The Plant Cell, Vol. 1, 415–425, April 1989 © 1989 American Society of Plant Physiologists

Soybean β-Conglycinin Genes Are Clustered in Several DNA Regions and Are Regulated by Transcriptional and Posttranscriptional Processes

John J. Harada,[1] Susan J. Barker, and Robert B. Goldberg[2]

Department of Biology, University of California, Los Angeles, California 90024-1606

We investigated the chromosomal organization and developmental regulation of soybean β-conglycinin genes. The β-conglycinin gene family contains at least 15 members divided into two major groups encoding 2.5-kilobase and 1.7-kilobase embryo mRNAs. β-Conglycinin genes are clustered in several DNA regions and are highly homologous along their entire lengths. The two groups differ by the presence or absence of specific DNA segments. These DNA segments account for the size differences in β-conglycinin mRNAs. The 2.5-kilobase and 1.7-kilobase β-conglycinin mRNAs accumulate and decay at different times during embryogenesis. By contrast, genes encoding these mRNAs are transcriptionally activated and repressed at the same time periods. Our studies indicate that the β-conglycinin family evolved by both duplication and insertion/deletion events, and that β-conglycinin gene expression is regulated at both the transcriptional and posttranscriptional levels.

INTRODUCTION

The soybean β-conglycinin storage protein is a highly prevalent protein that accumulates during seed development, is sequestered within protein bodies, and is hydrolyzed after germination to provide a carbon and nitrogen source for the developing seedling (Higgins, 1984). β-Conglycinin is unique among abundant soybean seed proteins in that it is encoded by two homologous mRNA classes that differ significantly in size (Goldberg et al., 1981; Schuler et al., 1982). The 2.5-kb and 1.7-kb β-conglycinin mRNAs encode two distinct-sized protein subunits designated as α'/α, and β, respectively (Beachy, Jarvis, and Barton, 1981; Schuler et al., 1982; Coates et al., 1985). These subunits interact to form a trimeric protein with different subunit compositions (Thanh and Shibasaki, 1976, 1978). Each mRNA size class represents a family of related mRNAs (Goldberg et al., 1981; Schuler et al., 1982). β-Conglycinin subunit heterogeneity in the seed is due to the translation of these divergent mRNAs as well as posttranslational modification of newly synthesized polypeptides (Beachy et al., 1981; Sengupta et al., 1981; Beachy et al., 1985).

We showed previously that β-conglycinin mRNAs accumulate and decay during specific embryonic periods, are highly prevalent at the midmaturation stage, and are undetectable in mature plant organ systems, and that the accumulation of β-conglycinin mRNAs is controlled in part at the transcriptional level (Goldberg et al., 1981; Walling, Drews, and Goldberg, 1986). Others have shown that individual β-conglycinin genes are differentially expressed during the soybean life cycle. For example, α'/α-subunit and β-subunit mRNAs and polypeptides accumulate at different embryonic stages (Gayler and Sykes, 1981; Meinke, Chen, and Beachy, 1981; Sengupta et al., 1981; Naito, Dubé, and Beachy, 1988), have different prevalences in the embryonic axis and cotyledons (Ladin et al., 1987; Sugimoto et al., 1987), and are differentially affected by sulfur deficiency (Gayler and Sykes, 1985) and by methionine deficiency (Holowach, Madison, and Thompson, 1986). To begin to unravel the precise mechanisms that regulate β-conglycinin genes, we investigated the organization and expression of the β-conglycinin gene family. We found that two distinct gene classes encode β-conglycinin mRNAs, that the β-conglycinin genes differ in size by insertion and/or deletion events, and that β-conglycinin genes are clustered in different genomic regions. We further show that posttranscriptional events play an important role in the differential expression of β-conglycinin gene classes during embryogenesis.

RESULTS

β-Conglycinin Genes Constitute a Large Gene Family

To characterize the β-conglycinin gene family, we selected clones from a λ Charon 4 EcoRI soybean genome library

[1] Current address: Department of Botany, University of California, Davis, CA 95616.

[2] To whom correspondence should be addressed.

(Fischer and Goldberg, 1982) using the cDNA clone A-16 (Goldberg et al., 1981). We showed previously that A-16 represents the 2.5-kb β-conglycinin mRNA (Goldberg, Crouch, and Walling, 1983), and that it contains sequences complementary to the 1.7-kb β-conglycinin mRNA as well (Goldberg et al., 1981). Sequencing studies presented in this paper, and by others (Schuler et al., 1982; Doyle et al., 1986; Tierney et al., 1987), demonstrated that these mRNAs encode the α'/α- and β-subunits of the β-conglycinin protein, respectively. Figure 1A shows that six β-conglycinin genes, located in three non-overlapping genomic regions, were obtained from this screen. Hybridization of A-16 with the cloned DNA fragments produced variable hybridization signals, suggesting that additional β-conglycinin genes existed in the soybean genome (data not shown). To test this possibility, the gene fragments shown in Figure 1A (bracketed regions) were hybridized individually, or in combination, with leaf nuclear DNA gel blots. As shown in Figure 1B, the mixed probe hybridized with approximately 25 distinct DNA fragments, most of which were not represented in the genomic clones presented schematically in Figure 1A. We were able to correlate the DNA fragments with specific β-conglycinin genes by determining that they were present approximately once per soybean genome (Figure 1B, compare lane L with lanes R1 and R2), and by showing that they were present in specific genome clones (see below). We designated the β-conglycinin genes as CG-1 through CG-15 on the basis of their order in the DNA gel blot shown in Figure 1B. As shown in Figure 1C, individual β-conglycinin gene probes reacted with a subset of genomic DNA fragments at a reduced criterion, and these subsets were narrowed even further when the DNA gel blots were washed at a stringent criterion. Together, these data suggest that there are at least 15 distinct β-conglycinin genes in the soybean genome, and that these genes have various levels of relatedness to each other.

β-Conglycinin Genes Are Clustered in the Soybean Genome

To describe β-conglycinin gene organization in soybean chromosomes, we screened a λ Charon 4 AluI/HaeIII genome library (Jofuku, 1987; Jofuku and Goldberg, 1989) with the probes shown in Figure 1A. We then ordered all of the phage DNA clones isolated in this screen, and in the previous one (Figure 1A), into linear arrays by using restriction site mapping, gel blot hybridization between DNA fragments, and heteroduplex analysis (Nielsen et al., 1989). DNA regions defined by the overlapping phages were extended by screening the λ Charon 4 AluI/HaeIII library a second time with single-copy restriction fragments from the ends of each domain. The three screens yielded a total of 44 recombinant phages. We mapped the phage DNAs with at least 12 restriction enzymes to distinguish

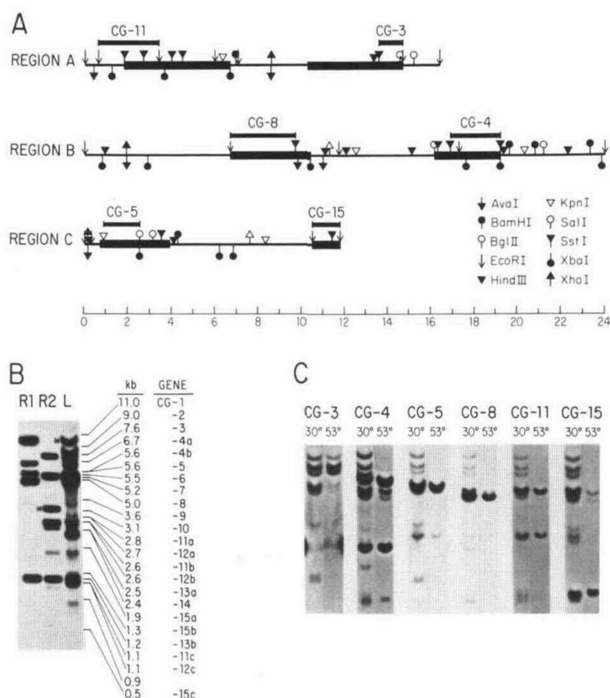

Figure 1. β-Conglycinin Is Encoded by a Large Gene Family.

