The Plant Cell, Vol. 2, 469–478, May 1990 © 1990 American Society of Plant Physiologists

Differential Expression of the *Arabidopsis* 2S Albumin Genes and the Effect of Increasing Gene Family Size

Philippe Guerche,[a,1] Christine Tire,[b] Fatima Grossi de Sa,[a,2] Ann De Clercq,[b] Marc Van Montagu,[b] and Enno Krebbers[a,3]

[a] Plant Genetic Systems, Josef Plateaustraat 22, B-9000 Gent, Belgium
[b] Rijksuniversiteit Gent, Laboratorium Genetika, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

We studied the expression of the four genes encoding 2S albumin seed storage proteins (*at2S1* to *at2S4*) in *Arabidopsis thaliana*. All four genes followed similar temporal profiles throughout development, but *at2S2* and *at2S3* were expressed at significantly higher levels than *at2S1* or *at2S4*. In situ hybridization showed that *at2S2* to *at2S4* mRNAs were present throughout the embryo, whereas *at2S1* was expressed at levels similar to *at2S2* and *at2S3* in the embryo axis but at only insignificant levels in the cotyledons. The different members of the gene family are, thus, likely to be regulated by different combinations of *cis*-acting elements, but it cannot be ruled out that post-transcriptional factors play a role. We studied the effect of enlarging the gene family by introducing an extra, nearly identical gene driven by the promoter of *at2S1*. The data were consistent with a model in which the expression of *at2S2* to *at2S4* is not affected by that of *at2S1*, and in which, at least at low copy numbers of the introduced gene, there is no limit on the overall amount of RNA that the *at2S* gene family can produce.

INTRODUCTION

A number of plant proteins are encoded by gene families ranging in size from a few genes to more than 100. It is unclear whether there are functional reasons for maintaining multiple genes encoding some proteins whereas other gene products required in similar quantities are produced from single genes. The presence of such gene families raises several fundamental questions. Do the individual genes play different roles in different developmental contexts? Are their levels of expression similar? Are the members of a gene family, or subfamily, coordinately regulated? If so, what might be the effect of introducing more genes driven by the same promoter? These questions have been addressed most extensively in plants using the *rbcS* genes, which encode the small subunit of ribulose bisphosphate carboxylase. Initial studies (Tobin, 1978) demonstrated that *rbcS* genes, aside from being expressed in specific tissues, are regulated by light in a process mediated by phytochrome. As more genes were cloned, however, it became clear that individual members of *rbcS* gene families, while encoding proteins of almost identical amino acid sequences, respond differently to light of different qualities or other developmental stimuli (reviewed by Dean et al., 1989). Assuming that these differences are functionally important, other solutions have been found in species closely related to those studied because analogous genes are not always present; *Nicotiana plumbaginifolia* has only two *rbcS* genes (Poulson et al., 1986). Similar observations have been made for genes encoding another light-regulated, photosynthetic protein, the chlorophyll *a/b*-binding protein (Sheen and Bogorad, 1986; Sullivan et al., 1989).

Another group of plant proteins often encoded by multigene families is the seed storage proteins. These genes are also expressed in a specific pattern, being transcribed in developing embryos and/or endosperm at specific stages of development (Okamuro and Goldberg, 1989). On the basis of solubility and size, different classes of storage proteins have been defined (Higgins, 1984); individual species may express more than one of these classes, which may be expressed at different stages of development (Walling et al., 1986). Each class may be encoded by occasionally one (Goldberg et al., 1983) but often more genes, ranging from a few (Krebbers et al., 1988) to 15 to 20 (Josefson et al., 1987; Scofield and Crouch, 1987), to more than 100 in the case of the maize prolamins. At present, less is known about differences in patterns of expression of individual storage protein gene family members than other gene families. In this work, the expression of all four members of the gene family encoding the 2S albumins of *Arabidopsis thaliana* has been studied.

[1] Current address: Laboratoire de Biologie Cellulaire, INRA, Centre de Versailles, Route de Saint-Cyr, F-78026 Versailles Cedex, France.
[2] Current address: CENARGEN/EMBRAPA, SAIN, Parque Rural, C.P. 10.2372, CEP 70.770, Brasilia DF, Brazil.
[3] To whom correspondence should be addressed.

2S albumins are small, water-soluble proteins consisting of 9-kD and 3-kD chains linked by disulfide bridges, and are found in the seeds of a variety of plants (Youle and Huang, 1981). The first 2S albumin clones were obtained from *Brassica napus* (Crouch et al., 1983), and further work showed that the napins are encoded by a large gene family of 12 to 19 genes (Josefson et al., 1987; Scofield and Crouch, 1987). In contrast, the closely related crucifer *Arabidopsis* has only four 2S albumin genes. It was reasoned that this system offered an opportunity to examine in detail expression of all the members of a seed storage protein gene family. The four genes are located in tandem over a 10-kb genomic region. Previous work has shown that at least one of the four genes is expressed. In this work it is demonstrated that all four genes are expressed, but at quantitatively different levels. In situ hybridizations with gene-specific probes show that the genes differ in the cell types in which they are expressed. Finally, transgenic *Arabidopsis* plants are used to show that the introduction of an "extra" member of the gene family has little or no effect on the expression of the original members.

RESULTS

Gene-Specific Probes for the *at2S* Genes

To follow the expression of the four *Arabidopsis* 2S albumin genes individually, gene-specific probes were made by subcloning fragments from the 3′ ends of the four genes into vectors allowing in vitro RNA synthesis from both strands. To verify that the probes were specific during RNA-RNA as well as DNA-DNA hybridizations, unlabeled sense strand RNA was made from each probe and hybridized with labeled antisense RNA. Figure 1 shows that even when an excess of unlabeled sense strand RNA is applied to the filter, hybridization is gene specific. It was further verified that the probes did not hybridize to other regions of the genes (data not shown).

