-scan mutations used were: LS 1 (sequences from −182 to −171 replaced by XbaI linker), LS 2 (−174 to −162 replaced), LS 3 (−168 to −156 replaced), LS 4 (−201 to −190 replaced), and LS 5 (−149 to −130 replaced) (Figure 4B). The relative affinity of these probes for SEF 3 was determined by liquid scintillation counting of bands cut from band shift gels like that in Figure 4A. Since the introduced mutations are located 5′ of the SEF 4 binding site, they do not affect the binding of SEF 4. Therefore, the level of SEF 4 binding was used as an internal control to normalize for differences in specific activity of the various probes.

The most 5′ mutation, LS 4, does not affect binding of SEF 3. This result was not unexpected since the lesion lies within a region that can be safely deleted without adversely affecting binding (see Figure 4C). Binding of SEF 3 to a probe that includes LS 1 is reduced significantly, to approximately 15% of wild-type level. The presence of LS 2 reduces binding to about 40% of wild-type levels, but LS 3 does not have a detectable effect on the binding of SEF 3. The mutation LS 5 also significantly affects binding of SEF 3, reducing it to about 5% of wild-type.

Deletion of sequences from the 5′ end to −187 has no detectable effect on SEF 3 binding (Figure 4C); however, further deletion to −181 reduces binding by about 50%. Deletion of sequences to −173 reduces binding still further, and, when sequences to −150 are deleted, SEF 3 binding is abolished. Corresponding deletions from the 3′ end through −133 did not affect binding. Partial loss of binding was observed when sequences to −146 were deleted, and complete loss of binding activity occurred with deletion of sequences to −166. Analysis of these 5′ and 3′ deletions, along with the linker-scan mutations, establishes the locations of two sequence elements that are involved in binding SEF 3. The 5′ SEF 3 binding site is located between −183 and −169 and the 3′ site between −153 and −134. Since the LS 3 probe retains high affinity for SEF 3, sequences between the two active sites do not appear to interact specifically with SEF 3. These data correlate well with the locations of the exonuclease III protected regions discussed above. We believe that this analysis demonstrates that the SEF 3 binding domain is composed of two elements located between −183 and −169 bp and between −153 and −134 bp from the transcription start site. Comparison of these two elements reveals a region of significant homology, as shown in Figure 5. The hexanucleotide AACCCA is present in both the 5′ and 3′ SEF 3 binding sites. It is likely that this element may be the core of the SEF 3 recognition sequence.

The sequence requirements for SEF 3 binding were further characterized using synthetic double-stranded oligonucleotides. A 32-bp double-stranded oligonucleotide

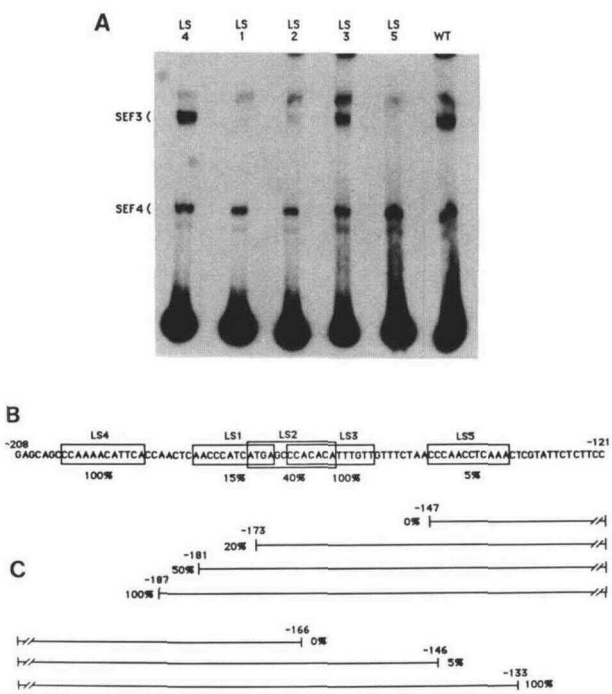


**Figure 4.** Analysis of SEF 3 Binding Site by Linker-Scan Mutation.

**(A)** Electrophoretic band shift assay showing differential binding of SEF 3 and SEF 4 to probes that span −257 to −77 and include linker-scan mutations (LS 1 to LS 5). Reaction conditions were as described for Figure 1. Locations of linker-scan mutations are given in Results and in **(B)**. WT = wild-type probe (−257 to −77 probe with no introduced linker-scan mutations).