(A) Restriction endonuclease maps of β-conglycinin genomic clones. Phage DNA restriction endonuclease sites were determined as described in Methods. The phages and insert sizes that defined each genomic region were: Region A, λA16-A32 (14.7 kb) and λA16-A33 (15.8 kb); Region B, λA16-B7 (17.3 kb) and λA16-B3 (17.3 kb); and Region C, λA16-C1 (11.8 kb). Overlapping genomic segments were deduced by aligning restriction sites and by heteroduplex analysis (Nielsen et al., 1989). Gene positions were localized by R-loop analysis (Barker et al., 1988) and by DNA gel blot hybridization studies. Brackets over genes represent fragments used in DNA gel blot experiments.

(B) Representation of β-conglycinin genes in the soybean genome. Leaf nuclear DNA and phage DNA single-copy equivalents were digested with EcoRI, fractionated by electrophoresis on 0.7% agarose gels, transferred to nitrocellulose filters, and hybridized at 30°C with the probes shown in **(A)**. Lane L contained 10 μg of leaf DNA; lanes R1 and R2 contained 450 pg each of λA16-B3 and λA16-C1, and λA16-A33 and λA16-C18 phage DNAs, respectively. Asterisks denote a λA16-C18 EcoRI end fragment that corresponds to β-conglycinin gene CG-1 (see Figure 2). Because the λA16-C18 genomic clone was isolated from the λ Charon 4 AluI/HaeIII library, EcoRI end fragments do not correspond with those present in the soybean genome.

(C) Divergence of β-conglycinin gene family members. Leaf DNA gel blots were hybridized at 30°C with each of the β-conglycinin gene probes shown in **(A)**. After hybridization, the gel blots were washed at both 30°C and 53°C. Following each wash the gel blots were exposed to x-ray film.

between linked genomic regions and highly conserved homoeologous regions that were unlinked in the genome (Nielsen et al., 1989). DNA fragments contained within the recombinant phages were correlated with those observed in leaf DNA blots to ensure that no rearrangements occurred during cloning (Figures 1B and C).

Figure 2 summarizes the results of our chromosome walking experiments. Ten of 13 cloned β-conglycinin genes were clustered in three genomic regions. Five genes were present in divergent transcriptional orientations within one 44-kb domain designated as Region A. Genes within Region A were separated by 1.5 kb to 4 kb. By contrast, Regions B and C contained two and three β-conglycinin genes each, respectively, and genes in these regions were present in the same transcriptional orientation. Heteroduplex analysis showed that DNA sequences contiguous to the CG-8 and CG-4 β-conglycinin genes in Region B did not hybridize extensively with those flanking the CG-5 and CG-15 β-conglycinin genes in Region C (data not shown). Regions D, E, and F contained only one β-conglycinin gene each.

We determined that no other highly expressed seed protein genes were present in the β-conglycinin gene regions by hybridizing relevant phage DNA fragments with embryo mRNA gel blots and with labeled midmaturation stage cDNA (data not shown). By contrast, we identified a DNA sequence that was represented at low levels in leaf mRNA (NSP gene, Region B), indicating that some β-conglycinin genes were linked with genes expressed in the mature plant. Together, these experiments show that β-conglycinin genes are present in at least six genomic regions, and that some regions contain several closely linked β-conglycinin genes.

2.5-kb and 1.7-kb β-Conglycinin mRNAs Are Encoded by Distinct Genes

We hybridized DNA fragments containing each β-conglycinin gene with embryo mRNA gel blots to determine whether the β-conglycinin genes were complementary to one or both mRNA size classes. Figure 2 shows that, at a 42°C criterion, each fragment reacted with both the 2.5-kb and 1.7-kb mRNAs, irrespective of where the fragment was located within the gene (e.g., CG-11 fragments a, b, and c). The CG-1, CG-2, and CG-3 β-conglycinin genes hybridized preferentially with the 2.5-kb mRNA, whereas the CG-4, CG-8, CG-11, CG-12, CG-13, and CG-15 genes reacted more intensely with the 1.7-kb mRNA (Figure 2). Increasing the gel blot wash temperature to 60°C resulted in a signal specific for one mRNA size class (Figure 2), indicating that the hybridization signals reflected sequence homologies rather than differences in 2.5-kb and 1.7-kb mRNA prevalences. By contrast, the CG-5, CG-6, CG-7, and CG-14 probes hybridized equally to both mRNA size classes (Figure 2). Together, these findings indicate that

