Plant Molecular Biology **22**: 873–885, 1993.

Upstream regulatory sequences from two β-conglycinin genes

Philip A. Lessard [1], Randy D. Allen [2], Toru Fujiwara [3] and Roger N. Beachy [4,*]
Department of Biology, Washington University, St. Louis, MO 63130, USA; Present addresses: [1] *Monsanto Company, 700 Chesterfield Village Parkway, Chesterfield, MO 63198, USA;* [2] *Department of Biological Sciences/Department of Agronomy, Texas Tech University, Lubbock, TX 79409, USA;* [3] *Laboratory of Plant Nutrition, Department of Agricultural Chemistry, Faculty of Agriculture, The University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113, Japan;* [4] *Division of Plant Biology-MRC7, Department of Cell Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA (* author for correspondence)*

Received 4 January 1993; accepted in revised form 21 April 1993

Key words: β-conglycinin, gene expression, legumin box, seed storage protein, vicilin box

Abstract

Genes encoding the β-conglycinin seed storage proteins of soybean are expressed only in seeds during specific stages of development. The different subunits of β-conglycinin, α′, α and β, are encoded by distinct members of a gene family. Yet there are marked differences in the regulation of the genes encoding the α′/α and β subunits. Previous work (Chen *et al.*, EMBO J 7: 297–302, 1988) identified a seed specific transcriptional enhancer upstream of a gene encoding the α′ subunit. Mutations were made within this region to discern its functional components. Among those identified is a 62 bp region (between −77 and −140) that contains a vicilin box consensus sequence as well as a sequence that binds the soybean nuclear factor SEF4 *in vitro*. A second region, which contains a sequence homologous to the core of the legumin box consensus (i.e., CATGCAT-like or RY repeat element) at −246, was also shown to affect the activity of this enhancer in transgenic plants. A series of 5′ terminal deletions were used to identify regulatory elements upstream of the β subunit gene. Two regions were identified (from −553 to −442 and from −308 to −72) that, when deleted, led to a marked reduction in gene expression. Both of these elements contain sequences that bind SEF4 *in vitro*. The distal element also contains an AT-rich segment that recognizes a second nuclear factor, SEF1, *in vitro*. Neither of these elements contains any homology to the vicilin box consensus.

Introduction

β-Conglycinin comprises approximately 30% of the total seed protein in soybeans [34]. This 7S vicilin-like storage protein is composed of three distinct subunits, labelled α′, α and β. Expression of the genes that encode these subunits is activated midway through seed development. Shortly thereafter, mRNA from the α′/α genes alone accounts for roughly 10% of the mRNA mass in cotyledons [21]. By maturity, these mRNAs are virtually undetectable [21, 34].

Although there is considerable sequence homology between the three major subunits of β-conglycinin [14], regulation of the α′ and β subunit genes clearly differs. Individual copies of

the α′ subunit gene appear to be expressed at much higher levels than the β subunit genes [7, 33, 35]. In developing soybean seeds, α′/α subunit mRNA begins to accumulate three to five days earlier than β subunit mRNAs [35]. In addition, the β subunit gene responds to physiological signals such as ABA and methionine while the α′ subunit gene does not [6, 17, 24, 25]. Finally, the subunits encoded by these genes are differentially distributed within the developing soybean embryo [32].

Much of the control of the β-conglycinin genes occurs at the level of transcription [2, 23], and the characteristic regulation of these genes is conserved when they are expressed in transgenic petunia (*Petunia hybrida*) or tobacco (*Nicotiana tabacum*) plants [2–4, 7, 17, 35]. Chen *et al.* [12, 13] used transgenic plants to identify a short (180 bp) element from the upstream region of a cloned α′ subunit gene that specifically enhances transcription in seeds midway through development. Considerably less is known about regulatory elements that act upon the β subunit gene. Within 1 kb upstream of the β subunit coding sequence, there is no region homologous to the enhancer found upstream of the α′ subunit gene. This implies that different mechanisms may be responsible for maintaining the strict pattern of expression for each gene.

In an effort to better understand the mechanisms that regulate expression of these genes sequences from the upstream regions of both genes were altered and linked to a reporter gene. Expression of these chimeric genes was monitored in transgenic plants. Mutagenesis of the α′ subunit gene enhancer revealed two or more functional elements. Terminal deletions of the β subunit gene were used to identify sequences responsible for expression of this gene.

Materials and methods

Plasmid construction, plant transformation

A reporter cassette plasmid was constructed in pUC19 [39] composed of the coding sequence of the GUS gene (from pRAJ260) [28] linked to the termination signal from the α′ subunit gene [12]. These segments are preceded in the cassette by six unique restriction sites. URS fragments from the α′ subunit gene were derived from a genomic clone [13], while fragments from the β subunits gene were derived from the plasmid pTZ18Rpβ C79 [18]. These were inserted into the cassette by standard procedures. Plasmid constructions were confirmed by DNA sequencing. URS-reporter-termination signal units were then excised as *Eco* RI-*Hin*d III fragments, ligated into the binary vector pMON505 [27] and introduced into *Agrobacterium tumefaciens* strain GV3111SE [16] via triparental mating, using the helper plasmid pRK2013 [15]. Tobacco (*Nicotiana tabacum* L. cv. Xanthi; for linker scan experiments) or *N. benthamiana* (for all other transformations) were transformed according to the leaf disk protocol of Horsch *et al.* [26]. Transformation was confirmed by assaying the ability of leaf discs from young leaves to callus on medium containing kanamycin [26]. Genomic southerns revealed low insert copy number among the plants tested. The GUS phenotype segregated as 3:1 among seeds of nearly all expressing plants. Developing seeds from transgenic plants were harvested at the appropriate developmental stages and frozen at −80 °C until they were assayed for GUS activity.

β-glucuronidase (GUS) assay

The fluorometric assay for GUS activity was carried out essentially as described by Jefferson [29]. Protein concentrations were determined by the method of Bradford [5]. The lower limit of detection in these assays is defined as ca. 200 pmol MU per hour per mg as this level of activity cannot be distinguished from fluctuations in assays of 'no protein' controls. Seeds from untransformed plants generated GUS activity at or below this level.