All Four *at2S* Genes Are Expressed in a Temporally Regulated Fashion

To determine whether all four genes are expressed and to compare their expression throughout development, RNA was prepared from seedpods in different stages of development. Five stages of seedpod maturation were defined. Stage A consists of flowers; it is not possible to differentiate pollinated and unpollinated flowers. Seedpods of less than 5 mm long were defined as being in stage B (approximately 2 days to 4 days after pollination, DAP); those of 5 mm to 10 mm as stage C (4 DAP to 6 DAP); those longer than 10 mm, but still green, stage D (7 DAP to 9 DAP); and seedpods that were just beginning to turn yellow and in which seeds were visible, stage E (10 or more DAP). These stage definitions are necessarily arbitrary divisions of a continuum and may not be applicable to all varieties of *Arabidopsis*. Up to 50 individual pods were used for each extraction.

Initially, the four gene-specific probes were used in an RNA gel blot against RNA isolated from the five stages as well as from leaves, as shown in Figure 2. All four probes gave signals in seed RNA from stages C to E, but the signals varied in intensity. No expression was detected in leaves (slot L) or flowers (data not shown). To quantitate the relative expression of the four genes throughout development, RNA from each stage was applied to slot blots, and the filters were hybridized with in vitro synthesized labeled RNA from the four probes. No expression of any of the genes was detected in either stage A or stage B. RNA from all four genes was detected in increasing amounts through stages C, D, and E, as illustrated in Figure 3. In dried seed, RNA levels dropped again (data not shown). The relative amounts of RNA derived from each gene were not identical. The mRNAs from genes 2 and 3 were always present at statistically significant (Friedman classification test; Friedman, 1937) greater amounts

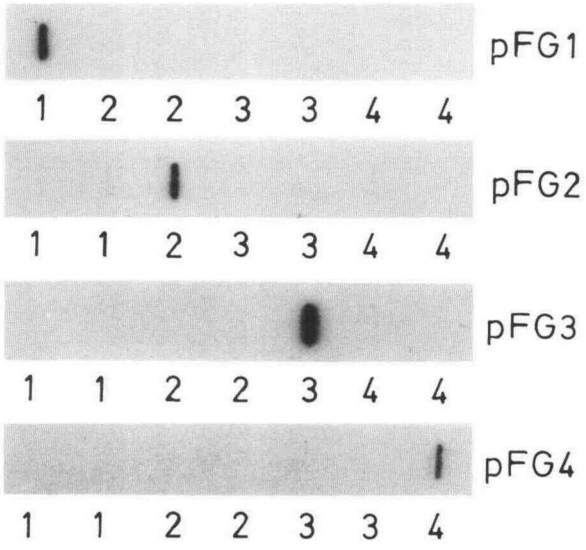

Figure 1. Four Gene-Specific *at2S* Probes.

Each of four filters was loaded pairwise with 10 ng and 1 ng of unlabeled sense strand RNA made from the four *at2S* probes (labeled 1 to 4 under the filter), except that the 10-ng aliquot was omitted for the homologous probe for each filter. The filters were then hybridized with the antisense strand, ^{32}P-labeled riboprobe indicated to the right of each filter. See Methods for the hybridization conditions.

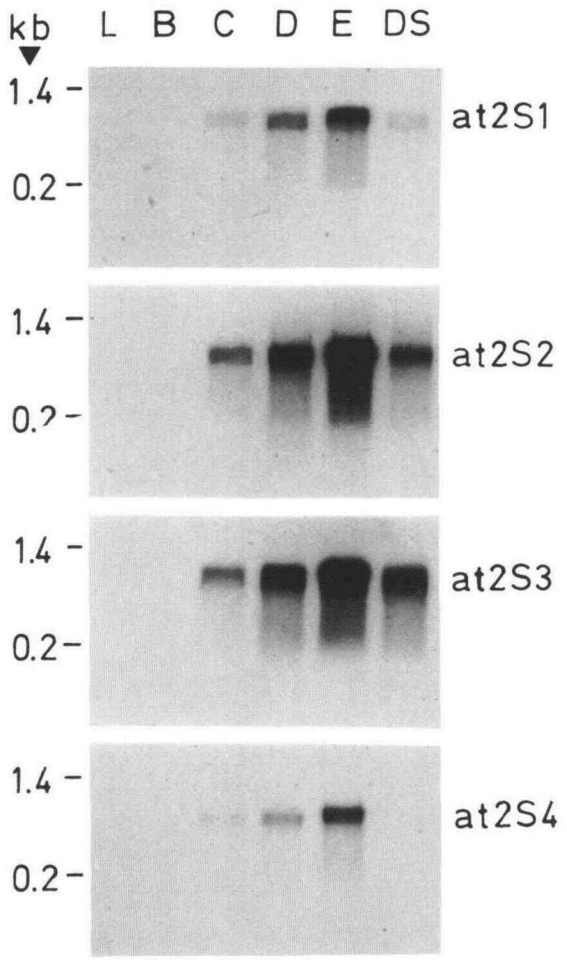

Figure 2. RNA Gel Blot Analysis of Total *Arabidopsis* Seed RNA Using *at2S* Gene-Specific Probes.

Total RNA from leaves (L), developing seedpods of stages B, C, D, and E, and desiccating seeds (DS) was separated on a denaturing gel and blotted, and the resulting filters were hybridized with riboprobes specific for the gene indicated to the right of each autoradiograph. The numbers on the left show the position of RNA size markers, indicated in kilobases.

than those from genes 1 or 4. With minor variations, the ratios of the steady-state levels of mRNA from each gene relative to the others remained approximately constant throughout development (gene 2 > gene 3 > gene 1 ≥ gene 4), even as the total amount of mRNA rose. Very similar results were obtained when two other *A. thaliana* lines, Columbia and K85, were tested (data not shown). These results demonstrated that all four *Arabidopsis* 2S albumin genes were expressed and followed similar temporal profiles. However, there were significant differences in their levels of expression.

The Four *at2S* Genes Are Differently Expressed in Different Tissues

The slot-blot techniques used in the experiments allowed careful quantitation, but because of the amount of tissue needed to isolate sufficient quantities of RNA, it was necessary to pool seedpods; thus, one cannot be certain that the time course obtained was accurate. This problem can be overcome in part using in situ hybridization techniques in which single seeds can be examined, albeit not in a quantitative fashion. This technique also allows study of the tissue and cell type specificity of expression. A single inflorescence was labeled on each of 10 individual plants, and the flowers developing on that inflorescence were followed day by day. The day of pollination was defined as the day the stigma extruded from the corolla. Newly pollinated flowers were labeled for 2 weeks so that at the end of that period, flowers and seedpods of varying, defined ages could be isolated. Preference was given to those inflorescences that produced new flowers on the most regular basis so that sections could be made wherever possible on pods from the same inflorescence.