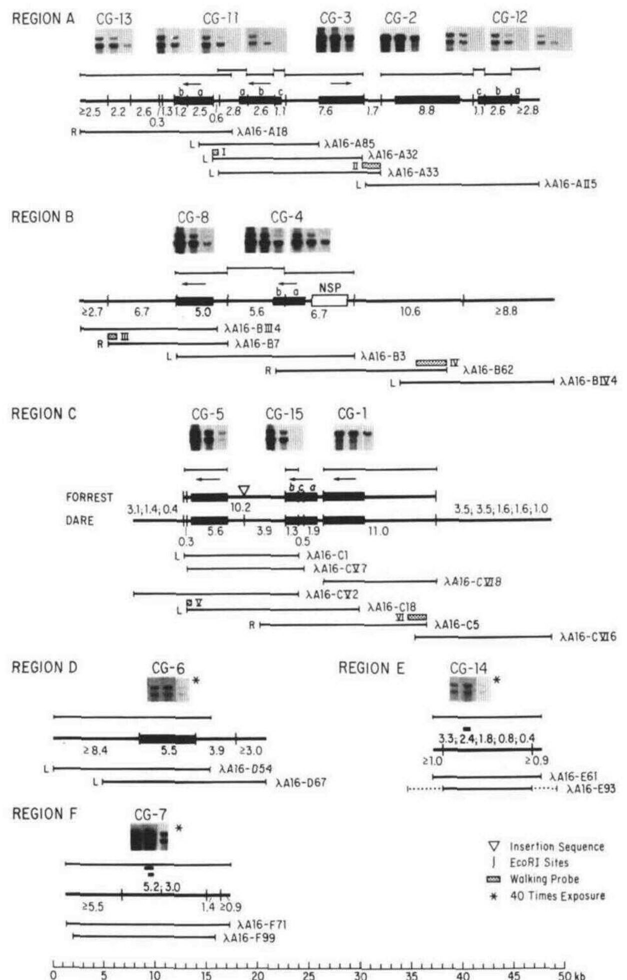

Figure 2. Organization of the β-Conglycinin Gene Family.
Restriction maps showing EcoRI sites (vertical lines) and gene positions (boxed areas) were determined as described in Methods. Transcriptional orientations were established by DNA gel blot hybridization with random oligonucleotide-primed and oligo-dT-primed cDNAs, and by heteroduplex analysis (Nielsen et al., 1989). Numbers indicate EcoRI fragment sizes (kb). In Region C, the triangle shows the position of a small insertion/deletion polymorphism present in Forrest variety DNA that eliminated an EcoRI restriction site present in Dare variety DNA. Bars below restriction maps show representative phage inserts used to construct the maps. L and R indicate the left and right λ phage arms, respectively. Stippled boxes labeled with Roman numerals represent single-copy restriction fragments that were recloned and used for the chromosome walking experiments. Brackets above the restriction maps show cloned fragments that were used as probes in RNA gel blot experiments. For these studies, midmaturation stage embryo mRNA was fractionated on 1% agarose gels containing 10% mM methylmercury hydroxide, transferred to APT paper, hybridized with each probe at 42°C, and then exposed to x-ray film. The RNA gel blots were then washed at 50°C and exposed a second time to autoradiographic film. This procedure was repeated again at 60°C. RNA gel blot autoradiograms were exposed for 1.5 hr, except for those marked by asterisks. These RNA gel blots were exposed for 40 hr.

each β-conglycinin gene contains sequences complementary to both the 2.5-kb and 1.7-kb mRNAs, that these homologies are present in 5′ and 3′ gene regions, and that three subfamilies of β-conglycinin genes exist: 2.5-kb α'/α-subunit mRNA genes; 1.7-kb β-subunit mRNA genes, and genes equally homologous to both mRNA size classes.

The CG-4 β-Conglycinin Gene Encodes a Functional 1.7-kb mRNA

We showed previously using R-loop analysis that the CG-4 β-conglycinin gene (Region B, Figure 2) contains at least five introns (Barker, Harada, and Goldberg, 1988). Figure 2 shows that this gene preferentially hybridized with the 1.7-kb β-conglycinin mRNA, suggesting that it encodes the β-subunit polypeptide. We sequenced the CG-4 β-conglycinin gene to precisely characterize β-conglycinin gene structure. Figure 3 shows that the CG-4 β-conglycinin gene contained six exons and five introns that ranged from 85 to 268 base pairs in length. Comparison of amino acid sequences obtained from the DNA sequence analysis with those obtained from N-terminal and internal peptides of the purified β-conglycinin β-subunit (Coates et al., 1985; Hirano et al., 1987) indicated that the CG-4 gene encoded a β-subunit as predicted from the mRNA gel blot studies (Figure 2). Transformation studies with the CG-4 β-conglycinin gene by us (Barker et al., 1988), and by others (Bray et al., 1987), directly demonstrated that CG-4 is a functional gene that directs the synthesis of the 1.7-kb β-subunit mRNA in transgenic tobacco and petunia plants.

β-Conglycinin Gene Classes Differ by a Unique Insertion Sequence

To analyze differences between β-conglycinin genes encoding the 2.5-kb and 1.7-kb mRNAs, we compared the CG-4 DNA sequence with that of the CG-1 gene present in Region C (Figure 2). The CG-1 β-conglycinin gene was sequenced by Doyle et al. (1986), and was shown to specify the 2.5-kb α'-subunit β-conglycinin mRNA. Figure 4 shows that the structures of these genes were similar, and that each gene contained six exons and five introns. Table 1 indicates that the exons averaged 85% in sequence homology, whereas the introns were only 50% homologous on the average. Exon homologies translated to approximately 75% identity at the amino acid sequence level. With the exception of several shared consensus sequences, and two blocks of highly homologous sequences (Figure 4, stippled areas), the 5′ regions of the CG-1 and CG-4 genes did not have a high degree of sequence similarity.

Figure 4 shows that the CG-1 and CG-4 β-conglycinin genes differed primarily by the presence of a 0.56-kb protein coding element in the first exon of the CG-1 gene. This element is represented in 2.5-kb β-conglycinin mRNAs and α'/α-subunit polypeptides, and is responsible for the size differences between 2.5-kb and 1.7-kb β-conglycinin mRNAs and α'/α-subunit and β-subunit polypeptides. We analyzed the DNA sequences present in the 0.56-kb CG-1 insertion element and observed a mosaic array of repeated sequences (Figure 4, arrows). The 5′ portions of each repeat block within the element (Figure 4, hatched areas below the arrows) were highly homologous with each other, and with a 27-base pair sequence in exon one of the CG-4 β-conglycinin gene (Figure 3, arrow underline). The remainder of each repeat block consisted of short

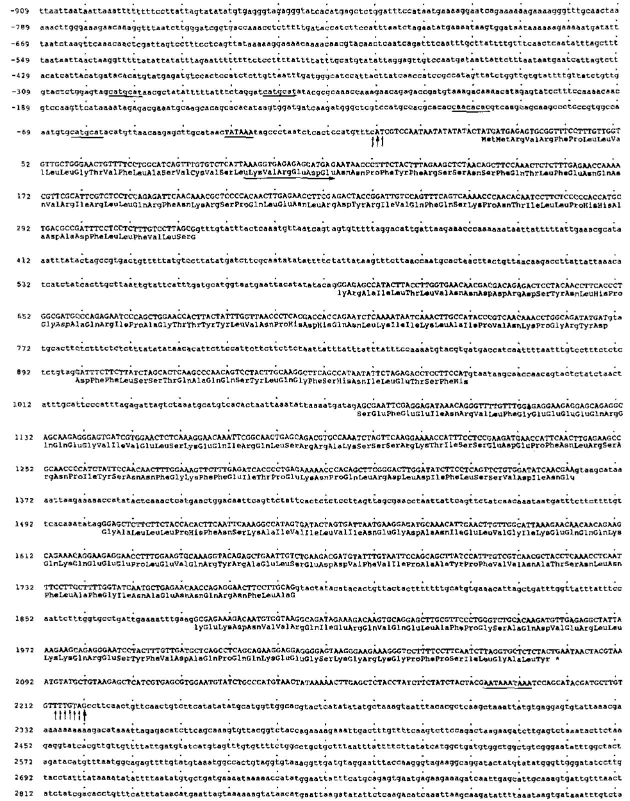

Figure 3. DNA Sequence of the CG-4 β-Conglycinin Gene.

Capital letters represent sequences present in the 1.7-kb β-conglycinin mRNA, except for the putative TATA box, which is also underlined. Lower case letters represent untranscribed sequences and intervening sequences. The assumed polyadenylation consensus sequence is also underlined. The 5′ and 3′ gene ends were mapped by S1 nuclease, and are designated by arrows. The strongest transcription start signal is indicated by the largest arrow. Short 5′ sequences that are conserved between different seed protein genes are underlined (Goldberg, 1986; Dickinson et al., 1988; Goldberg et al., 1989; Nielsen et al., 1989). A sequence that is repeated in the CG-1 gene 0.56-kb insertion region (see Figure 4) is underlined with an arrow.