LacZ/gusA *fusion protein expression in* Escherichia coli

The plasmids pUC118 and pBluescript KS+ were digested with *Acc* I, 3′ recessed ends were blunted using Klenow fragment of *E. coli* DNA polymerase and dNTPs, and then religated. This adjusted the reading frame of *lacZ* (encoded by each plasmid) by inserting two basepairs. Plasmid DNA from strains of *Agrobacterium* that hold the LS1, LS2 and LS3 constructs was digested with *Eco* RI and *Hin*d III, and the appropriate fragments were ligated into pUC19. The coding sequence from the *gusA* gene was removed from each of these as a *Pst* I fragment (sites preserved from original pRAJ260 plasmid) [28] and ligated into the frameshift-adjusted pUC118 and pBluescript KS+ vectors (as parallel experiments). *E. coli* (DH5αF′) cells transformed with these plasmids were selected on ampicillin-containing (200 mg/l) LB plates augmented with 20 mg/ml 5-bromo-4-chloro-3-indolyl-β-*D*-glucuronide (40 μl per plate; Jersey Lab & Glove Supply) and 200 mg/ml IPTG (4 μl per plate; Fisher). Blue colonies were selected, and the polylinker and 5′ portion of the *gusA*-coding sequence was examined in each construct by DNA sequencing. GUS activity was measured in pUC118-or pBluescript-derived clones by pelleting for 8 h 1.5 ml bacterial cultures and lysing the pellet in 300 μl GUS extraction buffer [29] + 0.1 mg/ml lysozyme (Sigma) for 15 min at room temperature. These lysates were centrifuged and 5 or 10 μl aliquots from the supernatant were tested via the fluorometric GUS assay [29]. Negative controls included extracts from bacteria carrying plasmids with no GUS insert or with the GUS gene inserted in the antisense orientation.

Statistical treatment of data

The distributions of values for GUS activity were compared for plants carrying each construct to determine whether differences in expression between constructs were significant. These comparisons were carried out using SAS software (SAS Institute). Because the values for GUS activity among plants carrying the same construct did not follow a normal distribution, a variety of non-parametric tests were applied to the samples of data, including the Wilcoxon test, Kruskal-Wallis test, Median 2-sample test and the Kolmogorov-Smirnov 2-sample test [37]. In each case, the level of significance was set at $P \leq 0.05$.

Results

Linker scan experiments

Extracts from developing soybean seeds contain a factor, termed 'SEF3', that binds to sequences within the seed specific enhancer of the α′ subunit gene [1]. A similar factor in tobacco seed extracts binds to the same region [33]. These observations led to the hypothesis that SEF3 may mediate enhancer activity. When the binding site for SEF3 was originally identified [1], a series of linker scan (LS) mutations were generated in which specific sequences in the enhancer element were replaced with synthetic linkers. The position of each of these mutations is indicated in Fig. 1. These modified enhancer elements were then tested for their ability to bind SEF3 in gel mobility shift assays [1]. In that study, linkers that disrupted the AACCCA repeats in this region also reduced SEF3 binding. A series of experiments were conducted to utilize these linker scan mutations to assess the role of the binding site in enhancer function. Each of the linker-mutagenized enhancers (which represent sequences from −257 to −77 relative to the start of transcription) were linked to the α′ subunit gene 'minimal promoter (−76 to +13; identical to fragment 'd' below) and used to drive expression of the *gusA* gene in transgenic tobacco plants. β-Glucuronidase enzyme (GUS) activity was measured in mature seeds [24 days after anthesis (daa)] from seven to twelve plants for each of the mutations and compared with the wild-type control, 'C6'; results are shown in Fig. 2.

It is clear from these results that the ability of each mutant to bind SEF3 *in vitro* does not correlate with enhancer activity *in vivo*. LS1 and LS5,

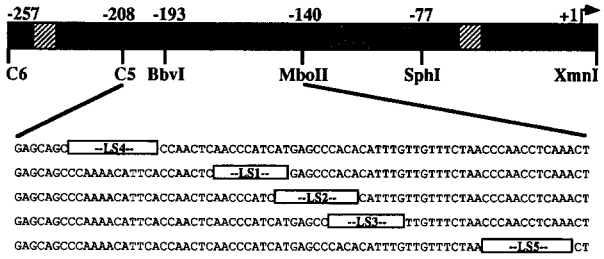

Fig. 1. Linker scan (LS) mutagenesis of the seed-specific enhancer from the α′ subunit gene. *Xba* I linkers (white boxes) replaced native sequence as described elsewhere [1]. In the context of the entire URS (−257 to + 13), each of these mutations as well as the unaltered control ('C6') were used to drive *gusA* expression in transgenic plants. The name and position of each LS mutant is indicated relative to the surrounding sequence; for clarity, only the sequence between the C5 terminus (−208) and the *Mbo* II site (−140) is presented. Notice that each of these mutations (except LS4) lies within fragment 'b' (see Fig. 3). Numbers indicate nucleotide position relative to transcription start; C6 and C5, endpoints of 5′ terminal deletions generated through *Bal* 31 digestion [13]. Hatched boxes represent conserved sequences described in Fig. 10.

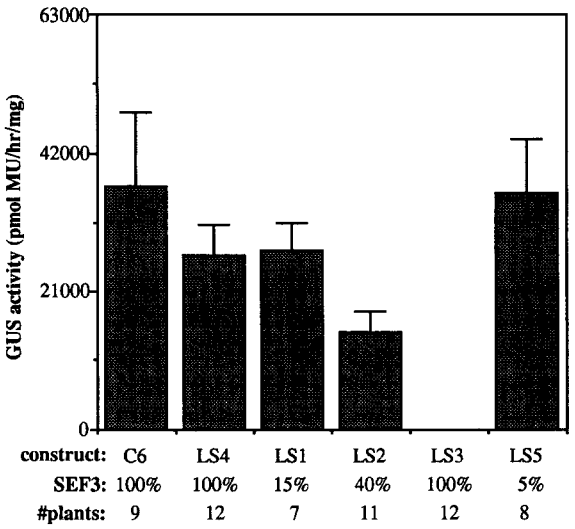

Fig. 2. GUS activity in transgenic LS plants. GUS activity was measured in 24 daa seeds from plants carrying reporter genes with the LS constructs described in Fig. 1. Shaded boxes represent the mean of these results (with standard error bars) for each construct. Statistical tests could not distinguish the GUS activities from LS1, LS2, LS4, LS5 and C6 with greater than 90% certainty. Only LS3 had significantly lower activity ($P = 0.0001$). SEF3: the ability of each of these mutants to bind SEF3 *in vitro* [1]. The number of independently transformed plants examined in each case is listed.

each of which had very reduced affinity for SEF3 in gel shift experiments [1], had activities *in vivo* near wild-type (C6) as did LS4, which binds SEF3 as well as the wild-type sequence. Surprisingly, plants with LS3, which had wild-type SEF3-binding activity, produced no detectable GUS activity in seeds. LS2 had an intermediate affinity for SEF3 [1]. Although the average GUS activity in seeds of these and control plants suggest that the plants carrying the LS2 mutation also had intermediate expression, statistical tests did not support this interpretation (see Materials and methods).