Because of the difficulty in obtaining histological sections from tissues with a high water content, only sections from pods older than 6 days (the end of stage C) could be analyzed. As expected, Figure 4 shows that genes *at2S2* and *at2S3* were expressed strongly in all tissues. The level of expression was already high at day 7, and remained

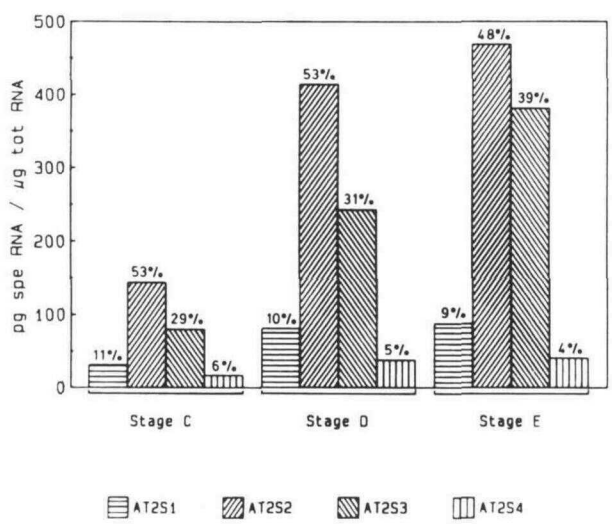

Figure 3. Quantitative Analysis of the *at2S* Genes Throughout Development.

Total seedpod RNA from the indicated stages was analyzed using slot-blot analysis with the four gene-specific probes. The figures at the top of each bar indicate the percentage of total 2S albumin RNA represented by that gene at that stage.

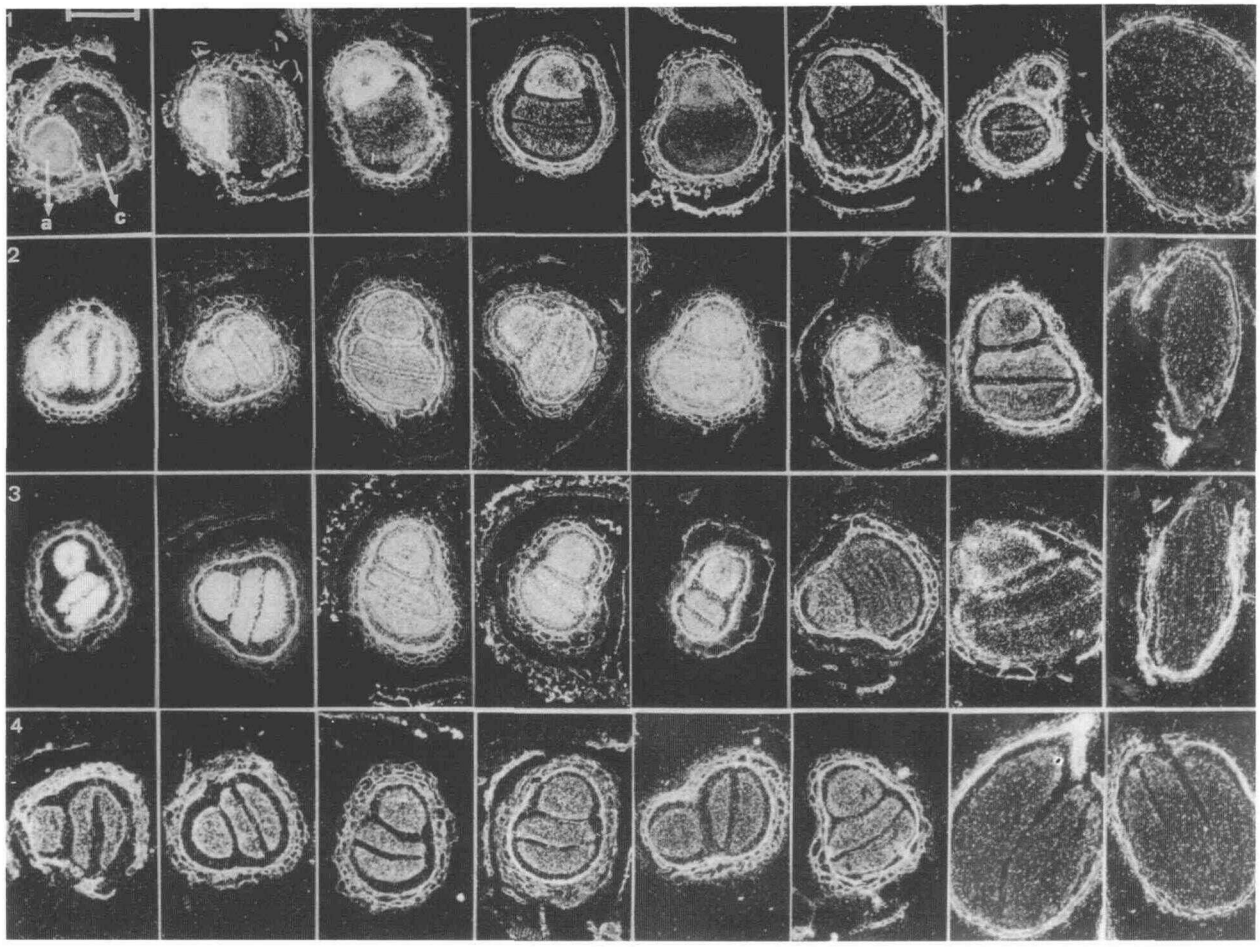

Figure 4. In Situ Analysis of *at2S* Gene Expression.