**(B)** Sequence of α′ subunit gene from −208 to −121 with positions of linker-scan mutations (LS 1 to LS 5) indicated by boxes. Relative intensities of SEF 3:DNA complex bands for linker-scan containing probes were determined and are presented as percentages of SEF 3 binding to wild-type probe.

**(C)** Deletions from 5′ and 3′ ends of the −257 to −77 probe. SEF 3 binding efficiencies relative to full-length probe are given. Deletion endpoints are indicated.

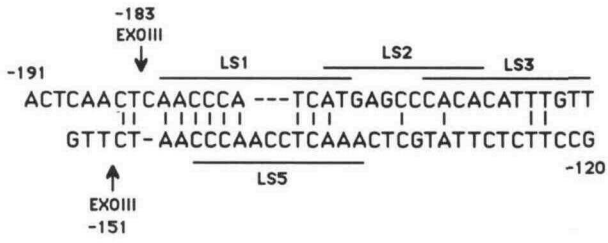


**Figure 5.** Sequence Comparison of Two SEF 3 Binding Sites within the −257 bp to −77 bp Region.

Gaps introduced to maximize the match are indicated with dashes. Positions of linker-scan mutations LS 1, LS 2, LS 3, and LS 5 are indicated. The hexanucleotide sequence AACCCA is the only significant homology apparent between the two sites. Positions of two exonuclease III stop sites are indicated (see Figure 3).

was prepared that included the putative 3′ SEF 3 binding element. This molecule was cloned into pUC19 and a second copy of the oligonucleotide was subsequently added. The dimer included two SEF 3 binding elements separated by 27 bp. This spacing is identical to that in the native α′ gene upstream sequence. A second double-stranded oligonucleotide was prepared in which the AACCCA sequence was altered to AAGGCA. A dimer of this molecule was also subcloned into pUC19. Probes prepared from these plasmids were tested for their ability to bind SEF 3 by band shift assays, as shown in Figure 6. The single element probe did not bind SEF 3 detectably under standard conditions (Figure 6, 1); however, significant binding could be detected if the amount of poly dI:dC in the reaction is lowered from 2 μg to 10 ng (not shown). The dimerized probe interacts strongly with SEF 3 (Figure 6, 2). The affinity of this probe for SEF 3 is comparable to that of the native sequence (i.e., probe α4a). The probe that includes the dimer of the mutated SEF 3 binding element has no detectable affinity for SEF 3 (Figure 6, MT). These results indicate that the first two cytosine residues of the AACCCA element are an essential part of the SEF 3 recognition sequence and that duplication of this sequence element is required for efficient binding.

### Developmental Expression of SEF 3 Binding Activity

The relative levels of SEF 3 DNA binding activity in nuclear extracts of maturing soybean embryos were determined by band shift assays using the α4a probe, as shown in Figure 7. Low levels of SEF 3 binding activity were first detected in extracts prepared from embryos 22 days after pollination (DAP). Peak levels occurred in extracts from seeds between 32 DAP and 38 DAP, and SEF 3 activity was significantly reduced by 42 DAP. The presence of SEF 3 DNA binding activity in developing seeds correlates well with the period of β-conglycinin mRNA accumulation. Synthesis of α′ subunit mRNA begins in maturing soybean seeds (cv Provar) at approximately 18 DAP; mRNA levels peak at about 35 DAP and decline thereafter (Meinke, Chen, and Beachy, 1981; Naito, Dube, and Beachy, 1988). SEF 3 activity is detectable only in nuclear extracts from embryos during middle and late stages of maturation (when β-conglycinin mRNA is synthesized). Nuclear extracts from soybean leaves and from pea leaves, petioles, and stems were also tested for SEF 3 activity, as shown in Figure 8. These extracts did not contain SEF 3 activity, but the pea extract did have large amounts of DNA binding activity when assayed with a probe that includes four tandem repeats of the GT-1 binding site from the *RbcS-3A* gene (Figure 8, 682; Green et al., 1989). The correlation of the presence of SEF 3 DNA binding activity with expression of the α′ gene supports the hypothesis that SEF 3 is involved in positively regulating the transcription of β-conglycinin genes.