It was possible that the lack of expression in LS3 plants was caused by a frame-shift or point mutation that occurred while the transgene was in *Agrobacterium*. To test this the *gusA*-coding sequences were excised from the binary plasmid vectors in *Agrobacterium* hosts carrying LS1, LS2 and LS3. Each of these was then linked in-frame to the *lacZ* coding sequences in pUC119 and pBluescript KS + (Stratagene). The genes thus expressed in *E. coli* produced functional β-glucuronidase enzyme (data not shown). Regardless of whether the gene was derived from LS1, LS2 or LS3 *Agrobacterium* strains, the activity of GUS produced in *E. coli* was several-fold higher than the 'background' GUS activity normally associated with the *E. coli* strain used for these studies (DH5αF′). These results ruled out the possibility that the reporter gene had been rendered inactive prior to plant transformation. In addition, DNA sequence analysis indicated that the ca. 270 bp upstream regulatory sequences (URSs) from LS3, LS1 and LS2 had not been altered by introduction into *Agrobacterium* (data not shown).

Reconstruction of the α′ gene seed-specific enhancer

The results described above indicated that the sequences that bind SEF3 in soybean nuclear extracts [1] and a similar factor in tobacco seed nuclear extracts [33] are not required for expression in transgenic tobacco. The region that was

disrupted by LS3 is, however, likely to be important for expression. These unexpected results led to a broader examination of sequences in 180 bp seed-specific enhancer. To test whether other regions in the enhancer bear functional elements, the enhancer was dissected into three fragments (labelled 'a', 'b' and 'c') by digestion with the appropriate restriction enzymes as indicated in Fig. 3. Thus fragment 'a' lies entirely upstream of the region examined by the linker scan experiments, fragment 'b' contains the SEF3-binding site and fragment 'c' contains sequences downstream of the region examined by the linker scan mutations. A fourth fragment, 'd', comprises the minimal promoter of the α′ subunit gene and contains the CCAAAT and TATAAA elements as well as the transcription start site (Fig. 3). The 3′ terminus of fragment 'd' is located at + 13 (*Xmn* I site). These four fragments were reassembled as depicted in Fig. 3. Two additional constructs were included which represent short terminal deletions of fragment 'a'. The first of these (denoted with the 'a' component underlined: a̲bcd) has its 5′ terminus at −228 (*Rsa* I site) while the second (denoted with the 'a' component double-underlined: a̳bcd) extends only to −208 (the 'C5' terminus).

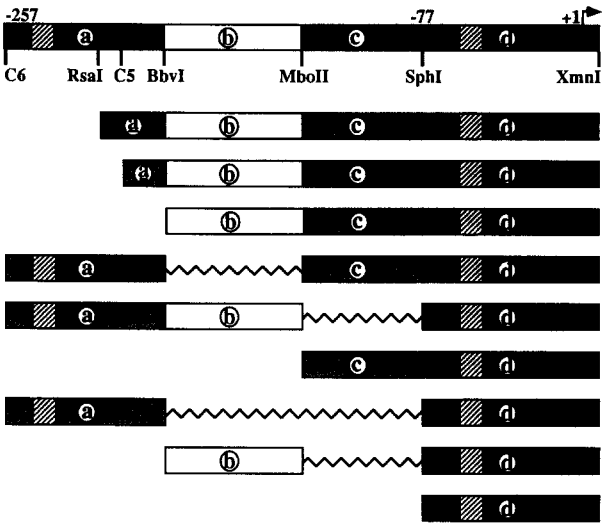

Fig. 3. Reconstruction of the seed specific enhancer from the α′ subunit gene. The enhancer was divided into three fragments, 'a', 'b', and 'c', and reassembled as depicted. These were then linked to the minimal promoter, fragment 'd'. Zigzag lines indicate regions where fragments were replaced with polylinker sequence; spacing was not rigorously conserved in these constructions. Hatched boxes represent conserved sequences described in Fig. 10.

These chimeric URSs were inserted into a *gusA* reporter cassette and introduced into *Nicotiana benthamiana*. Seeds were harvested at 24 daa (in *N. benthamiana* most seed capsules dehisce by 26 or 28 daa), and GUS activity was measured for each plant. The results of these assays are shown in Fig. 4. Statistical tests were also applied to the data to evaluate these results (see Materials and methods).

The activity associated with the intact enhancer ('abcd') was defined as wild-type. Several interesting results emerged from a comparison of the different transgenic lines. First, the minimal promoter alone, '---d' did not generate GUS activity above the lower limit of detection. Sec-

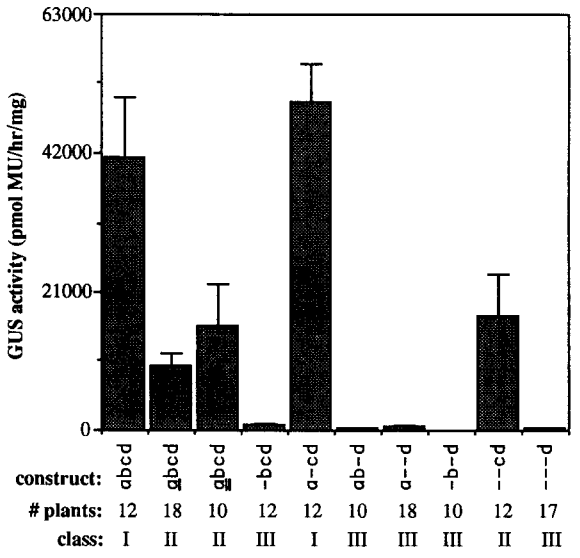

Fig. 4. GUS activity in plants carrying the reconstructed enhancers. GUS activity was measured in seeds (24 daa) from plants carrying reporter genes with the constructs described in Fig. 3. Shaded boxes represent the mean of these results (with standard error bars) for each construct. Statistical tests distinguished three classes of expression: Class I plants had wild-type GUS activity; Class II plants had significantly less activity ($P < 0.05$); and Class III plants had still less activity ($P < 0.001$) and could not be distinguished from background. Underlining indicates partial deletion of fragment 'a' as depicted in Fig. 3. Number of independently transformed plants examined for each construct is indicated.