Rows 1 through 4 show cross-sections of *A. thaliana* embryos hybridized with probes specific for messages derived from genes *at2S*1, *at2S*2, *at2S*3, and *at2S*4, respectively. The embryos depicted in the first column are 7 days old; the age increases by 1 day per column. The "a" indicates the embryo axis; the "c" the cotyledon. The scale bar represents 100 μm; the enlargement is 80-fold. The white grains represent hybridization of the gene-specific probe with mRNA. The light color observed in some sections around the seed coat is an artifact of the use of dark-field microscopy.

high at least through day 11, falling off thereafter as the seed ripened. RNA from *at2S*2 appeared to be present at least 1 day longer than that of the other three genes. Similarly, *at2S*4 was expressed in all tissues and, as predicted by the slot-blot analysis, the level of expression appeared to be less than that of *at2S*2 or *at2S*3. (The probes were adjusted to approximately the same specific activity.) In contrast, *at2S*1 showed a markedly different pattern of expression. It was expressed in the embryo axis at levels which, within the quantitative sensitivity of the technique, appeared similar to the level of expression of *at2S*2 or *at2S*3. However, it appeared to be expressed at only very low levels, if at all, in the cotyledons. In this case, it appears that the slot blots (Figure 3) gave a misleading picture; the apparent overall low level of expression was caused by a difference in expression pattern rather than low levels of expression throughout the embryo, as for *at2S*4. One can conclude that *at2S*1 is differentially regulated with regard to tissue specificity relative to the other three genes and that the four genes do not follow exactly the same temporal expression profile.

Analysis of *at2S* Expression in Transgenic *Arabidopsis* Plants Carrying Modified 2S Albumin Genes

We were interested to determine the effect, if any, of introducing another gene driven by one of the *at2S* gene

promoters. Two questions were asked: Would the relative levels of expression of the four endogenous genes be altered? What would be the effect on the total pool of 2S albumin encoding mRNA? The answers to these questions were approached using a series of transgenic plants into which three different modified *at2S1* genes were introduced. All three modifications were made in the region encoding the large subunit. Thus, the modified genes had the same 3′ end as the endogenous genes, and the same probe used to detect mRNA from *at2S1* itself (pFG1) would detect mRNA from the introduced genes. For one of the three modifications (expressed in the series of plants designated A20), a separate probe was available complementary to the modified region. This probe could be used specifically to detect expression of the introduced gene. The copy number of the introduced gene was usually one, except as noted in the figure legends (data not shown).

Total seedpod RNA was prepared from 35 independent transformants and used for slot-blot analysis. Because the level of expression of the four genes relative to each other was constant throughout development, no attempt was made to isolate RNA from specific stages; therefore, the variations in total amount of 2S albumin RNA detected are not significant. Figures 5A to 5C show data for 18 of the 35 individual plants tested. As in the wild-type plants, mRNAs complementary to *at2S2* and *at2S3* were the most abundant, whereas *at2S4* made the smallest contribution to the 2S albumin message pool. (Note that if the total amount of mRNA were higher, one would expect the individual contributions to be slightly lower.) The percentage of 2S albumin mRNA complementary to pFG1 was significantly increased in all of the transgenic plants, as would be expected if pFG1 detects mRNA produced both by the endogenous *at2S1* and the introduced gene. No correlation was observed with the copy number of the introduced gene; this is presumably lost in the normal variation in expression of such genes (Jones et al., 1985; De Almeida et al., 1989). In plant series A20, a fifth probe was used to detect mRNA complementary to the introduced gene (Figure 5C). In the six plants tested in this way, the percentage of 2S albumin mRNA derived from the introduced gene ranged from 6% to 21%. The ratio of mRNA from the introduced gene to mRNA complementary to pFG1 (i.e., both the introduced gene and endogenous *at2S1*) was in several cases higher than one would expect, suggesting that the mRNA levels for the introduced gene may be overestimated in these experiments. This finding may be related to technical difficulties in accurately determining the concentration of the unlabeled RNA used as a standard (see Methods). Overall, the results were consistent with a model in which the expression levels of the four endogenous genes relative to each other are unchanged.

Because of variations in individual RNA preparations and the normal variation in expression levels of introduced genes in transgenic plants (Jones et al., 1985), it was difficult to determine whether the total amount of 2S albumin mRNA was increased in transgenic plants simply by adding up the absolute amounts of RNA. Instead, the ratio of the total mRNA contributed by *at2S2*, *at2S3*, and *at2S4* to that contributed by *at2S1* and, in transgenic

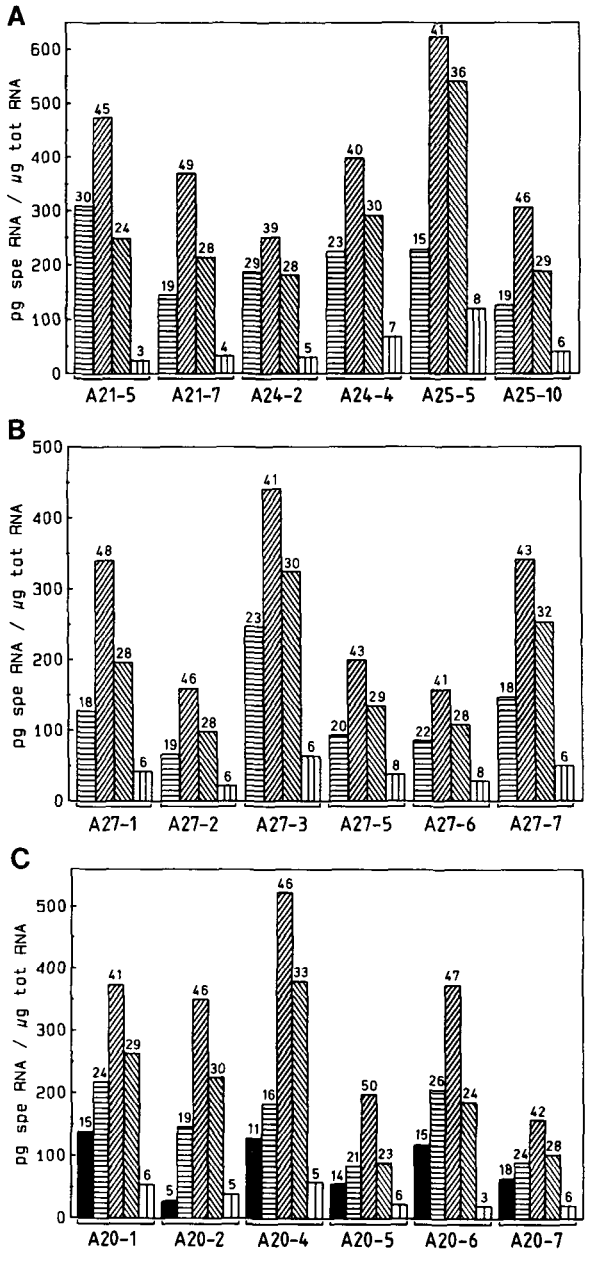

Figure 5. Quantitative Analysis of *at2S* Gene Expression in Transgenic Plants.