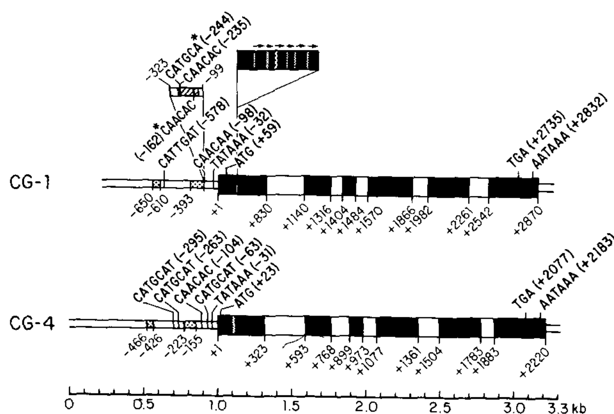

Figure 4. Sequence Homology between the CG-4 and CG-1 β-Conglycinin Genes.

The sequence of the CG-4 1.7-kb β-subunit mRNA gene was compared with that of the CG-1 2.5-kb α′-subunit mRNA gene (Doyle et al., 1986). Black boxes represent exon sequences and white boxes represent introns. Consensus sequences are numbered relative to the transcription start site. The DNA segment above the CG-1 gene corresponds to the insertion element that is present in exon 1 of the CG-1 gene but is absent from the CG-4 gene. Arrows above the CG-1 insertion element indicate blocks of repetitive sequences that have homologies with each other. The grey areas at the 5′ end of each repeat block contain sequences complementary to a 27-base pair sequence in exon 1 of the CG-4 gene (Figure 3, arrow underline). Stippled areas in the CG-1 and CG-4 5′ regions indicate regions that are longer than 40 base pairs in length, and are greater than 80% similar. The raised portion of the CG-1 5′ region contains sequences necessary for elevated expression of the CG-1 gene during seed development (Chen et al., 1986, 1988). Diagonal lines show the location of the quantitative element (Chen et al., 1988). The CG-4 5′ region does not contain contiguous sequences corresponding to those in the raised CG-1 5′ region. Consensus sequences marked with asterisks are present in the glycinin Gy_7 gene quantitative region (Nielsen et al., 1989).

sequence elements that were represented in varying combinations in contiguous blocks. These repeated sequences manifested themselves as groups of similar amino acids in the α′-polypeptide (Doyle et al., 1986). Together, these comparisons indicate that β-conglycinin genes encoding the 2.5-kb α′/α-subunit mRNA and the 1.7-kb β-subunit mRNA are homologous along their entire lengths and differ by the presence or absence of an insertion element.

Differential Accumulation of β-Conglycinin mRNAs Is Regulated at the Posttranscriptional Level

We hybridized cDNA clones representing 2.5-kb and 1.7-kb β-conglycinin mRNAs with ^{32}P-labeled nuclear RNA and polysomal mRNA from different embryonic stages to determine whether β-conglycinin gene expression was regulated at the transcriptional or posttranscriptional levels. Figure 5A shows that the A-16 cDNA clone hybridized more intensely with the 2.5-kb α′/α-subunit mRNA class at a 42°C criterion. The specificity of A-16 for the 2.5-kb mRNA class was amplified at a higher stringency (compare Figure 5A, 42°C and 60°C lanes). By contrast, the pCG-24 cDNA clone produced a stronger hybridization signal with the 1.7-kb β-subunit mRNA class at both the 42°C and 60°C hybridization criteria. Densitometric analysis of the RNA gel blot autoradiograms indicated that there was approximately 25% cross-hybridization between the cDNA clones and their heterologous mRNAs at 42°C.

Figure 5B shows that the β-conglycinin mRNAs accumulated differentially during embryogenesis. The 2.5-kb α′/α-subunit mRNA class accumulated and decayed earlier in development than the 1.7-kb β-subunit mRNA class. Figure 5B also confirms that the A-16 and pCG-24 cDNA probes were specific for their homologous mRNAs at the 42°C hybridization temperature. For example, the pCG-24 probe produced strong signals with mRNAs from late embryonic stages, whereas the A-16 cDNA probe produced only a weak signal with the same mRNAs (Figure 5B, 91 and 93 days after flowering).

To measure the relative transcriptional activity of β-conglycinin genes during embryogenesis, we isolated nuclei from embryos at different developmental time periods, and then hybridized ^{32}P-labeled nuclear RNAs synthesized in vitro with the A-16 and pCG-24 cDNA clones (Walling et al., 1986; Cox and Goldberg, 1988). Figure 5C shows that the relative transcription rates of the 2.5-kb α′/α-subunit and 1.7-kb β-subunit mRNA genes increased and decreased during embryogenesis, indicating that both of these gene classes were activated and repressed during embryogenesis. In contrast with the mRNA accumulation results, both the timing of transcriptional activation and repression events, and the relative transcriptional levels of the 2.5-kb and 1.7-kb gene classes were similar to each other at each stage of development (compare Figure 5, B and C). For example, the relative transcription rate of the 2.5-kb α′/α-subunit gene class was not significantly different from that of the 1.7-kb β-subunit gene class late in development, even though the 2.5-kb α′/α-subunit mRNA was barely detectable (Figure 5, B and C, 91 days after flowering). Together, these results show that β-conglycinin genes are transcriptionally regulated and that posttranscriptional events (e.g., processing of primary transcripts, selective nuclear export, mRNA stability) play a significant role in the differential accumulation of 2.5-kb and 1.7-kb β-conglycinin mRNAs during embryogenesis.

DISCUSSION

β-Conglycinin Is Encoded by a Complex Gene Family

β-Conglycinin is a glycosylated storage protein that represents approximately 25% of the protein present in a

Table 1. Average Percent Similarity between CG-1 and CG-4 β-Conglycinin Genes

	Exon 1[a]											Exon 6[b]		
	5′	3′	Intron 1	Exon 2	Intron 2	Exon 3	Intron 3	Exon 4	Intron 4	Exon 5	Intron 5	S	A	E
% Similarity[c]	82	75	49	84	33	85	51	88	66	91	61	84	48	87

[a] 5′ and 3′ refer to regions before and after the 0.56-kb CG-1 insertion element.
[b] S, A, and E refer to regions before the stop codon, from the stop codon to the poly(A) consensus sequence, and from the poly(A) consensus sequence to the 3′ end of gene, respectively.
[c] CG-1 DNA sequence data taken from Doyle et al. (1986).

soybean seed (Hill and Breidenbach, 1974; Beachy et al., 1981; Coates et al., 1985). β-Conglycinin consists of three major subunits, α′, α, and β, that have apparent molecular weights of 76 kD, 72 kD, and 52 to 54 kD, respectively (Coates et al., 1985). Minor subunits, designated as β′, γ, and δ, have also been reported (Thanh and Shibasaki, 1976; Coates et al., 1985). The α′-, α-, and β-subunits associate in various combinations to form trimeric proteins that are present in embryo axis and cotyledon protein bodies (Thanh and Shibasaki, 1976, 1978; Sykes and Gayler, 1981; Ladin et al., 1987; Sugimoto et al., 1987). Two-dimensional gel electrophoresis studies showed that there is a larger number of unique β-subunits than α′/α-subunit counterparts (Lei et al., 1983; Ladin et al., 1987; Sugimoto et al., 1987). The α′/α-subunits accumulate earlier in seed development than the β-subunits (Gayler and Sykes, 1981; Meinke et al., 1981; Sengupta et al., 1981; Ladin et al., 1987; Tierney et al., 1987), and are more prevalent than the β-subunits in the embryonic axis (Ladin et al., 1987; Sugimoto et al., 1987).