ond, fragment 'c' is required for activity; every construct with significant activity included 'c', while every construct that lacked 'c' had undetectable activity. Fragment 'c' was active when linked alone to the minimal promoter ('--cd') and gave ca. 40% wild-type activity. Addition of fragment 'a' increased the activity of the '--cd' components to wild-type levels. The terminal deletions of fragment 'a' had less activity than the intact URS (compare 'abcd', 'a̲bcd' and 'a̲bcd'). In the absence of fragment 'a', fragment 'b' had a strong negative effect on expression (compare '--cd' and '-bcd'). This negative effect was alleviated by the presence of at least a portion of fragment 'a' (compare '-bcd' and 'abcd', 'a̲bcd' and 'a̲bcd').

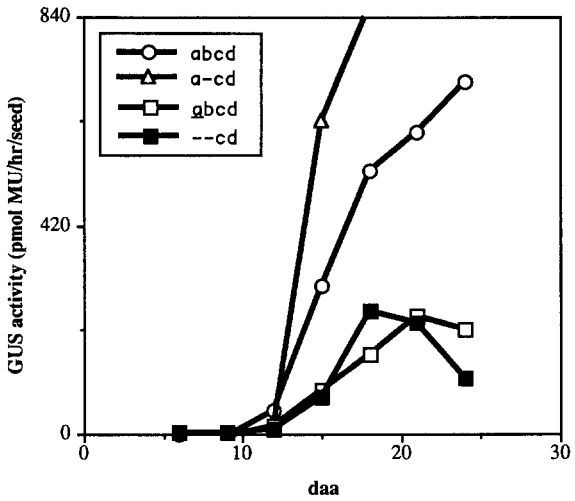

Fig. 5. Timecourse of GUS accumulation during seed development in plants carrying the reconstructed enhancers. Seeds were harvested at three-day intervals from 6 to 24 daa. GUS activity was measured at each stage. Number of plants averaged to determine each timecourse: 'abcd', 2 plants; '--cd', 3 plants; 'a-cd', 5 plants; 'a̲bcd', 6 plants. Both 'abcd' plants had highest activity at 24 daa; three 'a-cd' and three 'a̲bcd' plants had peak activity at 24 daa while the five other plants in these series peaked at 21 daa; all three '--cd' plants had peak activity at 18 daa and lost activity thereafter.

Temporal regulation of the reconstructed enhancer elements

To examine whether the timing of expression of any of these chimeric URSs deviated from the wild-type, GUS activity was monitored in developing seeds from several plants. This comparison is depicted in Fig. 5. Of these constructs, '--cd' demonstrated the most significant perturbation in timing of expression relative to the control ('abcd'). The 'a̲bcd' construct exhibited a somewhat intermediate timing of expression. With the '--cd' URS, GUS activity in each of the three plants examined peaked at 18 daa, whereas the 'abcd', 'a̲bcd' and 'a-cd' URSs all had peak activity between 21 and 24 daa. GUS activity in the '--cd' seeds decreased by ca. 50% by 24 daa. This loss of GUS activity suggests that in mature seeds GUS enzyme has a half-life of no greater than 72 h. This estimate is similar to the ca. 50 h half-life reported elsewhere [30].

Terminal deletions of the β subunit gene URS

Although genes encoding the α′ and β subunit genes of β-conglycinin are both expressed solely in seeds, upstream sequences from these two genes share very little homology. To determine what sequences regulate expression of the β subunit gene, a series of terminal deletions was prepared from the β URS. Nine terminal deletions were constructed by digesting the plasmid with restriction enzymes as presented in Fig. 6. These truncated URS fragments were ligated into the same *gusA* reporter cassette that had been used for the enhancer reconstruction experiments above. These constructs were then transformed into *Nicotiana benthamiana* cells, and fertile plants were regenerated.

Seeds were harvested from transformed plants at 24 daa and assayed for activity of the reporter enzyme. Results of these assays are depicted in Fig. 7. As deletions progressed from the *Hin*d III site at −1046 (βHB) to the *Sac* I site at −847 (βSB), there was a statistically significant increase in expression, indicating that the deletions may have disrupted some element that interfered with expression from this promoter. The expression from deletions with termini at −714 (βXB) or −553 (βAB) was similar to that of the longer −919 (βDB) URS. Statistical tests bore out this

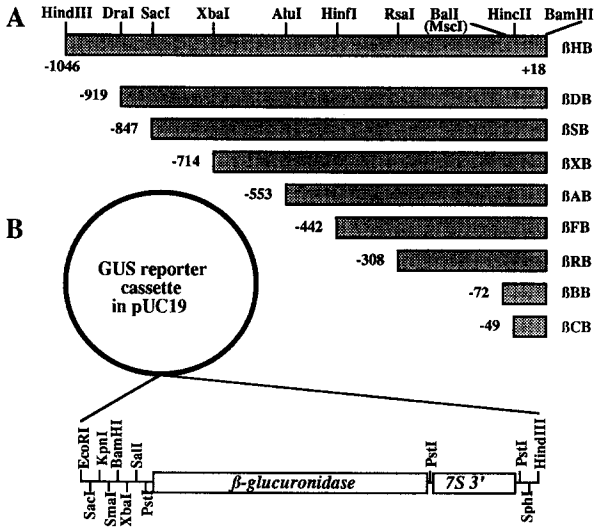

Fig. 6. Construction of β subunit gene terminal deletions. A. Terminal deletions of the β subunit URS. A plasmid carrying sequences from −1046 to + 18 from the β subunit gene as a *Hin*d III-*Bam* HI fragment has been described elsewhere [18]. Restriction fragments were excised by digesting with *Bam* HI and the appropriate second enzyme. B. GUS reporter gene cassette. Each of the truncated URS fragments was inserted into the polylinker of this cassette and transformed into *N. benthamiana* plants. This cassette was also used for the LS and enhancer reconstruction experiments described in Figs. 1 and 3.