(A) Total seedpod RNA from individual transgenic plants (indicated by the numbers under each group of graphs) was analyzed as in Figure 3. Labels are as in Figure 3, except that the probe specific for *at2S1* detects mRNA from both the endogenous and introduced gene(s), which are regulated by the promoter of *at2S1*. The copy number of the introduced gene was 1, with the exception of A21-7 (2 copies) and A21-5 (3 copies).

(B) As in **(A)**, except that data for different individual plants are shown. The copy number of the introduced gene was 1, with the exception of A27-5 (1 or 2 copies) and A27-3 (not determined).

(C) As in **(A)**, except that data for another set of individual plants are shown. In this case the solid bar represents RNA produced by the introduced gene only. The copy number of the introduced gene was 1, with the exceptions of A20-1, A20-2, and A20-7 (2 copies) and A20-4 (3 copies).

plants, the introduced gene, were plotted as shown in Figure 5. The dashed line represents a linear regression of data derived from three separate quantitation experiments using mRNA from different stages of maturation in wild-type plants, and the solid line represents similar data derived from 35 individual transgenic plants. The slope of the line representing wild-type plants is 5.62 (±0.67; correlation coefficient 0.95), whereas that of transgenic plants is 3.16 (±0.36; correlation coefficient 0.84). The difference in these slopes is statistically significant (Student's *t* test, 40 *df*, P = 0.01), indicating that the ratio of mRNA from genes *at2S2* to *at2S4* to that complementary to pFG1 has changed.

A change in this ratio could be brought about either by an increase in the relative steady-state level of mRNA complementary to pFG1 (i.e., mRNA from *at2S1* and the introduced gene) relative to the sum of the mRNA from *at2S2* to *at2S4*, a decrease in the latter relative to the former, or some combination of the two. All three of these models could be consistent with an increase, a decrease, or no change in the total amount of 2S albumin mRNA. Because of variations between individual transgenic plants and individual RNA preparations, it is very difficult to compare absolute amounts of RNA. It was hoped that by pooling the data from the large numbers of plants analyzed in these experiments, some of these factors would be mitigated and allow an approximation. Table 1 presents the average amount of RNA complementary to each of the genes in all of the wild-type and transgenic plants analyzed. Remarkably, the values for *at2S2* to *at2S4* were very similar in wild-type and transgenic plants, whereas that for *at2S1* increased by 93% in transgenic plants. These results are consistent with the interpretation that the total amount of mRNA produced by genes regulated by the *at2S1* promoter has increased in transgenic plants and that the expression of genes *at2S2* to *at2S4* is not affected.

DISCUSSION

In this work the *A. thaliana* 2S albumin genes were used as a model to study interrelationships among all the members of a seed storage protein multigene family. Quantitative analysis of RNA levels and in situ analysis using gene-specific probes demonstrated that the four closely linked *at2S* genes are not regulated in similar ways, indicating that, like more complex gene families studied in other species, the *Arabidopsis at2S* gene family, while encoding very similar proteins, expresses them at different levels in different tissues.

The initial quantitative analysis of steady-state RNA levels throughout development demonstrated that all four *at2S* genes were expressed. All four RNAs appeared to increase in level after similar temporal profiles so that at any given stage the relative contribution of each of the genes to the total *at2S* RNA pool was the same, within the precision possible with the techniques used. In situ analysis, in which individual seeds can be examined, suggested that this picture is oversimplified; RNA corresponding to *at2S2* persisted a day longer than that of the others. *at2S2* and *at2S3* contributed the bulk of the mRNA (40% to 55% and 35% to 45%, respectively), whereas *at2S1* and *at2S4* provided the remainder, usually from 5% to 10% each. In most RNA preparations, *at2S1* mRNA was present in equal or greater quantities than mRNA from *at2S4*. This dichotomy of relative expression levels is similar to that observed in other gene families. For example, gene SSU301 of the eight-member petunia *rbcS* gene family gives 47% of the total *rbcS* mRNA, whereas five other genes provide the remainder (Dean et al., 1985). In the leaves of tomato, *rbcS* 3B and 3C together provide 60% of *rbcS* mRNA; three other genes provide the remainder (Sugita and Gruissem, 1987).

In addition to quantitative differences, individual members of gene families may follow different developmental patterns. The in situ hybridization results (Figure 4) demonstrated that tissue specificity differed among the four genes. In particular, whereas *at2S2* to *at2S4* were expressed throughout the embryo, *at2S1* was expressed only in the embryo axis and not in the cotyledons at least after day 6 post-pollination. (Analysis at earlier timepoints was not possible in these experiments, but the slot-blot analysis suggested that little or no 2S albumin expression took place before stage C, 4 DAP to 6 DAP.) The differences in tissue specificity and expression levels suggested that the four genes, although highly homologous in their protein coding regions, differ in their *cis*-acting regulatory elements. Comparison of the 5′-flanking sequences of the four genes (Krebbers et al., 1988) shows significant homologies up to about −210 (relative to the initiation codon), but the functional significance of the differences found in

Table 1. RNA Levels in Wild-Type and Transgenic Plants

Plant	*at2S1*	*at2S2*	*at2S3*	*at2S4*
	pg specific RNA/μg total seed RNA			
Wild-type	89	369	255	69
Transformed	172	361	248	54

The mean levels of RNA corresponding to each gene in wild-type and transformed *Arabidopsis* plants are shown. In transgenic plants, the amount in column 1 includes RNA from both *at2S1* and introduced genes under the regulation of the promoter of *at2S1*. The data for the wild-type plants are derived from nine different RNA preparations from seedpods of pooled plant. Each determination was repeated three times. The data for transformed plants are derived from single RNA preparations from 35 individual transformants. In these cases, each determination was done once. The difference between the two values in column 1 is statistically significant (Student's *t* test, 42 *df*, P = 0.01).

this region and the mapping of *cis*-acting elements located further upstream await molecular analysis of the regions using deletion studies and DNA-protein binding studies. The evolutionary advantage of having a gene expressed only in the embryo axis while two other genes are already highly expressed there as well as in the cotyledons is not clear.