DNA gel blot studies (Figure 1) and characterization of isolated genomic clones (Figure 2) indicate that there are at least 15 distinct β-conglycinin genes in the soybean genome. This estimate is probably a minimum one because functionally related DNA sequences that are distant relatives would not have reacted with our probes, even at the reduced stringency (30°C) used for our DNA gel blots (Nielsen et al., 1989). mRNA gel blot experiments with the β-conglycinin genes identified in our chromosome walking studies (Figure 2) indicate that there are at least three α′/α-subunit genes (CG-1, CG-2, and CG-3), six β-subunit genes (CG-4, CG-8, CG-11, CG-12, CG-13, and CG-15), and four genes that are equally homologous with the 2.5-kb α′/α-subunit and 1.7-kb β-subunit mRNAs (CG-5, CG-6, CG-7, and CG-14). mRNA gel blots with the CG-6, CG-7, and CG-14 β-conglycinin gene probes required long exposure times to produce visible signals (Figure 2). This result indicates that these genes are either complementary to low prevalence β-conglycinin mRNAs, or are related distinctly to β-conglycinin mRNAs present at midmaturation.

With the exception of the CG-1 and CG-4 β-conglycinin genes (Figures 2 and 4), we do not know how many of the β-conglycinin genes identified in our studies specify functional β-conglycinin polypeptides. Our estimate of the number of α′/α-subunit and β-subunit genes agrees well with those obtained from two-dimensional protein gel electrophoresis studies (Lei et al., 1983; Ladin et al., 1987; Sugimoto et al., 1987). This suggests that the CG-2, CG-3, CG-8, CG-11, CG-12, and CG-15 β-conglycinin genes may also be functional. Regardless of the number of functional genes, it is clear that β-conglycinin is encoded by a complex gene family, and that subfamilies of genes with different extents of relatedness occur (Figure 1C). In comparison with other soybean seed protein gene families, the β-conglycinin family appears to be larger than those encoding glycinin (Nielsen et al., 1989) and seed lectin (Goldberg, Hoschek, and Vodkin, 1983; Okamuro, Jofuku, and Goldberg, 1986), but is equally complex as that specifying the Kunitz trypsin inhibitor (Jofuku and Goldberg, 1989).

2.5-kb α′/α-Subunit and 1.7-kb β-subunit mRNAs Are Specified by Distinct Genes

Each β-conglycinin gene identified in our chromosome walking studies (Figure 2) contains sequences complementary to both the 2.5-kb and 1.7-kb β-conglycinin mRNAs. These sequences are not confined to one region but are scattered throughout the β-conglycinin genes, and appear at both the 5′ and 3′ gene ends. Sequence analysis of genes specifying the 2.5-kb and 1.7-kb β-conglycinin mRNAs confirms this observation, and indicates that the primary difference between the two gene classes is the presence of a 0.56-kb element in exon 1 of the 2.5-kb α′/α-subunit mRNA gene class (Figures 3 and 4; Doyle et al., 1986; Tierney et al., 1987). Thus, the two size classes of β-conglycinin mRNAs do not arise from the differential processing of a common precursor transcript (Breitbart, Andreatis, and Nadal-Ginard, 1987), but result from the transcription of distinct genes.

The 556-base pair insertion sequence in the CG-1 α′-subunit β-conglycinin gene is not present in related vicilin-like storage protein genes of the common bean (Doyle et

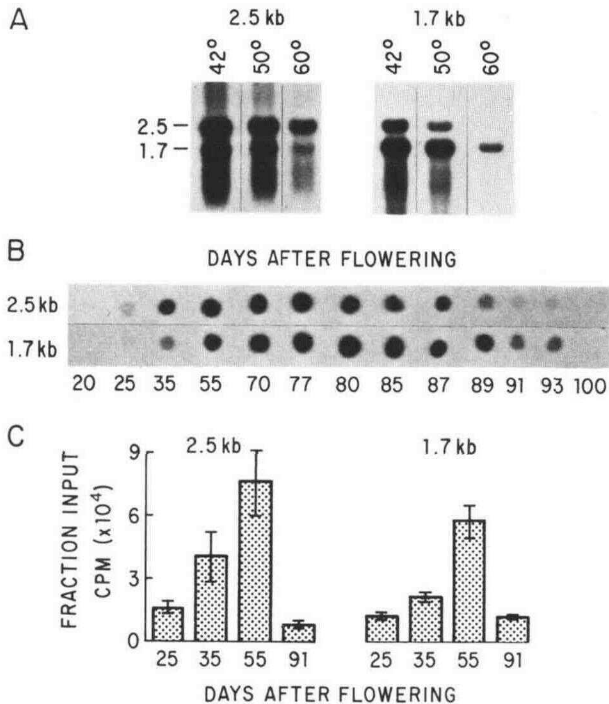

Figure 5. Transcriptional and Posttranscriptional Regulation of β-Conglycinin Gene Expression.

(A) Specificity of β-conglycinin gene probes. A-16 and pCG-24 cDNA clones were hybridized with midmaturation stage embryo mRNA gel blots at 42°C, 50°C, and 60°C. These cDNA clones were isolated from a library of midmaturation stage embryo mRNA (Goldberg et al., 1981). The gel blots were washed at the same hybridization criterion and then exposed to x-ray film for the same time periods.

(B) Differential accumulation of 2.5-kb α'/α-subunit and 1.7-kb β-subunit β-conglycinin mRNAs. One hundred nanograms of polysomal poly(A) mRNA from embryos at the indicated stages were spotted onto nitrocellulose filters, hybridized individually with a sequence excess of A-16 and pCG-24 DNA at 42°C, and then washed at the same temperature.

(C) Transcription of β-conglycinin genes during embryogenesis. Nuclei were isolated from embryos at 25, 35, 55, and 91 days after flowering and ^{32}P-labeled nuclear RNAs were synthesized in vitro according to Walling et al. (1986) and Cox and Goldberg (1988). The relative transcription rates were quantitated by hybridizing the ^{32}P-labeled nuclear RNAs with a filter-bound sequence excess of linearized A-16 and pCG-24 DNAs at 42°C. Hybridization to pBR322 DNA was used as a control and was subtracted as background. Fraction input cpm was calculated according to Walling et al. (1986) and Nielsen et al. (1989). Values represent the average of four determinations. Ranges of values are indicated by the brackets.

al., 1986) or pea (Lycett et al., 1983; Higgins et al., 1988). Nor is it present in a cotton vicilin-like storage protein gene, although an unrelated insertion is present in the 5' coding region of this gene as well (Borroto and Dure, 1987). Comparisons of the CG-4 β-subunit β-conglycinin gene with the vicilin-like 7S storage protein genes of pea and the common bean indicate that they are highly homologous. Each gene has six exons and five introns, the introns are located at similar positions relative to the translated protein sequence, the coding sequences show a high degree of homology, and each gene specifies a 1.7-kb mRNA (Doyle et al., 1986; Higgins et al., 1988). Two-dimensional protein gel electrophoresis studies (Ladin et al., 1987; Sugimoto et al., 1987), genetic analysis of β-conglycinin subunit variants (Davies, Coates, and Nielsen, 1985), DNA gel blot experiments (Figure 1; Tierney et al., 1987), and characterization of β-conglycinin genomic clones (Figure 2) indicate that there are more 1.7-kb β-subunit mRNA genes than 2.5-kb α'/α-subunit mRNA genes. *The simplest interpretation of these data is that the* β-subunit genes are the progenitor vicilin genes in soybean, and that the events leading to the 0.56-kb insertion in the α'/α-subunit genes occurred after the divergence of soybean from its legume relatives.