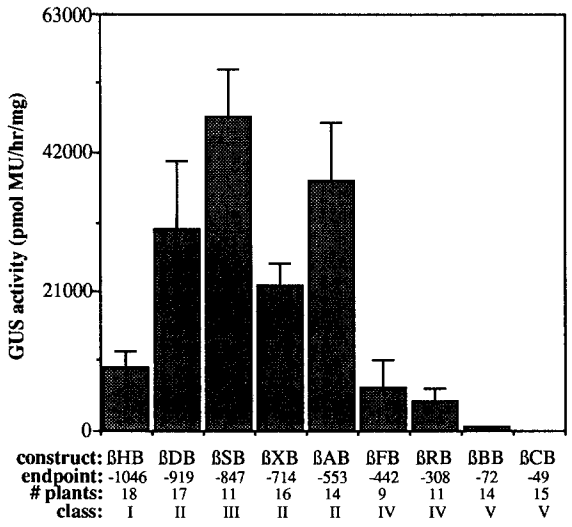

Fig. 7. Expression of the *gusA* reporter gene in transgenic plants. GUS activity was measured in seeds harvested at 24 daa from plants carrying the constructs described in Fig. 6. Statistical tests were used to distinguish five classes of expression. The activity of constructs within each class could not be distinguished while distinctions between classes were statistically valid ($P<0.05$). The number of independently transformed plants examined in each case is indicated. Shaded boxes represent the average of these results for each construct with standard error bars.

interpretation (see Materials and methods). Deletion from −553 to −442 (βFB), however, resulted in a dramatic decrease in promoter activity. Expression of the −442 (βFB) URS was significantly lower than that observed with the longest URS construct (−1046; βHB) as well ($P<0.01$), but could clearly be distinguished from background. This level of activity was reproduced by a shorter URS deletion to −308 (βRB). Removing sequences to −72 (βBB) or to −49 (βCB) produced URS fragments whose activity could not be distinguished from background. These results indicated the presence of major *cis*-acting regulatory elements between −553 and −442 and between −308 and −72.

Expression of terminal deletions during seed development

The expression of the *gusA* gene from many of these truncated URSs was also examined over the course of seed development. Seeds were harvested at three-day intervals from 6 daa until 24 daa. Fifty seeds were dissected from developing capsules at each stage and assayed for activity of the reporter enzyme. Results are presented in Fig. 8. Other than a few, relatively minor perturbations, 5′ terminal deletion of upstream sequences had very little effect on the timing of expression of each construct. No terminal deletion permitted expression in seeds earlier than the −919 (βDB) construct. GUS activity in seeds was very low for each of the contructs until 15 or 18 daa after which expression increased substantially. At 15 daa, the activity of the −442 construct (βFB) was slightly higher than that of the other constructs. This difference is probably not significant, however, because the standard error bars of activity of many of these constructs overlap at this timepoint (error bars were omitted from the graph for clarity). GUS activity in the shortest URS construct (−308; βRB) was undetect-

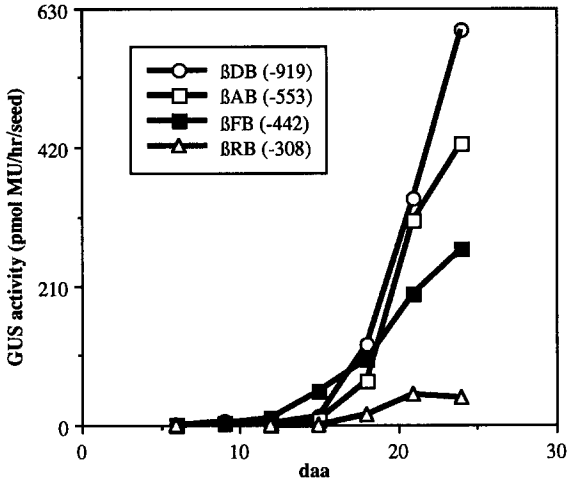

Fig. 8. Timecourse of GUS accumulation during seed development in plants carrying truncated β subunit URSs. GUS activity was compared among plants carrying four different terminal deletions. Seeds were harvested at three-day intervals from 6 to 24 daa. GUS activity was measured at each stage. Timecourses were obtained from multiple plants in each case and the results averaged. Number of plants examined in each case as follows: βDB, 8; βAB, 6; βFB, 5; βRB, 4.

able at 15 daa. This probably reflects the overall lower level of expression of the −308 deletion. Because expression from even the shortest URS examined in these comparisons does not increase until 15 daa or later, elements that control timing of this expression must lie downstream of the *Rsa* I site (−308).

Comparison of α′ and β URS activity

When expressed in transgenic plants, the genes that encode the α' and β subunits of β-conglycinin exhibit much of the differential expression they experience in soybeans [2, 17, 35]. Most notably, there is higher expression of the α' subunit gene than there is of the β subunit gene, and expression of the α' subunit gene can be detected earlier than that of the β subunit gene [35]. To determine whether URSs from each of these two genes could confer these properties to a reporter gene, expression of the β URS constructs was compared to similar constructs carrying URSs derived from the α' subunit gene. Table 1 presents the results

Table 1. Comparison of activities of URSs from α' and β subunit genes.

