The Plant Cell, Vol. 3, 561–571, June 1991 © 1991 American Society of Plant Physiologists

A Soybean Cell Wall Protein Is Affected by Seed Color Genotype

Jon T. Lindstrom and Lila O. Vodkin[1]

Department of Agronomy, University of Illinois, Urbana, Illinois 61801

The dominant *I* gene inhibits accumulation of anthocyanin pigments in epidermal cells of the soybean seed coat. We compared saline-soluble proteins extracted from developing seed coats and identified a 35-kilodalton protein that was abundant in Richland (genotype *I/I*, yellow) and much reduced in an isogenic mutant line T157 (genotype *i/i*, imperfect black seed coats). We purified the 35-kilodalton protein by a novel procedure using chromatography on insoluble polyvinylpolypyrrolidone. The 35-kilodalton protein was composed primarily of proline, hydroxyproline, valine, tyrosine, and lysine. Three criteria (N-terminal amino acid sequence, amino acid composition, and sequence of a cDNA) proved that the seed coat 35-kilodalton protein was PRP1, a member of a proline-rich gene family expressed in hypocotyls and other soybean tissues. The levels of soluble PRP1 polypeptides and PRP1 mRNA were reduced in young seed coats with the recessive *i/i* genotype. These data demonstrated an unexpected and novel correlation between an anthocyanin gene and the quantitative levels of a specific, developmentally regulated cell wall protein. In contrast, PRP2, a closely related cell wall protein, was synthesized later in seed coat development and was not affected by the genotype of the *I* locus.

INTRODUCTION

The soybean seed coat is composed of three distinct maternal cell types: an epidermal layer of palisade cells, a hypodermis with large hourglass-shaped cells, and a layer of spongy parenchyma that contains vascular bundles (Carlson and Lersten, 1973). In some genotypes, anthocyanin pigments accumulate in vacuoles of the palisade cells as controlled by at least two unlinked loci, *I* and *R* (reviewed in Bernard and Weiss, 1973). The *I* locus (for inhibitor) prevents anthocyanin production in a spatial manner and exists as a series of four alleles with the following phenotypes: *I* results in complete absence of seed coat pigment; i^i limits pigment to the narrow hilum area where seed and pod are attached; i^k restricts pigment to a saddle-shaped region over two-thirds of the seed coat, and *i* results in a completely pigmented seed coat. The color of the pigmented portion of the seed is black if the dominant *R* allele is present and brown in the recessive *r/r* genotype. The gene products of *I* and *R* are not known.

Interestingly, the dominance relations of the four *I* alleles are in the order *I*, i^i, i^k, *i*, where the relative absence of pigment is the dominant phenotype in each case. Most cultivated soybean varieties are homozygous for a dominant form of the *I* gene resulting in a yellow seed coat. However, spontaneous mutations from yellow seed to dark-colored seed with *i/i* genotype arise frequently within highly inbred soybean varieties (Wilcox, 1988). More than 30 of these independent, isogenic pairs of yellow and pigmented soybean lines exist and they provide an excellent genetic resource to investigate the *I* locus.

In this report, we compared seed coat development in the yellow seeded cultivar Richland (*I/I*) with that of a black seeded mutant line T157 (*i/i*) descended from Richland. Unexpectedly, we found that the levels of a highly abundant 35-kD protein and its mRNA were reduced in the mutant seed coats with *i/i* genotype. We purified the 35-kD protein based on its unusual property of binding to insoluble polyvinylpolypyrrolidone (PVPP) and identified it to be PRP1, a member of a proline-rich gene family in soybean (Hong et al., 1987, 1989, 1990). Thus, our results demonstrated direct detection and purification of the PRP1 polypeptide that had previously been identified only by cloning. We showed that PRP1 is approximately 20% hydroxyproline, a post-translationally modified amino acid found in cell wall proteins (Lamport and Northcote, 1960). Two forms of a similar but distinguishable protein, PRP2, have been isolated from the cell walls of soybean hypocotyl

[1] To whom correspondence should be addressed.