The complex structure of the 0.56-kb α'/α-subunit gene insertion element (Figure 4; Doyle et al., 1986) provides some clues as to how it may have been inserted into the CG-1 β-conglycinin gene coding sequence. Figure 3 shows that there is a 27-base pair sequence in exon 1 of the CG-4 β-subunit gene (nucleotides +94 to +112, underlined with an arrow) that is repeated, in part, several times within the CG-1 gene insertion element (Figure 4, hatched areas). The CG-4 gene region containing this sequence probably represents the site of element insertion within the progenitor 2.5-kb α'/α-subunit mRNA gene. Because the insertion element contains a mosaic of repeated sequences (Figure 4; Doyle et al., 1986), the simplest hypothesis is that the initial insertion was accompanied by a series of duplications 3' to the insert, giving rise to a coding region that specifies the N terminus of the mature α'/α-polypeptide. The insertion element does not have inverted and/or direct repeats at its presumed junction in the CG-1 α'/α-subunit gene. If these sequences were present initially, they have diverged during the course of time. The insertion of a protein coding element into a eukaryotic gene is unusual but not unprecedented. Analogous to the situation with the β-conglycinin gene family, the *Drosophila* 68C glue protein gene contains a 0.7-kb module in a protein coding region that is not present in two other glue protein gene family members (Garfinkel, Pruitt, and Meyerowitz, 1983).

β-Conglycinin Genes Are Clustered in Several Genomic Regions

The chromosome walking experiments summarized in Figure 2 spanned 175 kb of the soybean genome. β-Conglycinin genes have been localized to six DNA regions, although most β-conglycinin genes are clustered in three

domains (Regions A, B, and C). Within a given region, 2.5-kb α'/α-subunit mRNA genes and 1.7-kb β-subunit mRNA genes are both tandemly organized (e.g., Region A, CG-13 and CG-11) and interspersed among each other (e.g., Region C, CG-15 and CG-1). In addition, β-conglycinin genes are contiguous to each other in both divergent transcriptional orientations (e.g., Region A, CG-11 and CG-3) and similar transcriptional orientations (e.g., Region B, CG-4 and CG-8, and Region A, CG-11 and CG-13). These observations, and the close linkage of the CG-4 β-conglycinin gene with a differentially regulated non-seed protein gene in Region B, suggest that each gene within a cluster is programmed by its own developmental control element. Developmental-specific expression of both the CG-1 β-conglycinin gene (Beachy et al., 1985; Chen, Schuler, and Beachy, 1986; Chen, Pan, and Beachy, 1988) and the CG-4 β-conglycinin gene (Bray et al., 1987; Barker et al., 1988) in transformed tobacco plants indicates that this is indeed the case.

At the present time we do not know how the β-conglycinin gene regions are organized with respect to each other in the soybean genome. Genetic analysis of β-conglycinin polypeptide variants indicates that α'- and α-subunits are encoded by genes that assort independently of each other (Davies et al., 1985). By contrast, linkage studies indicate that a gene encoding an α-subunit is linked to one encoding a β-subunit (Davies et al., 1985). These genetic studies are in close agreement with our molecular analysis of β-conglycinin gene organization, and suggest that Region A containing the five clustered α'/α-subunit and β-subunit genes is not linked with Region C that contains the CG-1 α'-subunit gene (Figure 2). Recently, we showed that glycinin storage protein genes are located on two homologous DNA regions that probably represent relics of an ancient tetraploidization event (Hymowitz and Singh, 1987; Nielson et al., 1989). Together, these studies strongly suggest that the β-conglycinin gene clusters shown in Figure 2 are probably present on at least two soybean chromosomes.

β-Conglycinin Genes Are Regulated at the Transcriptional Level

Transcription studies presented here (Figure 5C) and elsewhere (Goldberg et al., 1981; Walling et al., 1986) show that β-conglycinin genes are regulated in part by transcriptional processes. β-Conglycinin genes are activated transcriptionally early in embryogenesis, attain a relatively high transcription rate by midmaturation, and are repressed transcriptionally prior to seed dormancy (Figure 5C; Walling et al., 1986). β-Conglycinin gene transcription is not detectable in cotyledons of the germinating seedling (Walling et al., 1986) or in organ systems of the mature plant (Goldberg et al., 1981; Walling et al., 1986). These observations indicate that the accumulation and decay of β-conglycinin mRNAs during embryogenesis (Figure 5B) are due in part to transcriptional events, and that transcriptional processes controlling β-conglycinin genes are similar to those operating on seed protein genes in general (Goldberg, Barker, and Perez-Grau, 1989).

The DNA sequence comparisons presented in Figure 4 and deletion analysis of the CG-1 β-conglycinin gene in transformed plants (Chen et al., 1986, 1988) provide clues as to the control sequences that might program β-conglycinin gene transcription during development. Chen et al. (1986, 1988) showed that only 159 base pairs contiguous to the transcription start site of the CG-1 β-conglycinin gene are required for expression during seed development. A specific 5′ region (nucleotides −159 to −257) flanking the CG-1 gene is required, however, for high expression levels (Figure 4, hatched box within −323 to −99 region). Sequence comparisons indicate that the CG-1 quantitative control region is absent from the CG-4 5′ flanking region (Figure 4). By contrast, short sequences present in the CG-1 quantitative control region are represented in the 5′ region of the CG-4 β-conglycinin gene (Figure 4). These sequences include the 5′-CATGCAT-3′ and 5′-CAACACA-3′ consensus elements found in the 5′ regions of several seed protein genes investigated to date (Goldberg, 1986; Dickinson, Evans, and Nielsen, 1988; Goldberg et al, 1989; Nielsen et al., 1989). Recently, deletion analysis of the Gy_1 glycinin gene also revealed the presence of a 5′ quantitative region (nucleotides −464 to −66) necessary for elevated expression levels during seed development (R. B. Goldberg, J. Truettner, and T. Sims, unpublished results; Goldberg et al., 1989). The nucleotide sequences of the CG-1 β-conglycinin and Gy_1 glycinin quantitative control regions are different. They do share, however, the 5′-CATGCAT-3′ and 5′-CAACACA-3′ consensus sequences (Figure 4). DNA binding studies showed recently that the 5′-CAACACA-3′ sequence within the Gy_1 glycinin quantitative control region interacts with an embryo DNA-binding protein (L. Perez-Grau and R. B. Goldberg, unpublished results). These studies suggest that short consensus sequences shared by β-conglycinin and glycinin genes may be important for regulating seed protein gene transcription levels during development.

Posttranscriptional Events Are Important for the Temporal Accumulation of β-Conglycinin mRNAs

The most surprising aspect of our results is the importance of posttranscriptional processes in the selective accumulation of 2.5-kb and 1.7-kb β-conglycinin mRNAs during embryogenesis. The experiments shown in Figure 5B, as well as those of Naito et al. (1988), indicate that the 2.5-kb α'/α-subunit and 1.7-kb β-subunit mRNAs accumulate differentially during seed development. The 2.5-kb mRNA class accumulates earlier than the 1.7-kb mRNA class, and this is paralleled by an earlier accumulation of α'/α-subunit polypeptides (Gayler and Sykes, 1981; Meinke et al., 1981; Ladin et al., 1987; Tierney et al., 1987). By

contrast, genes encoding the 2.5-kb and 1.7-kb β-conglycinin mRNA classes are transcribed at approximately the same relative rates, and are activated and repressed at the same times during embryogenesis (Figure 5C). Thus, the differential accumulation of 2.5-kb and 1.7-kb β-conglycinin mRNAs appears to be due to events that occur after transcription has taken place.