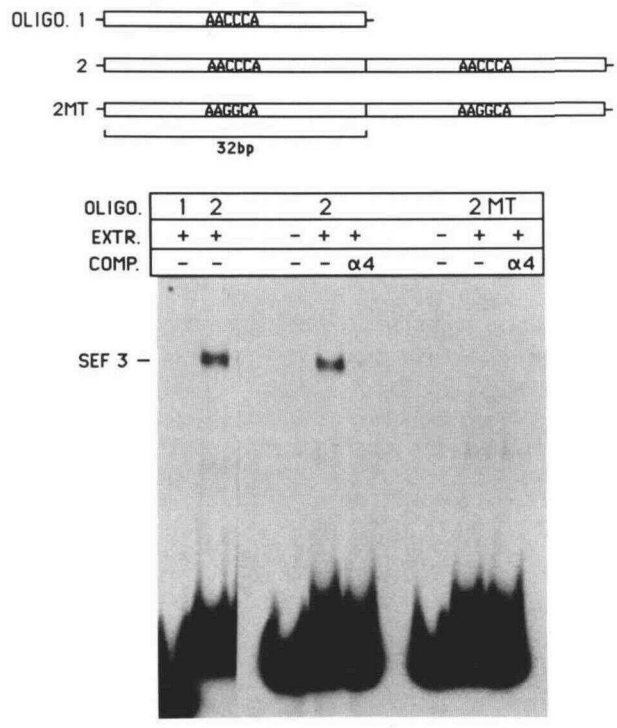


**Figure 6.** Band Shift Assay That Shows Relative Binding of SEF 3 to Double-Stranded Oligonucleotide Probes.

Monomer (1) and dimer (2) probes that include the AACCCA element were tested along wtih a dimer in which the AACCCA elements have been mutated to AAGGCA (2MT). COMP, 50-fold molar excess of α4 DNA.

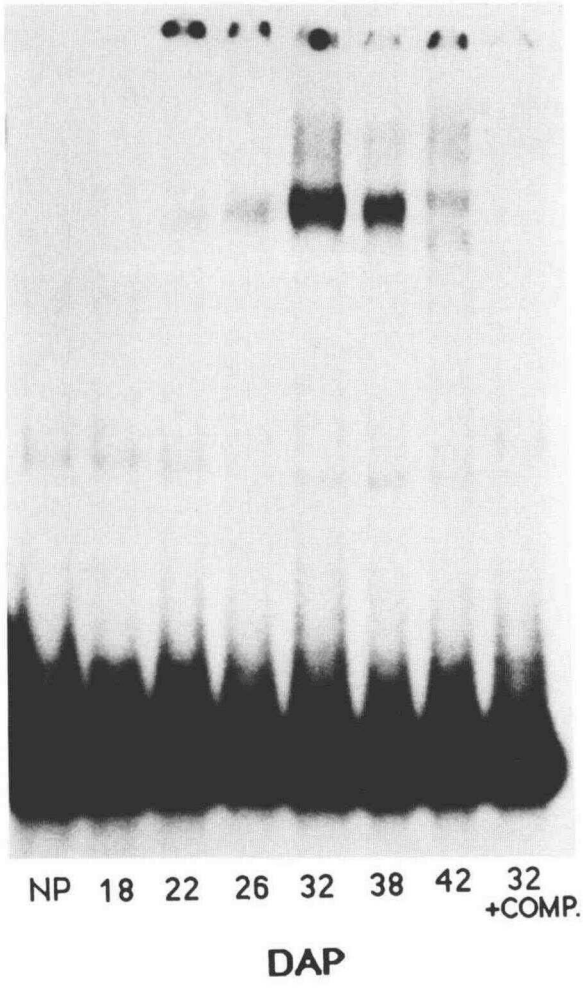


**Figure 7.** Relative Levels of SEF 3 Binding Activity in Nuclear Extracts of Maturing Soybean Seeds Determined by Electrophoretic Band Shift Assay.