Similar results are observed in some, but not all, other plant gene families. In petunia the relative level of expression of the eight *rbcS* genes does not vary significantly in the different organs in which these genes are expressed (Dean et al., 1985). Similar results have been reported in pea (Fluhr et al., 1986), but in this case individual genes responded differently to light quality. In tomato both light response and organ distribution differ between *rbcS* genes (Sugita and Gruissem, 1987); comparable results have been reported for the chlorophyll *a/b* binding protein genes in maize (Sheen and Bogorad, 1986; Sullivan et al., 1989).

Three of 10 members of the Kunitz trypsin inhibitor (KTi) genes of soybean (Jofuku and Goldberg, 1989; Perez-Grau and Goldberg, 1989) have also been examined. The KTi3 gene was the most abundant form found in seeds. All three genes were also expressed in other tissues, but KTi3 differed from KTi1 and KTi2 in its organ distribution. As in the present study, it was concluded that different members of the gene family have different combinations of *cis*-acting control elements. The KTi genes differ from the 2S albumin genes in their expression at significant levels in tissues other than seeds. Both RNA gel blot analysis (Figure 2 and Krebbers et al., 1988) and expression of reporter genes using the *at2S1* promoter in transgenic plants (De Clercq et al., 1990) suggest that, with the possible exception of an extremely low level of expression in leaves, the 2S albumins are expressed exclusively in seeds. Similarly, the two isozymes of sucrose synthase have been observed to be expressed in different tissues of the developing maize kernel (Heinlein and Starlinger, 1989); sucrose synthase 1 is expressed only in the endosperm, whereas form 2 is expressed throughout the kernel, with expression levels depending on the cell type and developmental stage. Whether such differences are achieved through the use of different *cis*-acting elements or the addition of specific elements for each tissue or cell type is not yet understood. Harada et al. (1989) studied the 15-member soybean β-conglycinin gene family, and showed that it produces two classes of mRNA whose differential accumulation depends at least in part on post-transcriptional effects. Ladin et al. (1987) reported that the closely related genes encoding the three subunits of β-conglycinin are differentially regulated, being differently distributed between soybean cotyledons and axis. These differences were at least in part post-translational, the β subunit being synthesized but degraded in the soybean embryo axis.

There is widespread interest in the expression of seed storage protein genes, some of them modified, in transgenic plants. It is not clear how, if at all, such introduced genes will interact with endogenous storage protein genes. If all genes of this class compete for the same *trans*-acting factors, one can envision that there will be a limit to the total amount of seed storage protein encoding mRNA that can be made. It is probable, given the complex patterns of differential expression observed in different systems, that individual genes interact with a variety of such factors, and so far there is no evidence that these factors may be limiting. The small 2S albumin gene family and the availability of gene-specific probes make it a promising system to approach this question. In an initial attempt to do so, the steady-state mRNA levels of all four genes, plus an introduced gene regulated by the *at2S1* promoter, were studied. Figure 6 shows that the ratio of the mRNA produced by *at2S2* to *at2S4* to that produced by the *at2S1* promoter (both the endogenous and introduced genes) changed. This means that the level of the latter increased, the former decreased, or both. To differentiate among these possibilities requires looking at absolute versus relative RNA levels, but excludes a model in which there is an absolute limit (at the level observed in wild-type plants) on the amount of RNA that *at2S1* can produce and in which this limit is independent of the expression of the other three. (In such a model, the slopes would be identical.)

Because of variations between individual transgenic plants and RNA preparations, the comparison of absolute RNA levels is difficult, and such results should be regarded

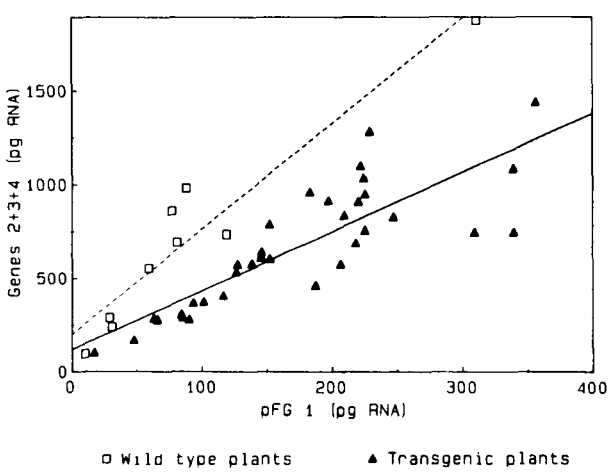

Figure 6. Comparison of Relative Expression Levels of Different Genes in Wild-Type and Transgenic Plants.

Linear regressions of the expression levels of genes driven by the *at2S1* promoter versus the sum of the levels of *at2S2*, *at2S3*, and *at2S4* in wild-type and transgenic *Arabidopsis* plants. Open squares, wild-type plants; filled triangles, transgenic plants. See text for determination of slopes.

with caution. With this in mind, the data of Table 1, which contains average values of a large number of plants, suggest that the level of *at2S2* to *at2S4* RNA remains approximately constant (relative to wild-type plants) in transgenic plants carrying "extra" genes driven by the promoter of *at2S1*, whereas the total amount of RNA from genes regulated by the *at2S1* promoter is increased. This observation is consistent with a model in which *at2S1* is expressed independently of the other three genes and in which there is no overall limitation on the amount of 2S albumin RNA that can be produced, or at least that this limitation has not yet been reached at the copy numbers of introduced genes in these experiments. It is not clear whether the result would be the same if 10 or 20 genes were introduced.

It would be more difficult to draw conclusions from the magnitude of the average increase in *at2S1* promoter-produced mRNA. In principle, if the introduced gene were expressed independently of its endogenous homolog, a plant with one copy of the introduced gene would have a 100% increase in RNA produced by *at2S1* promoters. This is approximately what is observed in the limited data presented here, but in some cases the copy number of the introduced gene was greater than one. In any case, the situation is far more complex; levels of gene expression are likely to be influenced not only by local *cis*-acting elements but also by chromosomal environment, activity of nearby genes, etc. Analysis of such questions would require the ability to insert genes at a known location, in this case preferably adjacent to the *Arabidopsis at2S* locus, at which all four genes are present in a tandem array (Krebbers et al., 1988).