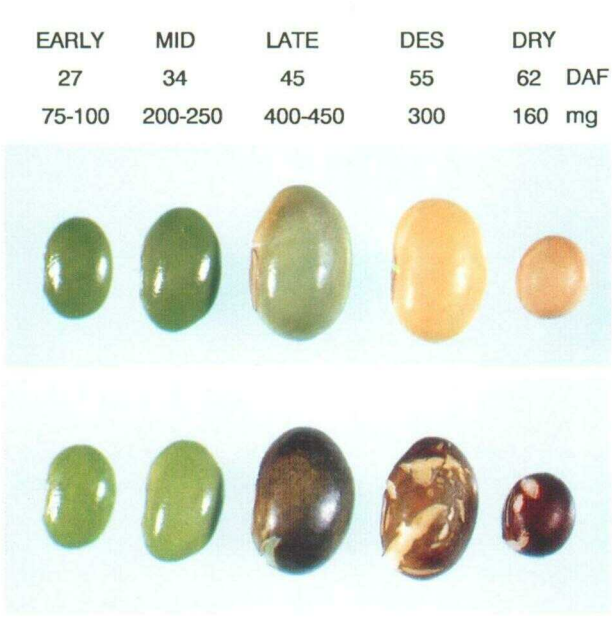

Figure 1. Changes in Seed Coat Color and Structure during Seed Maturation in Isogenic Lines of Soybean.

The top row is Richland with a yellow seed coat (genotype *I*// *R*/*R*/ *t*/*t*), and the bottom row is the spontaneous mutant line T157 (*i*/*i R*/*R*/ *t*/*t*) with the characteristic imperfect black and defective seed coat at maturity. Days after flowering (DAF) and approximate fresh weight of the whole seed are indicated. Des = desiccating seed; Dry = mature harvested seed. Purple anthocyanin pigments begin to accumulate in T157 after approximately 300 mg, fresh weight, and are concomitant with the appearance of cracks in the seed coat structure. The imperfect black and defective seed coat is a result of pleiotropic interaction between the recessive *i* and *t* genes (Bernard and Weiss, 1973), resulting in reduced pigmentation and changes in the types of anthocyanins as compared with black seeded soybean with *i*/*i T*/*T* genotype. The level of cyanidin-3-glucoside pigments is reduced in imperfect black seed coats (Buzzell et al., 1987), but the reason for the structural defects in *i*/*i t*/*t* genotypes is unknown. Although there can be considerable variation in the extent of defects on individual seed of the same plant, the T157 line is highly defective when grown under either field or greenhouse conditions.

hooks and cell cultures (Averyhart-Fullard et al., 1988; Datta et al., 1989; Kleis-San Francisco and Tierney, 1990).

The novel effect of an anthocyanin gene on a hydroxyproline/proline-rich cell wall protein suggests that the *I* locus may influence other cellular functions in the developing seed coat in addition to control of anthocyanin synthesis. The effect of *I* genotype on the PRP1 cell wall protein preceded accumulation of visible anthocyanin pigments. We found that the influence of *I* genotype was specific for PRP1 in young seed coats and did not affect the similar PRP2 protein that was expressed very late in the mature seed coat and during seed desiccation.

RESULTS

A 35-kD Protein Is Affected by Seed Color Genotype

We examined the SDS-PAGE profiles of total proteins extracted from developing seed coats of the cultivar Richland (*I*/*I*, yellow mature seed coat) and its isogenic mutant line, T157 (*i*/*i*, imperfect black and defective mature seed coat). Figure 1 shows the developmental stages and accumulation of anthocyanin pigments during seed maturation in the two cultivars as measured by seed fresh weight and approximate days after flowering (DAF). We detected an unexpected variation in the abundance of a 35-kD protein. Figure 2 shows that the 35-kD protein was abundant in Richland but was absent or considerably reduced in T157 seed coats. Figure 2 also shows that accumulation of the 35-kD protein was developmentally regulated. It was extractable with low-salt buffer from seed coats harvested 21 DAF to 34 DAF (50 mg to 200 mg, fresh weight, of the