All reactions included 2.5 μL of nuclear extract except where noted. Age of seeds is indicated as days after pollination (DAP). Probe was $^{32}$P-labeled α4a DNA fragment (−227 to −133) and does not include the SEF 4 binding site. NP, no nuclear extract; COMP, competition with 50-fold molar excess of α4 DNA fragments.

## DISCUSSION

Several factors that interact with upstream sequences of the β-conglycinin α′ gene have been identified in crude nuclear extracts from immature soybean seeds. Certain characteristics of these factors (SEF 1 through 4) are given in Table 1. SEF 1, which is not described in this paper, binds to A-T-rich sequences located between −607 and −823. Since these far upstream sequences can be deleted without adversely affecting expression of the α′ subunit gene in transgenic tobacco plants (Chen, Schuler, and Beachy, 1986), it seems that this factor does not play a primary role in regulating the expression of this gene. The sequence motif recognized by this factor is similar to that of a binding factor found to interact with A-T-rich sequences upstream of a soybean lectin gene (Jofuku, Okamuro, and Goldberg, 1987). An additional DNA binding activity, SEF 2, appears to bind nonspecifically with DNA fragments throughout the upstream region of this gene.

Two other factors, SEF 3 and 4, interact with sequences between −257 bp and −77 bp from the transcription start site of the α′ subunit gene. This region has been previously found to include *cis*-acting elements involved in the transcriptional regulation of this gene in transgenic tobacco plants (Chen, Schuler, and Beachy, 1986; Chen, Pan, and Beachy, 1988). One of these factors, SEF 3, binds specifically to a site composed of two elements: (a) −183 to −169 and (b) −153 to −131 as determined by exonuclease

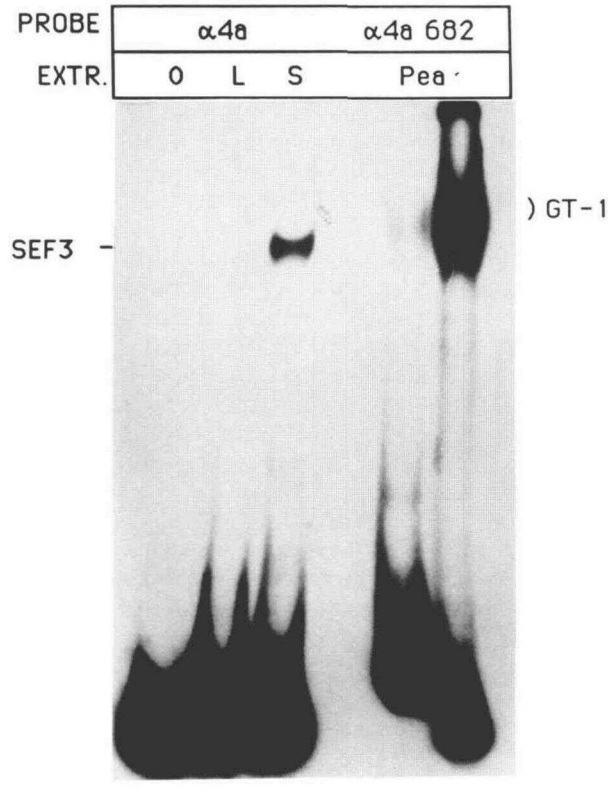


**Figure 8.** Analysis of Nuclear Extracts from Soybean Leaves (L), Immature Soybean Seeds (S), and Pea Leaves and Stems (Pea) for SEF 3 Activity Using the α4a Probe.

The pea extract was also tested for GT-1 activity (Green et al., 1989) using a probe that contains four copies of the core GT-1 binding site (682).

**Table 1.** Sequence-Specific DNA Binding Factors That Interact with Upstream Sequences of the β-Conglycinin α′ Subunit Gene

| Factor | Location of Binding Site(s) | Putative Recognition Motif | Reference |
|---|---|---|---|
| SEF 1 | −670 to −640 | ATATTTAT/AA/T | [a,b] |
| | −790 to −765 | | |
| SEF 3 | −186 to −130 | AACCCA---AACCCA | [c] |
| SEF 4 | −115 to −105 | A/GTTTTTA/G | [a] |
| | −340 to −325 | | |
| | −690 to −670 | | |

[a] R.D. Allen and R.N. Beachy, unpublished data.
[b] Consensus SEF 1 binding site; an analogous binding site was reported by Jofuku, Okamuro, and Goldberg (1987) in a soybean lectin gene.
[c] This report.