Posttranscriptional processes have been shown to be important for regulating the expression of other seed protein genes (Goldberg et al., 1989), as well as plant genes in general (Okamuro and Goldberg, 1989). Previously, posttranscriptional processes have been demonstrated to play a role in establishing seed protein mRNA prevalence levels rather than controlling the temporal programs of mRNA accumulation during development (Walling et al., 1986; Goldberg et al., 1989). For example, the Gy_1, Gy_2, Gy_4, and Gy_5 glycinin mRNAs accumulate and decay coordinately during soybean embryogenesis, and their corresponding genes are activated and repressed transcriptionally at the same developmental time periods (Nielsen et al., 1989). In contrast with the observations with seed protein genes, posttranscriptional selection events have been shown to control the establishment of developmental-specific mRNA populations in some tobacco organ systems (Kamalay and Goldberg, 1980, 1984). For example, stem mRNA contains thousands of diverse sequences that are not detectable on leaf polysomes but are represented in leaf nuclear RNA (Kamalay and Goldberg, 1980, 1984). These selection events appear analogous to those observed here for β-conglycinin mRNAs.

Posttranscriptional control sequences, by definition, must be located within the gene transcript. Tobacco and petunia plants transformed with both the CG-1 α'/α-subunit and CG-4 β-subunit β-conglycinin genes show the same differential accumulation of the 2.5-kb and 1.7-kb mRNAs during seed development as observed in soybean (S. J. Barker and R. B. Goldberg, unpublished results; Naito et al., 1988). These findings suggest that mechanisms responsible for differential β-conglycinin mRNA accumulation are highly conserved between legume and solanaceous plants. Comparisons of the CG-1 and CG-4 β-conglycinin gene sequences (Figure 4 and Table 1) show many nucleotide differences in exon and intron regions. One obvious candidate for a region involved in the differential accumulation of β-conglycinin mRNAs is the CG-1 gene 0.56-kb insertion element (Figure 4). Clearly, the precise mechanisms that control seed protein gene expression at the transcriptional and posttranscriptional levels remain to be determined.

METHODS

Isolation and Labeling of Nucleic Acids

Soybean, plasmid, and phage DNAs were isolated and labeled as previously described (Fischer and Goldberg, 1982; Cox and Goldberg, 1988; Jofuku and Goldberg, 1988). Polysomal poly(A^+) mRNAs were isolated by procedures described elsewhere (Cox and Goldberg, 1988). Labeled nuclear RNAs were synthesized in isolated nuclei as described previously (Walling et al., 1986; Cox and Goldberg, 1988).

Isolation of β-Conglycinin Genomic Clones

Phage clones were obtained from a λ Charon 4 EcoRI library of nuclear DNA from the Forrest variety (Fischer and Goldberg, 1982) and a λ Charon 4 HaeIII-AluI DNA library from the Dare variety (Jofuku, 1987; Jofuku and Goldberg, 1989). Restriction fragments for chromosome walking experiments (Figure 2, stippled bars) were recloned and hybridized with soybean leaf DNA to check for repeated sequences. In only two cases was it necessary to increase the hybridization criterion to eliminate hybridization of probes with related sequences. In these cases, plaque hybridization studies were performed at the elevated criterion.

Restriction Endonuclease Mapping

Restriction site positions were deduced from the results of single and multiple digestions, from digestions with Bal-31 nuclease, and from partial digestion procedures (Fischer and Goldberg, 1982; Jofuku and Goldberg, 1988).

Gel Blot Hybridization Experiments

DNA fragments were blotted onto nitrocellulose filters and hybridized with labeled probes as described by Wahl, Stern, and Stark (1979). APT paper RNA gel blots were prepared and hybridized with ^{32}P-labeled DNA as outlined by Alwine et al. (1979) and by Seed (1982). Hybrids were melted by incubating the blots in hybridization buffer for 4 hr at the indicated temperature.

DNA Sequence Analysis and S1 Nuclease Mapping

Dideoxy sequencing of Bal-31-deleted M13 clones was performed as described by Nielsen et al. (1989). Sequence analysis was carried out on a VAX 11/780 computer using the programs of Staden (1982) and Kanehisa (1982). Ambiguous exon-intron borders were assigned on the basis of homology with the CG-1 β-conglycinin gene (Doyle et al., 1986). 5′ and 3′ transcription termini were determined by S1 nuclease mapping using the procedures of Rosbash et al. (1979) and Berk and Sharp (1977).

ACKNOWLEDGMENTS

We thank Dr. Linda Walling for providing the staged embryo nuclei. This research was supported by a United States Department of Agriculture Grant to R.B.G. S.J.B. was supported by UCLA and McKnight Foundation predoctoral fellowships.

Received February 22, 1989.

REFERENCES

Alwine, J. C., Kemp, D. J., Parker, B. A., Reisner, J., Stark, G. R., and Wahl, G. M. (1979). Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. Methods Enzymol. **68**, 220–242.

Barker, S. J., Harada, J. J., and Goldberg, R. B. (1988). Cellular localization of soybean storage protein mRNA in transformed tobacco seeds. Proc. Natl. Acad. Sci. USA **85**, 458–462.

Beachy, R. N., Jarvis, N. P., and Barton, K. A. (1981). Biosynthesis of subunits of the soybean 7S storage protein. J. Mol. Appl. Genet. **1**, 19–27.

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

Berk, A. J., and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoreses of S1 endonuclease-digested hybrids. Cell **12**, 721–732.

Borroto, K., and Dure, L. (1987). The globulin storage protein genes of flowering plants are derived from two ancestral genes. Plant Mol. Biol. **8**, 113–131.

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

Breitbart, R. E., Andreatis, A., and Nadal-Ginard, B. (1987). Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. **56**, 467–496.

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 of a constitutive promoter. EMBO J. **7**, 297–302.

Coates, J. B., Medeiros, J. S., Thanh, V. H., and Nielsen, N. C. (1985). Characterization of the subunits of β-conglycinin. Arch. Biochem. Biophys. **243**, 184–194.

Cox, K. H., and Goldberg, R. B. (1988). Analysis of gene expression. In Plant Molecular Biology: A Practical Approach, C. H. Shaw, ed (Oxford: IRL Press), pp. 1–35.

Davies, C. S., Coates, J. B., and Nielsen, N. C. (1985). Inheritance and biochemical analysis of four electrophoretic variants from soybean. Theor. Appl. Genet. **71**, 351–358.

Dickinson, C. D., Evans, R. P., and Nielsen, N. C. (1988). RY-repeats are conserved in the 5′-flanking regions of legume seed-protein genes. Nucl. Acids Res. **16**, 371.

Doyle, J. J., Schuler, M. A., Godette, W. D., Zenger, V., Beachy, R. N., and Slightom, J. L. (1986). The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*: Structural homologies of genes and proteins. J. Biol. Chem. **261**, 9228–9238.

Fischer, R. L., and Goldberg, R. B. (1982). Structure and flanking regions of soybean seed protein genes. Cell **29**, 651–660.