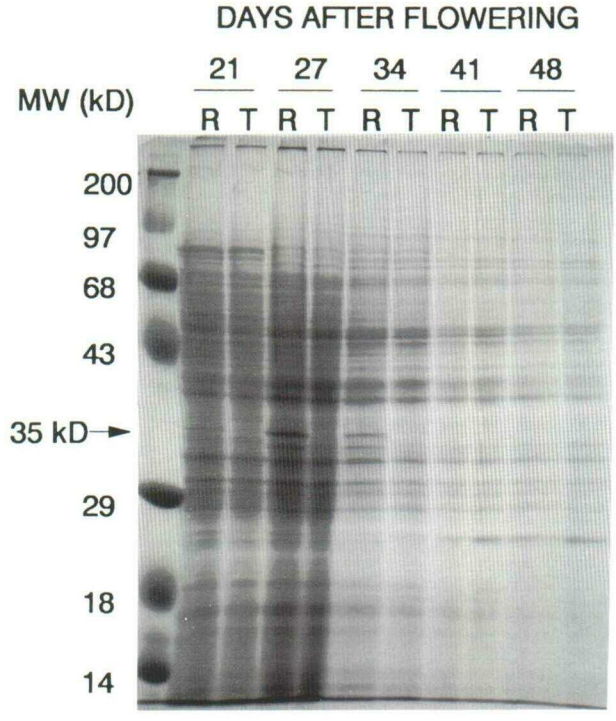

Figure 2. Comparison of Proteins in Developing Seed Coats of Isogenic Lines.

Seed coats were dissected from seed collected at the indicated DAF, extracted with PBS, and analyzed by SDS-PAGE and staining with Coomassie Blue. R = Richland; T = T157. Molecular mass markers (in kilodaltons) are shown in lane 1, and the arrow indicates the 35-kD protein that is more abundant in Richland as compared with T157 seed coats.

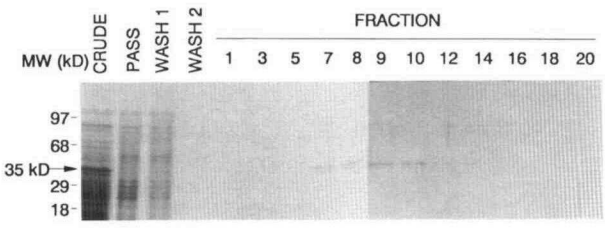

Figure 3. Purification of the 35-kD Protein Using Chromatography on Insoluble PVPP.

An SDS-PAGE gel stained with Coomassie Blue illustrates the affinity of the seed coat 35-kD protein for insoluble PVPP. Details of the protocol are presented in Methods. Molecular mass markers (in kilodaltons) are shown at left. Proteins in the supernatant after the seed coats were ground in PBS are shown in the lane marked "crude." The lane "pass" contains proteins not bound by PVPP after the crude supernatant has flowed through the column. Lanes "wash 1" and "wash 2" are proteins eluted by PBS washes. Fractions 1 to 20 (approximately 150 μL each) were obtained during elution with acetic acid. The 35-kD protein eluted in fractions 7 to 14, immediately after the pH changes.

intact seed) and declined thereafter. At this stage, the embryos and seed coats of both genotypes are green and anthocyanin pigmentation in T157 is not apparent until at least 40 DAF, when the seed are at least 300 mg, fresh weight (Figure 1). Only the 35-kD protein, and a slightly smaller protein of 34 kD, varied between the two cultivars at the level of sensitivity of Coomassie Blue dye staining.

We made crosses between Richland and T157 and extracted proteins from seed coats of the F1 and F2 generations to correlate seed color genotype with the 35-kD protein. Because the seed coat represents the maternal genotype, all seed coats from a single plant have the same color. Therefore, we sacrificed some seed at the 75 mg to 100 mg weight range, scored these for the 35-kD protein, and subsequently scored the color of remaining seed once the plant matured. The immature seed coats on F1 plants (*I*/*i*) contained the 35-kD protein and yielded yellow seed as expected because *I* is dominant. The 35-kD protein was predictive of the mature seed coat color in F2 plants (data not shown). A total of 31 plants with the 35-kD protein yielded yellow seed, versus 11 plants in which the 35-kD protein was not detected. These plants yielded imperfect black and defective seed that are characteristic of the recessive *i*/*i* genotype.