III protection assays and analysis of deletions and linker-scan mutations. No other SEF 3 recognition domains were detected between −830 and −77. The hexanucleotide AACCCA, which is present in both sites, has been defined as a likely recognition sequence. The repeating sequence AGCCCA-16 bp-AACCCA, which is present twice between −208 bp and −77 bp of the α′ subunit gene, was proposed by Chen, Schuler, and Beachy (1986) as a possible regulatory sequence. Interestingly, the binding site for SEF 3 centers on each of the AACCCA sequence elements, but the AGCCCA regions do not appear to be important. Finally, SEF 4 interacts with sequences between −115 and −105. This factor also binds to sequences further upstream between −340 and −325 and between −690 and −670. Although deletion of the upstream sequences to −257 does not adversely affect expression of the gene in transgenic plants (Chen, Schuler, and Beachy, 1986), the downstream SEF 4 binding site does lie within the previously described enhancer region (Chen, Pan, and Beachy, 1988). Therefore, the possibility that SEF 4 is involved in regulating the expression of the α′ subunit gene cannot be discounted. This point is currently under investigation.

Chen, Schuler, and Beachy (1986) found that accurate and quantitative expression of the α′ subunit gene in transgenic tobacco plants was detected if the 5′-flanking sequences upstream of −257 were deleted. Further deletion to −208 bp resulted in an approximately 20% reduction in the level of expression, whereas deletion to −159 reduced expression to about 5% that of the full-length gene, and expression was undetectable if sequences to −69 were removed. The two SEF 3 binding elements defined in this study flank the −159 deletion. We believe that the correlation of the SEF 3 binding site with sequence elements required for maximal expression strongly indicates that SEF 3 is directly involved in transcriptional regulation of this gene. Although sequences between −257 and −208 are also necessary for maximum gene expression, we have been unable to detect sequence-specific nuclear factors that bind within this region. It is possible that a yet undetected binding site for SEF 3, or another factor, exists in this region, or that these sequences play a structural role in the recognition of the SEF 3 domain in vivo. The −257 to −208 region does include homology to a consensus sequence known as the legumin box (Bäumlein et al., 1986). Since this sequence is conserved in a number of seed protein genes, it is possible that it may function in a regulatory manner, perhaps in conjunction with the SEF 3 binding site.

Analysis of the binding of SEF 3 to fragments that include LS1 or LS 5 mutations (Figure 4) indicates that loss of one of the two recognition sites reduces the intensity of the SEF 3-DNA complex but does not change its relative mobility. This may indicate that the two SEF 3 binding sites interact with a single, large (probably complex) factor. Since disruption of individual SEF 3 sites does not increase the electrophoretic mobility of the complex, it does not appear that factors interact independently with these sites. This interpretation is supported by two additional observations: (a) The reduction in affinity for SEF 3 in probes with mutations in individual binding sites does not appear to be additive. Disruption of the 5′ site (LS 1) reduces binding by about 85%, whereas loss of the 3′ site (LS 5) reduces binding to an even greater degree. (b) Even under conditions of high probe excess, SEF 3-DNA complexes with increased electrophoretic mobility are not observed. For these reasons, we believe that, at least in vitro, the SEF 3 binding domain is composed of two recognition sites that cooperate to bind a single factor.

The presence of multiple binding sites for *trans*-acting factors is a common feature of eukaryotic enhancers (Schaffner et al., 1988). Heat shock genes from a number of organisms, including soybean, are regulated by multiple heat shock elements (Baumann et al., 1987; Bienz and Pelham, 1987), and an important *cis*-acting region in a soybean leghemoglobin gene includes two *trans*-acting factor binding domains (Jensen et al., 1988). The lightresponsive elements of a pea ribulose-bisphosphate carboxylase gene are repeated (Green, Kay, and Chua, 1987; Kuhlemeier et al., 1988). These light-responsive elements include binding sites for a nuclear factor named GT-1. The core recognition site for GT-1 must be repeated at least three times to bind GT-1 significantly (Green et al., 1989). Our results suggest that two elements are necessary for efficient binding of SEF 3, and additional experiments using oligonucleotides that include single or duplicated SEF 3 elements support this interpretation. Our results also establish that at least part of the AACCCA sequence is required for SEF 3 interaction since alteration of the sequence to AAGGCA abolishes SEF 3 binding.