METHODS

Construction of Gene-Specific Probes

Four fragments from the 3′ ends of genes *at2S1*, *at2S2*, *at2S3*, and *at2S4* were cloned into pGEM1 (Promega Biotec). The clones contained the following fragments (numbers refer to coordinants of the sequences for each of the genes in the EMBL data bank): pFG1, FokI (1404)-HindIII (1548); pFG2, FokI (850)-HpaI (943); pFG3, HinfI (707)-HindIII (just beyond end of sequenced region); pFG4, Sau3aI (973)-HpaI (1041). pEK9 contains sequences homologous to the modified region of one of the introduced genes, and no *at2S1* sequences, in the same vector. Recombinant DNA techniques were carried out as described (Maniatis et al., 1982).

The specificity of the probes under RNA-RNA hybridization conditions was verified using slot blots (Williams and Mason, 1985). One nanogram and 10 ng of sense strand unlabeled RNA of each probe (only 1 ng of the homologous probe) were applied to four separate filters, and each filter was hybridized with labeled antisense riboprobes (Melton et al., 1984). All hybridizations were carried out in 50% formamide, 5 × SSPE, 5× Denhardt's solution, and 0.1 mg/mL sonicated herring sperm DNA at 65°C overnight. The filters were washed at 65°C in three 15-min steps in 0.1% SDS and decreasing salt concentrations (5×, 2×, and 0.1× SSPE). To avoid cross-hybridization between the probes, it was necessary to purify the labeled probes on a 12% denaturing polyacrylamide gel (Maniatis et al., 1982). After autoradiography the bands corresponding to full-size RNA were cut from the gel and eluted for 4 hr at 65°C in 500 mM NH_4OAc, 10 mM MgOAc, 1 mM EDTA, and 0.1% SDS.

Plant Material

The *Arabidopsis* line C24 (Valvekens et al., 1988) was used for most of the quantitative analysis, the in situ analysis, and plant transformation experiments. Quantitative analysis was also carried out using the Columbia and K85 (Peleman et al., 1989) lines. Plants were grown under greenhouse conditions.

RNA Isolation, RNA Gel Blot, and Slot-Blot Analysis

Total RNA from seedpods (Beachy et al., 1985) or other tissues (Dean et al., 1985) was separated on 1.2% formaldehyde-agarose gels (Thomas, 1980), using an RNA ladder (Bethesda Research Laboratories) as size marker, and transferred to nylon membranes (Hybond N, Amersham, Amersham, United Kingdom). Hybridizations with the four gene-specific riboprobes were carried out as for the slot blots.

The amount of RNA corresponding to each gene was determined using slot blots. Standard curves were generated by hybridization of the antisense probes to a dilution series of the corresponding sense strand RNA. Each sample was adjusted to a total of 1 μg of RNA using total tobacco leaf RNA. For each of the four probes, a filter was made carrying the appropriate standard curve dilution series and hybridized together with a filter containing 1 μg of total denatured RNA from each of the preparations. After hybridization and autoradiography, each portion of filter corresponding to a slot was cut out and counted in a scintillation counter using scintillation fluid (Packard, Groningen, The Netherlands). Standard curves resulting from linear regression always yielded coefficients of correlation of 0.98 or better. Although the probes were usually of approximately the same activity, note that the standard curves are defined for each probe individually, and, thus, the data are not dependent on the specific activity of the probes. Each RNA preparation was examined at least three times, using different probe preparations and filters. The standard deviations were less than 20% of the mean values. Although RNA from transgenic plants was usually analyzed twice, it was occasionally analyzed only once if material was limiting.

In Situ Hybridizations

^{35}S-labeled riboprobes were prepared as described for the ^{32}P-riboprobes, using ^{35}S-UTP (1160 Ci/mmol, Amersham). The specific activity was 8.6×10^8 dpm/μg. *Arabidopsis thaliana* seedpods were harvested and fixed for 2 hr in 1% glutaraldehyde in 0.05 M cacodylate buffer, pH 7. The tissue was then embedded in paraplast (Chadwick and McGinnis, 1987). Sections of 10 μm were made with a Microtome 2050 Supercut (Reichert-Jung, Nussloch,

Federal Republic of Germany) and pretreated according to Cox et al. (1984). Hybridization of the probes to the slides was performed essentially as described by Barker et al. (1988). Duplicate sections were hybridized with the four gene-specific probes. Autoradiography and developing were as described by Angerer and Angerer (1981).

Transgenic Plants

The transgenic plants used in this study carried chimeric genes in which a portion of the region encoding the large subunit was replaced by sequences designed to alter the amino acid composition of the protein or to encode a biological peptide. The 5′- and 3′-flanking sequences, as well as the majority of the coding sequence, were, thus, identical to the endogenous *at2S1* gene. The constructions were all analogous to those described by Vandekerckhove et al. (1989), and the plant vectors and transformation techniques were identical to those described by Vandekerckhove et al. (1989). Data concerning the expression of the modified proteins will be published elsewhere (A. DeClercq, M. Vandewiele, J. Van Damme, P. Guerche, M. Van Montagu, J. Vandekerckhove, and E. Krebbers, manuscript submitted).

ACKNOWLEDGMENTS

P.G. and F.G.D.S. thank members of both groups for their hospitality and assistance during their stays in Gent. We thank Jack Okamuro, Diane Jofuku, and Gilbert Engler for helpful discussions and communication of results before publication; Georges Pelletier for assistance with statistical analysis; Ian Small, Johan Botterman, and Titti Mariani for critical reading of the manuscript; and Jan Leemans for support and encouragement. P.G. was supported by a grant from the Institut National de la Recherche Agronomique (INRA, France), C.T. by a fellowship from the Nationale Fonds Voor Wettenschappelijk Onderzoek (NFWO, Belgium), F.G.D.S. by a grant from the Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA, Brazil), and A.D.C. by a grant from the Instituut Voor de Aanmoediging Van Wettenschappelijk Onderzoek in Het Nijverheid en Landbouw (IWONL, Belgium).

Received January 24, 1990; revised March 14, 1990.