URS	GUS activity	Number of plants
α' (−828)	399.4 ± 60.1	13
α' (−257)	41.2 ± 9.2	12
β (−919)	30.4 ± 10.3	17

Sequences from the α' subunit gene extending from −828 to +13 or from −257 to +13 (relative to the start of transcription) were inserted into the GUS reporter cassette (Fig. 6b) and introduced into transgenic plants. GUS activity in 24 daa seeds from these plants is expressed in 10^3 pmol MU per hour per mg and is compared to that from the −919 (βDB) construct described in Fig. 6. The number of independent transgenic plants examined in each case is indicated.

of this comparison. This results demonstrate that 828 bp of upstream sequence from the α' subunit gene can induce approximately 10-fold higher expression than 919 bp from the β subunit gene. A shorter segment from the α' subunit gene (whose 5′ end extends only to −257) conferred levels of expression similar to that of the longer β subunit gene URS. This indicates that elements accounting for the stronger activity of the α' URS lie upstream of −257 in that gene. The timing of activity of these URSs also differed in transgenic plants. As depicted in Fig. 9, expression from the −919 β subunit gene URS was first detected three days later than that of either of the two α' URSs examined.

Whereas −828 and −257 URSs from the α' subunit gene produced markedly different levels of reporter gene activity, Chen *et al.* [13] reported that −904 and −257 terminal deletions of a genomic clone of the α' subunit gene produced nearly equivalent levels of RNA and protein in transgenic plants. These results may differ in part because genomic clones rather than reporter genes were used in the earlier work.

Discussion

Analysis of the seed-specific enhancer

Several candidate regulatory elements lie between nucleotide positions −257 and +13 in the α'

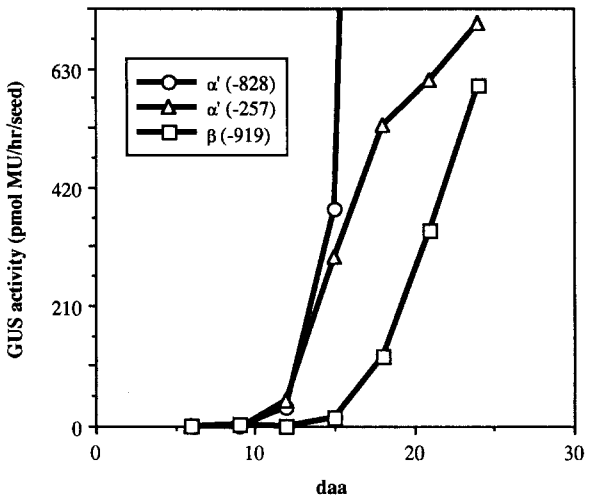

Fig. 9. Timing of α′ and β subunit promoter activity in transgenic plants. Curve for βDB is same as in Fig. 8. The −257 α′ promoter used is identical to the 'abcd' promoter described in Fig. 3 and in Fig. 5. In addition, thirteen plants were examined for the timecourse of the −828 α′ promoter.

subunit gene; the locations of these elements are shown in Fig. 10. The enhancer also possesses binding sites for two protein factors, SEF3 and SEF4, found in nuclear extracts of developing soybean seeds [1, 33]. Any of these or other unidentified components could play a role in mediating both the enhancing activity and the seed specificity of this fragment.

A vicilin box and SEF4-binding site lie within quantitative elements

From the enhancer reconstruction experiments (Figs. 3 and 4), it is clear that fragment 'c', in association with the minimal promoter, is necessary and sufficient to direct expression in the seed. Within this region is a segment that shares significant homology with the vicilin box elements described from other seed storage protein genes [19, 36]. Sequence conservation and the fact that the vicilin box comprises >70% of fragment 'c' suggest that this element is a major functional component of the enhancer. However, this element cannot, as yet, be separated from the SEF4-binding site. Lessard *et al.* [33] reported that

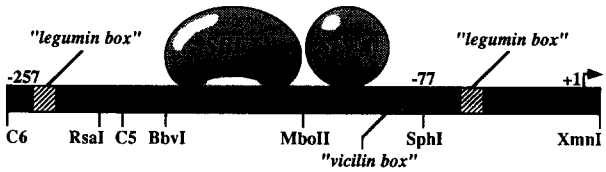

Fig. 10. Schematic representation of the immediate upstream region of the α′ subunit gene. The relative positions of various restriction sites used in Fig. 3 are indicated. SEF3 and SEF4 are positioned at the sequences each recognizes *in vitro* [33]. Locations of the vicilin box and RY repeat elements (legumin boxes) are also indicated.

SEF4 activity increased in extracts from developing soybean seeds midway through development, parallelling increases in β-conglycinin transcripts. This suggests that SEF4 could be involved in transcriptional regulation of the α′ subunit gene.

The SEF4-binding site is very similar to the AT-1-binding site found in several light regulated genes [20]. Although the independent function of the AT-1 site has not been determined, it has been correlated with positive transcriptional activity of the tomato *rbcS-3A* gene [38]. Additional SEF4 binding sites are found upstream of −257 in the α′ URS [33]. These sites are not associated with vicilin box elements, but they bind with high affinity indicating that SEF4 binding *in vitro* does not require the recognition site to be embedded within a vicilin box. These upstream sites may account for the higher activity associated with the longer α′ URS (Table 1).

A region containing an RY repeat can augment expression from the α′ promoter

The most distal component of the α′ subunit gene's seed specific enhancer, fragment 'a', can augment the activity of fragment 'c'. Unlike fragment 'c', however, fragment 'a' alone is not sufficient to enhance transcription from the minimal promoter. Within fragment 'a' is a short region homologous to the legumin box core sequence – the so-called CATGCAT or RY-repeat element [19, 36]. Deletion of a region containing this element reduced the activity of the enhancer

(Figs. 3 and 4). Other work has shown that specific alteration of the RY-repeat sequence in fragment 'a' led to a decrease in promoter activity [11] (T. Fujiwara and R.N. Beachy, unpublished results). Chamberland *et al.* [11] have also shown that mutation of a second RY-repeat element lying within fragment 'd' (the minimal promoter) led to a decrease in gene expression, and that mutations in both elements reduced expression to nearly one tenth the wild-type level. These results suggest that the RY repeat is responsible for the increase in activity when fragment 'a' is linked to construct '– –cd'. It is clear from our results, however, that fragment 'a' only acts in conjunction with other elements.

In addition to the RY repeat, fragment 'a' is likely to contain other functional elements. These sequences which lie near the 3′ end of fragment 'a' do not augment the activity associated with fragment 'c', but appear to be involved in overcoming the negative effect associated with fragment 'b' (see below). In addition sequences within fragment 'a' may also promote expression during the later stages of seed development. As shown in Figure 5, GUS activity in seeds carrying the 'abcd', 'abcd' and 'a–cd' constructs increase beyond 18 daa, whereas a shorter construct, '– –cd' lost GUS activity in these later stages.