The localization of binding sites for the nuclear factor SEF 3 within an upstream sequence with seed-specific

enhancer activity suggests that SEF 3 is likely to be a positive transcriptional regulatory protein. Although this factor may be necessary for normal expression of the α′ subunit gene, we do not know whether it, along with ubiquitous promoter factors, is sufficient to promote transcription. Since other specific DNA binding proteins interact with the upstream sequences of this gene, it seems possible that these factors (or other, as yet undetected factors) may also play a role in regulating its expression. Experiments are currently underway to address these questions.

## METHODS

### Materials

Restriction enzymes and DNA modifying enzymes were purchased from Promega Biotec and New England Biolabs. XbaI linkers were obtained from New England Biolabs and calf intestinal phosphatase from Boehringer Mannheim. Leupeptin hemisulfate and poly[d(I-C)]:poly[d(I-C)] were obtained from Sigma. Radioactive nucleotides were obtained from Du Pont-New England Nuclear or Amersham. Enzymes and other reagents were used according to the instructions of the suppliers.

### DNA Probes

Sequences upstream of the α′ subunit gene of β-conglycinin (Doyle et al., 1986) from −830 to −77 (relative to the transcription start site) were cleaved with RsaI and SphI to produce four fragments that were subcloned into pUC19. The inserts are labeled as follows: α1 = −830 to −608; α2 = −607 to −501; α3 = −500 to −228; α4 = −227 to −77 (Figure 1A). An additional fragment (α4a, −227 to −133) was derived from the α4 fragment by digestion with MboII. Probes for band shift assays were prepared from the α4 or α4a subclone by cutting with EcoRI (5′ of the insert) and HindIII (3′ of the insert). The ends of the fragments were filled in with α-$^{32}$P-dATP (3000 Ci/mmol), dCTP, dGTP, and TTP using the Klenow fragment of DNA polymerase I. Labeled fragments were purified by polyacrylamide gel electrophoresis and recovered by electroelution. Unlabeled fragments from all four subclones were purified similarly for use as competitors. Labeled fragment for exonuclease III protection assays was prepared by digesting the plasmid containing α4 with HindIII and labeling with γ-$^{32}$P-ATP (5000 Ci/mmol) using polynucleotide kinase after treatment with calf intestinal phosphatase. The labeled DNA was then cleaved with EcoRI, and the fragment was purified as described above.

Libraries of 5′ and 3′ deletions were prepared by Bal31 digestion of a subclone that contained the −257 to −77 fragment of the α′ subunit gene. The plasmid was first linearized with either EcoRI (for 5′ deletions) or HindIII (for 3′ deletions). Aliquots of the Bal31 reaction mixtures were stopped at 30-sec intervals and pooled and the ends were repaired with Klenow fragment in the presence of deoxynucleotide triphosphates. A 12-bp synthetic XbaI linker was ligated to the repaired ends and the fragments were released by digestion with EcoRI or HindIII. Fragments were sized by polyacrylamide gel electrophoresis and recovered by electroelution. These fragments were subcloned into pUC19 digested with either EcoRI and XbaI (for 5′ deletions) or HindIII and XbaI (for 3′ deletions) and the inserts were accurately sized by electrophoresis on sequencing gels.

Linker-scan mutations were prepared by subcloning appropriate 5′ deletion fragments into corresponding plasmids that contained 3′ deletion fragments with XbaI and HindIII. This procedure reconstituted full-length subclones of the −257 to −77 region in which selected segments were replaced by the XbaI linker.

A synthetic, double-stranded oligonucleotide was prepared that included the 3′ AACCCA element and surrounding sequences. The complete sequence of this oligonucleotide is shown below.

5′-GATCTGTTGTTTCTAACCCAACCTCAAACTCG-3′
3′-ACAACAAAGATTGGGTTGGAGTTTGAGCCTAG-5′

A similar oligonucleotide was also made in which the AACCCA element was changed to AAGGCA. Single and double copies of these molecules were inserted into the BamHI site of pUC19. The inserts were excised and $^{32}$P-labeled as described above.