Garfinkel, M. D., Pruitt, R. E., and Meyerowitz, E. M. (1983). DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168**, 765–789.

Gayler, K. R., and Sykes, G. E. (1981). β-Conglycinin in developing soybean seeds. Plant Physiol. **67**, 958–961.

Gayler, K. R., and Sykes, G. E. (1985). Effects of nutritional stress on the storage proteins of soybeans. Plant Physiol. **78**, 582–585.

Goldberg, R. B. (1986). Regulation of plant gene expression. Philos. Trans. R. Soc. Lond. B Biol. Sci. **314**, 343–353.

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

Goldberg, R. B., Crouch, M. L., and Walling, L. (1983). Regulation of soybean seed protein gene expression. In: Manipulation and Expression of Genes in Eukaryotes. P. Nagley, A. W. Linnane, W. J. Peacock, and J. A. Pateman, eds (Sydney: Academic Press) pp. 193–201.

Goldberg, R. B., Hoschek, G., and Vodkin, L. O. (1983). An insertion sequence blocks the expression of a soybean lectin gene. Cell **33**, 465–475.

Goldberg, R. B., Hoschek, G., Ditta, G. S., and Breidenbach, R. W. (1981). Developmental regulation of cloned superabundant embryo mRNAs in soybean. Dev. Biol. **83**, 218–231.

Higgins, T. J. V. (1984). Synthesis and regulation of major proteins in seeds. Annu. Rev. Plant Physiol. **35**, 191–221.

Higgins, T. J. V., Newbigin, E. J., Spencer, D., Llewellyn, D. J., and Craig, S. (1988). The sequence of the pea vicilin gene and its expression in transgenic tobacco plants. Plant Mol. Biol. **11**, 683–689.

Hill, J. E., and Breidenbach, R. W. (1974). Proteins of soybean seeds. I. Isolation and characterization of major components. Plant Physiol. **53**, 742–746.

Hirano, H., Kagawa, H., Kamata, Y., and Yamaguchi, F. (1987). Structural homology among the major 7S globulin subunits of soybean seed storage proteins. Phytochem. **26**, 41–45.

Holowach, L. P., Madison, J. T., and Thompson, J. F. (1986). Studies on the mechanism of regulation of the mRNA level for a soybean storage protein subunit by exogenous L-methionine. Plant Physiol. **80**, 561–567.

Hymowitz, T., and Singh, R. J. (1987). Taxonomy and speciation. In Soybean: Improvement, Production, and Use. J. R. Wilcox, ed (Madison, WI: American Society of Agronomy), pp. 23–48.

Jofuku, K. D. (1987). Developmental regulation of soybean seed protein gene expression. Ph.D. Dissertation, University of California, Los Angeles, California.

Jofuku, K. D., and Goldberg, R. B. (1988). Analysis of plant gene structure. In Plant Molecular Biology: A Practical Approach, C. H. Shaw, ed (Oxford: IRL Press), pp. 37–66.

Jofuku, K. D., and Goldberg, R. B. (1989). Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. Plant Cell, in press.

Kamalay, J. C., and Goldberg, R. B. (1980). Regulation of structural gene expression in tobacco. Cell **19**, 935–946.

Kamalay, J. C., and Goldberg, R. B. (1984). Organ-specific nuclear RNAs in tobacco. Proc. Natl. Acad. Sci. USA **81**, 2801–2805.

Kanehisa, M. I. (1982). Los Alamos sequence analysis package for nucleic acids and proteins. Nucl. Acids Res. **10**, 183–196.

Ladin, B. F., Tierney, M. L., Meinke, D. W., Hosangadi, P., Veith, M., and Beachy, R. N. (1987). Developmental regulation of β-

conglycinin in soybean axes and cotyledons. Plant Physiol. **84,** 35–41.

Lei, M. G., Tyrell, D., Bassette, R., and Reeck, G. R. (1983). Two-dimensional electrophoretic analysis of soybean proteins. J. Agric. Food Chem. **31,** 963–968.

Lycett, G. W., DeLauney, A. J., Gatehouse, J. A., Gilroy, J., Croy, R. R. D., and Boulter, D. (1983). The vicilin gene family of pea: A complete cDNA coding sequence for preprovicilin. Nucl. Acids Res. **11,** 2367–2380.

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., Dubé, P. H., and Beachy, R. N. (1988). Differential expression of conglycinin α' and β subunits in transgenic plants. Plant Mol. Biol. **11,** 109–123.

Nielsen, N. C., Dickinson, C. D., Cho, T.-J., Thanh, V. H., Scallon, B. J., Fischer, R. L., Sims, T. L., Drews, G. N., and Goldberg, R. B. (1989). Characterization of the glycinin gene family in soybean. Plant Cell **1,** 313–328.

Okamuro, J. K., and Goldberg, R. B. (1989). Regulation of plant gene expression: General principles. In The Biochemistry of Plants, Vol. **15,** A. Marcus, ed (San Diego: Academic Press), pp. 1–82.

Okamuro, J. K., Jofuku, K. D., and Goldberg, R. B. (1986). Soybean seed lectin gene and flanking nonseed protein genes are developmentally regulated in transformed tobacco plants. Proc. Natl. Acad. Sci. USA **83,** 8240–8244.

Rosbash, M., Blank, D., Fahrner, K., Hereford, L., Ricciardi, R., Roberts, B., Ruby, S., and Woolford, J. (1979). R-looping and structural gene identification of recombinant DNA. Methods Enzymol. **68,** 454–469.

Schuler, M. A., Ladin, B. F., Fryer, G., Pollaco, J. G., and Beachy, R. N. (1982) Structural sequences are conserved in the genes coding for the α', α, and β subunits of the soybean seed storage protein and other seed proteins. Nucl. Acids Res. **10,** 8245–8260.

Seed, B. (1982). Diazotizable arylamine cellulose papers for the coupling and hybridization of nucleic acids. Nucl. Acids Res. **10,** 1799–1810.

Sengupta, C., Deluca, V., Bailey, D. S., and Verma, D. P. S. (1981). Posttranscriptional processing of 7S and 11S components of soybean storage proteins. Plant Mol. Biol. **1,** 19–34.

Staden, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Acids Res. **10,** 2951–2961.

Sugimoto, T., Momma, M., Hashizume, K., and Saio, K. (1987). Components of storage protein in hypocotyl-radicle axis of soybean seeds. Agric. Biol. Chem. **51,** 1231–1238.

Sykes, G. E., and Gayler, K. R. (1981). Detection and characterization of a new β-conglycinin from soybean seeds. Arch. Biochem. Biophys. **210,** 525–530.

Thanh, V. H., and Shibasaki, K. (1976). Heterogeneity of β-conglycinin. Biochim. Biophys. Acta **439,** 326–338.

Thanh, V. H., and Shibasaki, K. (1978). Major proteins of soybean seeds: Subunit structure of β-conglycinin. J. Agric. Food Chem. **26,** 692–695.

Tierney, M. L., Bray, E. A., Allen, R. O., Ma, Y., Drong, R. F., Slightom, J., and Beachy, R. N. (1987). Isolation and characterization of a genomic clone encoding the β-subunit of β-conglycinin. Planta **172,** 356–363.

Wahl, G. M., Stern, M., and Stark, G. R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenyloxymethyl paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76,** 3693–3697.

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