Fragment 'b' from the α′ promoter contains both positive and negative elements

Fragment 'b' appears to negatively affect transcriptional activity. As described above, this negative activity can be overcome by inclusion of all or part of fragment 'a' (Fig. 4). Other researchers [8, 9] have identified negative elements upstream of β-phaseolin genes that regulate timing or tissue specificity of expression. The negative component of fragment 'b' does not appear to serve either function. The loss of fragment 'b' does not alter the timing of gene expression in seeds (Fig. 5), nor did it lead to aberrant expression of the reporter gene in non-seed tissues. When seedlings from each of these constructs were examined, none had detectable GUS activity in roots, stems or leaves (unpublished observations). Chamberland *et al.* [11] found that even the weak expression they could detect from fragment 'd' alone was restricted to seeds, further indicating that fragment 'b' is not involved in maintaining tissue specificity.

The most unexpected result of the present study came from the experiments with linker-scan-mutagenized URSs. It is important to note that these mutations (except LS4 which had no effect on expression) were confined to the region described as fragment 'b' above. Since deletion of the entire fragment 'b' had little effect on expression of the reporter gene (compare 'abcd' and 'a–cd' in Fig. 4), one would predict that individual mutations within fragment 'b' should also have had little effect on expression. Surprisingly, the LS3 mutation had no detectable expression in 12 independently isolated transgenic plants. This indicates the presence of a critical positive regulatory element between −168 and −157.

Since constructs that lack fragment 'b' entirely have expression patterns that are both quantitatively and temporally identical to that from the intact 'abcd' URS, it is difficult to attribute functional significance to this region. Interplay of the positive and negative components of fragment 'b' might be involved in modulating expression of the α′ subunit gene. However, models to accomodate these observations are only speculative.

Comparison of the α′ and β subunit gene URSs

The present study has confirmed that the differences in timing and level of expression of the α′ and β subunit genes can be conferred to the *gusA* reporter gene by URSs from these genes. GUS accumulation begins three to five days later with the β subunit gene URS than with the α′ subunit gene URS (Fig. 9). Furthermore, 828 bp of upstream sequence from the α′ subunit gene was able to direct ca. ten times as much GUS expression as the strongest β subunit URS (Table 1, Fig. 7). The observed differences in timing and level of expression of these two genes may be due in part to differences in where these genes are

expressed within the developing seed. In transgenic tobacco, expression from the α' subunit gene can be detected in the endosperm tissue as well as in the embryo while expression of the β subunit gene is restricted to portions of the embryo [2, 3]. That this pattern of expression is regulated at the level of transcription has been demonstrated by promoter swapping experiments and work with reporter gene constructs [2] (P.A. Lessard and R.N. Beachy, unpublished results).

While a SEF4-binding site has been identified within an important positive element (fragment 'c') from the α' subunit gene, additional SEF4-binding sites lie within two positive elements upstream of the β subunit gene [33]. While deletion of the distal (−553 to −442) element led to a marked reduction in gene expression, removal of the proximal element (−308 to −72) resulted in a complete loss of activity (Fig. 7). Neither element from the β subunit gene, however, bears any homology to the vicilin box that surrounds the SEF4 site from the α' subunit gene.

The distal element of the β subunit gene also harbors a binding site for SEF1, a nuclear factor that recognizes AT-rich sequences [33]. Bustos *et al.* [10] and Jordano *et al.* [31] described factors that recognize similar AT-rich regions in regulatory elements upstream of the phaseolin and helianthinin genes. It has been proposed that such AT-rich elements are recognized by components of the nuclear scaffolding that may stimulate transcriptional activity in the region [22]. SEF1 or the AT-rich sequences it recognizes [33] may have a similar activity in the β subunit gene.

In addition to a SEF4 site, the proximal element also contains two RY-repeat sequences, one at −295 and the second at −263. Given the role of these elements in regulation of the α' subunit gene enhancer [11], it is possible that these elements regulate the expression of the β subunit gene as well.

Conclusions

In summary, we have identified at least two functional regions within the seed-specific enhancer element of the soybean α' subunit gene of β-conglycinin. These regions are responsible for transcriptional enhancement of this element in transgenic tobacco seeds. Sequences located between −140 and −77, when fused to a minimal promoter, are necessary and sufficient to direct seed-specific expression of an attached reporter gene. This region contains a binding site for SEF4 as well as a conserved vicilin box motif. Sequences located between −257 and −193 (which include an RY-repeat element) provide increased levels of expression and are implicated in determining the timing of expression. However, this region alone cannot activate transcription from the minimal promoter. The role of the intervening region (−193 to −140) remains enigmatic.

Two upstream regions have been implicated in regulating the level of transcription of the β subunit gene. The distal region (between −553 and −442) harbors binding sites for SEF1 and SEF4 while the proximal region (between −308 and −72) contains an additional SEF4-binding site and two RY-repeat elements. The major differences in expression between the α' and β subunit genes are determined by upstream sequences. Further work is needed to determine which sequences in the β subunit gene confer this gene's responsiveness to ABA and nutritional stress.

Acknowledgements

We would like to thank Dr Z-L Chen and Monsanto Company for providing plasmid DNA and starting materials; D. Droste, J.C. Lessard and M. Scobell for indispensable assistance with plant transformations; M. Dyer, D. Droste and S.J. Leitner for maintenance of greenhouse plants; J. DePaulo for technical assistance; H. Chou and Drs A. Templeton, F. Bernier and L. Farrell for many helpful discussions during the course of this study. This work was supported by US Department of Energy Biological Energy Research Grant DE-FG02-88ER13877.