### Nuclear Extracts

Nuclei were prepared from immature soybean seeds (*Glycine max* cv Provar) essentially as described by Walling, Drews, and Goldberg (1986). Embryos (5 g) were frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. This powder was transferred to a beaker and suspended in 50 mL of buffer A (2.5% ficoll, 5% dextran, 25 mM Tris-HCl, pH 8.5, 10 mM $MgCl_2$, 0.5% Triton X-100, 440 mM sucrose, 10 mM β-mercaptoethanol, 0.2 mM EDTA, 2 mM spermine-HCl, and 10 μg/mL leupeptin hemisulfate). This and all subsequent procedures were performed on ice; all centrifugations were run at 4°C. The slurry was homogenized by three 30-sec bursts with a polytron and filtered through three layers of 70-μm mesh Nytex. Crude nuclei were pelleted by centrifugation for 5 min at 3000*g* and resuspended in buffer B (identical to buffer A but without spermine-HCl). Nuclei were pelleted as before, resuspended in buffer C (20 mM HEPES, pH 7.5, 420 mM NaCl, 25% glycerol, 0.5 mM DTT, 0.2 mM EDTA, and 10 μg/mL leupeptin), and subjected to Dounce homogenization (10 strokes with B pestle). The solution was cleared by centrifugation at 22,000*g* for 30 min, and the supernatant was dialyzed for 5 hr against buffer D (20 mM HEPES, pH 7.5, 50 mM KCl, 20% glycerol, 0.5 mM DTT, 0.2 mM EDTA, and 10 μg/mL leupeptin). The dialysate was recentrifuged at 12,000*g* for 10 min to remove precipitated globulin storage proteins and to float much of the remaining lipid, and the aqueous supernatant was stored at −80°C. These crude nuclear extracts contained about 10 mg/mL total protein.

### Band Shift Assays

Binding reactions were performed in a volume of 20 μL containing 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2 mM EDTA, 1 mM DTT, 100 μg/mL poly[d(I-C)]:poly[d(I-C)] plus 2.5 μL of nuclear extract (25 μg of total nuclear protein). This mixture was incubated for 10 min at room temperature, and then 0.5 ng to 1 ng of probe (2500 cpm to 5000 cpm) was added, and incubation continued for 15 min. After incubation, the reaction mixtures were loaded on 4%

polyacrylamide gels and run in 0.5 × TBE (1 × TBE = 89 mM Tris-borate, pH 8, 2 mM EDTA) at 100 V for approximately 2 hr. Gels were fixed, dried, and exposed to x-ray film with an intensifying screen for about 12 hr at −50°C.

### Exonuclease Protection Assays

Exonuclease III protection assays were performed according to the method described by Hennighausen and Lubon (1987) with the exception that the reaction buffer used was identical to that used for band shift assays. Briefly, 50 μL binding reactions, including 5 μL of nuclear extract, were prepared as described above. After incubation with probe, $MgCl_2$ was added to a final concentration of 5 mM, and 6 units of exonuclease III were added per microliter of reaction mixture. Incubation was continued for an additional 15 min at 30°C. Reactions were stopped by addition of 50 μL of 20 mM EDTA and 1% SDS, and products were purified by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol precipitation. Reaction products were run on 7% sequencing gels that were then dried and exposed to x-ray film for 12 hr at −50°C with an intensifying screen.

## ACKNOWLEDGMENTS

We thank Phat M. Dang for technical assistance and Thomas J. Dietz and Drs. Leigh B. Farrell and David S. Gilmour for reading the manuscript. We also thank Dr. Pamela S. Green for providing the pea nuclear extract and p682. This work was supported by U.S. Department of Energy Biological Energy Research Grants DE-AC02-81ER10888 and DE-FG02-88ER13877 (to R.N.B.). R.D.A. was supported by National Institutes of Health postdoctoral fellowship 1 F32 GM12294-01 and F.B. by a postdoctoral fellowship from Natural Sciences and Engineering Research Council of Canada.

Received March 21, 1989; revised April 17, 1989.