REFERENCES

Angerer, L.M., and Angerer, R.C. (1981). Detection of poly A^+ RNA in sea urchin eggs and embryos by quantitative in situ hybridization. Nucl. Acids Res. **9,** 2819–2840.

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., Chen, Z.-L., Horsch, R.B., Rogers, S.G., Hoffmann, 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.

Chadwick, R., and McGinnis, W. (1987). Temporal and spatial distribution of transcripts from the *Deformed* gene of *Drosophila*. EMBO J. **6,** 779–789.

Cox, K., DeLeon, D.V., Angerer, L.M., and Angerer, R.C. (1984). Detection of mRNAs in sea urchin embryos by in situ hybridization using asymmetric RNA probes. Dev. Biol. **101,** 485–502.

Crouch, M.L., Tenbarge, K.M., Simon, A.E., and Ferl, R. (1983). cDNA clones for *Brassica napus* seed storage proteins: Evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. J. Mol. Appl. Genet. **2,** 273–283.

De Almeida, E.R.P., Gossele, V., Muller, C.G., Dockx, J., Reynaerts, A., Botterman, J., Krebbers, E., and Timko, M.P. (1989). Transgenic expression of two marker genes under the control of an *Arabidopsis rbcS* promoter: Sequences encoding the Rubisco transit peptide increase expression levels. Mol. Gen. Genet. **218,** 78–86.

Dean, C., Pichersky, E., and Dunsmuir, P. (1989). Structure, evolution, and regulation of *RbcS* genes in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **40,** 415–439.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P., and Bedbrook, J. (1985). Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multi-gene family. EMBO J. **4,** 3055–3061.

De Clercq, A., Vandewiele, M., De Rycke, R., Van Damme, J., Van Montagu, M., Krebbers, E., and Vandekerckhove, J. (1990). Expression and processing of an *Arabidopsis* 2S albumin in transgenic tobacco. Plant Physiol. **92,** 899–907.

Fluhr, R., Moses, P., Morelli, G., Coruzzi, G., and Chua, N.-H. (1986). Expression dynamics of the pea *rbcS* multigene family and organ distribution of the transcripts. EMBO J. **5,** 2063–2071.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Statist. Assoc. **32,** 688–689.

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.

Harada, J.J., Barker, S.J., and Goldberg, R.B. (1989). Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. Plant Cell **1,** 415–425.

Heinlein, M., and Starlinger, P. (1989). Tissue- and cell-specific expression of the two sucrose synthase isoenzymes in developing maize kernels. Mol. Gen. Genet. **215,** 441–446.

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

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 **1,** 1079–1093.

Jones, J.D.G., Dunsmuir, P., and Bedbrook, J. (1985). High level expression of introduced chimeric genes in regenerated transformed plants. EMBO J. **4,** 2411–2418.

Josefson, L.G., Lenman, M., Ericson, M.L., and Rask, L. (1987). Structure of a gene encoding the 1.7S storage protein, napin, from *Brassica napus*. J. Biol. Chem. **262,** 12196–12201.

Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans,

J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., and Vandekerckhove, J. (1988). Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol. **87,** 859–866.

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.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. **12,** 7035–7056.

Okamuro, J.K., and Goldberg, R.B. (1989). Regulation of plant gene expression: General principles. In The Biochemistry of Plants, Vol. 15: A Comprehensive Treatise, A. Marcus et al., eds (New York: Academic Press), pp. 1–82.

Peleman, J., Boerjan, W., Engler, G., Seurinck, J., Botterman, J., Alliotte, T., Van Montagu, M., and Inzé, D. (1989). Strong cellular preference in the expression of a housekeeping gene of *Arabidopsis thaliana* encoding S-adenosylmethionine synthetase. Plant Cell **1,** 81–93.

Perez-Grau, L., and Goldberg, R.B. (1989). Soybean seed protein genes are regulated spatially during embryogenesis. Plant Cell **1,** 1095–1109.

Poulsen, C., Fluhr, R., Kauffman, J.M., Boutry, M., and Chua, N.-H. (1986). Characterization of an *rbcS* gene from *Nicotiana plumbaginifolia* and expression of an *rbcS*-CAT chimeric gene in homologous and heterologous nuclear background. Mol. Gen. Genet. **205,** 193–200.

Scofield, S.R., and Crouch, M.L. (1987). Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. J. Biol. Chem. **262,** 12202–12208.

Sheen, J., and Bogorad, L. (1986). Differential expression of six light-harvesting chlorophyll *a/b* binding protein genes in maize leaf cell types. Proc. Natl. Acad. Sci. USA **83,** 7811–7815.

Sugita, M., and Gruissem, W. (1987). Developmental, organ-specific, and light-dependent expression of the tomato ribulose-1,5-bisphosphate carboxylase small subunit gene family. Proc. Natl. Acad. Sci. USA **84,** 7104–7108.

Sullivan, T.D., Chistensen, A.H., and Quail, P.H. (1989). Isolation and characterization of a maize chlorophyll *a/b* binding protein gene that produces high levels of mRNA in the dark. Mol. Gen. Genet. **215,** 431–440.

Thomas, P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77,** 5201–5205.

Tobin, E. (1978). Light regulation of specific mRNA species in *Lemna gibba* L. Proc. Natl. Acad. Sci. USA **75,** 4749–4753.

Valvekens, D., Van Montagu, M., and Van Lijsebettens, M. (1988). *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proc. Natl. Acad. Sci. USA **85,** 5536–5540.

Vandekerckhove, J., Van Damme, J., Van Lijsebettens, M., Botterman, J., De Block, M., Vandewiele, M., De Clercq, A., Leemans, J., Van Montagu, M., and Krebbers, E. (1989). Enkephalins produced in transgenic plants using modified 2S seed storage proteins. Biotechnology **7,** 929–932.

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

Williams, J.G., and Mason, P.J. (1985). Hybridisation in the analysis of RNA. In Nucleic Acid Hybridisation—A Practical Approach, B.D. Hames and S.J. Higgins, eds (Oxford: IRL Press), pp. 139–159.

Youle, R., and Huang, A.H.C. (1981). Occurrence of low molecular weight and high cysteine containing albumin storage proteins in oilseeds of diverse species. Am. J. Bot. **68,** 44–48.