References

1. Allen RD, Bernier F, Lessard PA, Beachy RN: Nuclear factors interact with a soybean β-conglycinin enhancer. Plant Cell 1: 623–631 (1989).
2. Barker SJ, Goldberg RB: β-conglycinin gene sequences differentially program tobacco endosperm and embryo expression patterns. J Cell Biochem, Suppl 15A: 87 (1991).
3. Barker SJ, Harada JJ, Goldberg RB: Cellular localization of soybean storage protein mRNA in transformed tobacco seeds. Proc Natl Acad Sci USA 85: 458–462 (1988).
4. Beachy RN, Chen Z-L, Horsch RB, Rogers SG, Hoffmann NL, Fraley, RT: Accumulation and assembly of soybean β-conglycinin in seeds of transformed petunia plants. EMBO J 4: 3047–3053 (1985).
5. Bradford MM: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254 (1976).
6. Bray EA, Beachy RN: Regulation by ABA of β-conglycinin expression in cultured developing soybean cotyledons. Plant Physiol 79: 746–750 (1985).
7. Bray EA, Naito S, Pan N-S, Anderson EA, Dubé P, Beachy RN: Expression of the β-subunit of β-conglycinin in seeds of transgenic plants. Planta 172: 364–370 (1987).
8. Burow MD, Sen P, Chlan CA, Murai N: Developmental control of the β-phaseolin gene requires positive, negative, and temporal seed-specific regulatory elements and a negative element for stem and root expression. Plant J 2: 537–548 (1992).
9. Bustos MM, Begum D, Kalkan F, Battraw MJ, Hall TC: Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J 10: 1469–1479 (1991).
10. Bustos MM, Guiltinan MJ, Jordano J, Begum F, Kalkan A, Hall TC: Regulation of β-glucuronidase expression in transgenic plants by an A/T rich *cis* acting sequence found upstream of a French bean β phaseolin gene. Plant Cell 1: 839–853 (1989).
11. Chamberland S, Daigle N, Bernier F: The legumin boxes and the 3′ part of a soybean β-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. Plant Mol Biol 19: 937–949 (1992).
12. Chen Z-L, Pan N-S, Beachy RN: A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. EMBO J 2: 297–302 (1988).
13. Chen Z-L, Schuler MA, Beachy RN: Functional analysis of regulatory elements in a plant embryo-specific gene. Proc Natl Acad Sci USA 83: 8560–8564 (1986).
14. Doyle JJ, Schuler MA, Godette WD, Zenger V, Beachy RN, Slightom JL: The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins. J Biol Chem 261: 9928–9238 (1986).
15. Figuvski DH, Helinski DR: Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc Natl Acad Sci USA 76: 1648–1652 (1979).
16. Fraley RT, Rogers SG, Horsch RB, Eicholtz DA, Flick JS, Fink CL, Hoffman NL, Sanders PR: The SEV system: a new disarmed Ti plasmid vector system for plant transformation. Bio/technology 3: 629–635 (1985).
17. Fujiwara T, Hirai MY, Chino M, Komeda Y, Naito S: Effects of sulfur nutrition on expression of the soybean seed storage protein genes in transgenic petunia. Plant Physiol 99: 263–268 (1992).
18. Fujiwara T, Naito S, Chino M, Nagata T: Electroporated protoplasts express seed specific gene promoters. Plant Cell Rep 9: 602–606 (1991).
19. Gatehouse JA, Evans IM, Croy RRD, Boulter D: Differential expression of genes during legume seed development. Phil Trans R Soc Lond B314: 367–384 (1986).
20. Gilmartin PM, Sarokin L, Memelink J, Chua N-H: Molecular light switches for plant genes. Plant Cell 2: 369–378 (1990).
21. Goldberg RB, Hoschek G, Ditta GS, Briedenbach RW: Developmental regulation of cloned superabundant embryo mRNAs in soybean. Devel Biol 83: 218–231 (1981).
22. Grasser KD, Maier U-G, Haass MM, Feix G: Maize high mobility group proteins bind to CCAAT and TATA boxes of a zein gene promoter. J Biol Chem 265: 4185–4188 (1990).
23. Harada JJ, Barker SJ, Goldberg RB: Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell 1: 415–425 (1989).
24. Holowach LP, Madison JT, Thompson JF: 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 (1986).
25. Holowach LP, Thompson JF, Madison JT: Effects of exogenous methionine on storage protein composition of soybean cotyledons cultured in vitro. Plant Physiol 74: 576–583 (1984).
26. Horsch RB, Fry JE, Hoffman NL, Eicholtz D, Rogers SG, Fraley RT: A simple and general method for transferring genes into plants. Science 227: 1229–1231 (1985).
27. Horsch RB, Klee HJ: Rapid assay of foreign gene expression in leaf discs transformed by *Agrobacterium tumefaciens*: Role of T-DNA borders in the transfer process. Proc Natl Acad Sci USA 83: 4428–4432 (1986).
28. Jefferson RA, Burgess SM, Hirsch D: β-glucuronidase from *E. coli* as a gene-fusion marker. Proc Natl Acad Sci USA 83: 8447–8451 (1986).
29. Jefferson RA: Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep 5: 387–405 (1987).
30. Jefferson RA, Kavanagh TA, Bevan MW: GUS fusions:

β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6: 3901–3907 (1987).
31. Jordano J, Concepción A, Thomas TL: A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. Plant Cell 1: 855–866 (1989).
32. Ladin BF, Tierney ML, Meinke DW, Hosángadi P, Vieth M, Beachy RN: Developmental regulation of β-conglycinin in soybean axes and cotyledons. Plant Physiol 84: 35–41 (1987).
33. Lessard PA, Allen RD, Bernier F, Crispino JD, Fujiwara T, Beachy RN: Multiple nuclear factors interact with upstream sequences of differentially regulated β-conglycinin genes. Plant Mol Biol 16: 397–413 (1991).
34. Meinke DW, Chen J, Beachy RN: Expression of storage-protein genes during soybean seed development. Planta 153: 130–139 (1981).
35. Naito S, Dubé PH, Beachy RN: Differential expression α′ and β subunit genes in transgenic plants. Plant Mol Biol 11: 109–123 (1988).
36. Shirsat AH: Control of gene expression in the developing seed. In: Grierson D (ed) Developmental Regulation of Plant Gene Expression, pp. 153–181. Chapman and Hall, New York (1991).
37. Siegel S, Castellan NJ Jr.: Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, New York (1988).
38. Ueda T, Pichersky E, Malik VS, Cashmore AR: Level of expression of the tomato *rbcS-3A* gene is modulated by a far upstream promoter element in a developmentally regulated manner. Plant Cell 1: 217–227 (1989).
39. Yanisch-Perron C, Vieira J, Messing J: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–119 (1985).