## REFERENCES

**Baumann, G., Raschke, E., Bevan, M., and Schöffl, F.** (1987). Functional analysis of sequences required for transcriptional activation of a soybean heat shock gene in transgenic tobacco plants. EMBO J. **6,** 1161–1166.

**Bäumlein H., Ulrich, W., Pustell, J., and Kafatos, F.C.** (1986). The legumin gene family: Structure of B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucl. Acids Res. **14,** 2702–2720.

**Beachy, R.N., Chen, Z.-L., Horsch, R.B., Hoffmann, S.G., and Fraley, R.T.** (1985). Accumulation and assembly of soybean β-conglycinin in seeds of transformed petunia plants. EMBO J. **4,** 3047–3053.

**Bienz, M., and Pelham, H.R.** (1987). Mechanisms of heat shock gene activation in higher eukaryotes. Adv. Genet. **24,** 31–72.

**Bray, E.A., Naito, S., Pan, N.-S., Dube, P.H., and Beachy, R.N.** (1987). Expression of the β subunit of β-conglycinin in seeds of transgenic plants. Planta **172,** 364–370.

**Chen, Z.-L., Schuler, M.A., and Beachy, R.N.** (1986). Functional analysis of regulatory elements in a plant embryo specific gene. Proc. Natl. Acad. Sci. USA **83,** 8560–8564.

**Chen, Z.-L., Pan, N.-S., and Beachy, R.N.** (1988). A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. EMBO J. **7,** 297–302.

**Doyle, J.J., Schuler, M.A., Godette, W.D., Zenge, V., Beachy, R.N., and Slightom, J.** (1986). The glycosylated seed proteins of *Gycine max* and *Phaseolus vulgaris*. J. Biol. Chem. **261,** 9228–9238.

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

**Green, P.J., Yong, M.-H., Cuozzo M., Kano-Murakami, Y. Silverstein, P., and Chua, N.-H.** (1989). Binding site requirements for pea nuclear factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *RbcS-3A* gene. EMBO J. **7,** 4035–4044.

**Hennighausen, L., and Lubon, H.** (1987). Interaction of protein with DNA in vitro. Methods Enzymol. **152,** 721–735.

**Hill, J.E., and Breidenbach, R.W.** (1974). Proteins of soybean seed. II. Accumulation of the major protein components during seed development and maturation. Plant Physiol. **53,** 747–751.

**Jensen, E.O., Marker, K.A., Schell, J., and de Bruijn, F.J.** (1988). Interaction of a nodule specific *trans*-acting factor with distinct DNA elements in the soybean leghaemoglobin IBC3 5′ upstream region. EMBO J. **7,** 1265–1271.

**Jofuku, K.D., Okamuro, J.K., and Goldberg, R.B.** (1987). Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature **328,** 734–737.

**Kuhlemeier, C., Green, P.J., and Chua, N.-H.** (1987). Regulation of gene expression in higher plants. Annu. Rev. Plant Physiol. **38,** 221–257.

**Kuhlemeier, C., Cuozzo, M., Green, P.J., Goyvaert, E., Ward, K., and Chua, N.-H.** (1988). Localization and conditional redundancy of regulatory elements in *rbcS-3A*, a pea gene encoding the small subunit of ribulose-bisphosphate carboxylase. Proc. Natl. Acad. Sci. USA **85,** 4662–4666.

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

**Meinke, D.W., Chen, J., and Beachy, R.N.** (1981). Expression of storage protein genes during soybean seed development. Planta **153,** 130–139.

**Naito, S., Dube, P.H., and Beachy, R.N.** (1988). Differential expression of conglycinin α′ and β subunit genes in transgenic plants. Plant Mol. Biol. **11,** 109–123.

**Schaffner, G., Schirm, S., Müller-Baden, B., Weber, F., and Schaffner, W.** (1988). Redundancy of information in enhancers as a principle of mammalian transcriptional control. J. Mol. Biol. **201,** 81–90.

**Schell, J.** (1987). Transgenic plants as tools to study the molecular organization of plant genes. Science **237,** 1176–1182.

**Walling, L., Drews, G.N., and Goldberg, R.B.** (1986). Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. Proc. Natl. Acad. Sci USA **83,** 2123–2127.