Purification of the 35-kD Protein and Its Identification as PRP1

We purified the 35-kD protein by a novel procedure based on a serendipitous result. The 35-kD protein was extractable in saline buffer containing soluble PVPP, but it did not appear in the crude supernatant fraction if insoluble polymerized PVPP was included in the extraction buffer. We tested whether insoluble PVPP could be used as an affinity matrix to purify the 35-kD protein. Figure 3 shows that "doublet" polypeptides consisting of the 35-kD protein and a less abundant 34-kD protein were bound selectively by the insoluble PVPP column and eluted with 0.5 M acetic acid. The procedure yielded between 25 μg and 50 μg of purified protein from 150 mg of freeze-dried seed coats, corresponding to approximately 30 seed.

The purified 35-kD and 34-kD doublet polypeptides were separated by SDS-PAGE, blotted to Immobilon membrane, and each was subjected separately to N-terminal amino acid analysis. Both the 35-kD and 34-kD sequences were identical and contained hydroxyproline. Figure 4 shows that a search of protein data banks revealed an exact match of the soybean seed coat N-terminal sequence to the mature protein sequence predicted from a soybean cDNA (SbPRP1 or PRP1, soybean proline-rich protein) cloned from auxin-treated suspension culture cells and 4-day-old hypocotyls (Hong et al., 1987; Averyhart-Fullard et al., 1988). Amino acid composition (data not shown) indicated that 5 amino acids were most prevalent in the seed coat protein (proline, hydroxyproline, valine, tyrosine, and lysine), thus suggesting a repeat unit motif similar to PRP1, a putative cell wall protein of the hypocotyl (Hong et al., 1987, 1990).

To confirm the structure, we constructed a soybean seed coat cDNA library using mRNA from Richland seed coats dissected from seeds of 75 mg to 100 mg, fresh weight. The library was screened with a partial clone representing the repeat unit of SbPRP1 from hypocotyls (Averyhart-Fullard et al., 1988). We isolated a near full-length cDNA of the entire coding region that contained a signal sequence and an N-terminal sequence that matched the protein N-terminal data. The seed coat cDNA (pB1-3) corresponded exactly to the SbPRP1 cDNA sequence (Hong et al., 1987) except for an exact duplication that adds five extra repeat units in the coding region of the seed coat cDNA (data not shown). The nature of this discrepancy is unknown and could reflect either errors by

A XXX-TYR-GLU-LYS-PRO-HYP-ILE-TYR-LYS-PRO-HYP-VAL-TYR-THR

B ASP-TYR-GLU-LYS-PRO-PRO-ILE-TYR-LYS-PRO-PRO-VAL-TYR-THR

Figure 4. N-Terminal Sequence of the Seed Coat 35-kD Protein.

(A) The directly determined sequence of the first 14 amino acids that were obtained repeatedly. The first amino acid was ambiguous.

(B) The predicted mature protein N-terminal sequence for an auxin-regulated cDNA (SbPRP1) expressed in soybean suspension culture cells and in 4-day-old soybean hypocotyls as described by Hong et al. (1987).

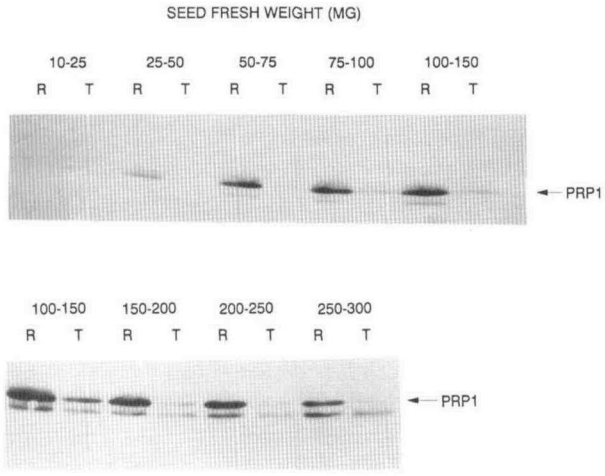

Figure 5. Comparison of PRP1 Levels in Richland and the T157 Mutant.

Seed coats were dissected from seeds with the indicated fresh weight (in milligrams) and extracted with PBS. Equal protein amounts (10 μg) were fractionated by SDS-PAGE, transferred to nitrocellulose, and probed with antibody. R and T identify extracts from Richland and T157, respectively. Because the upper and lower panels represent separate blots, relative intensities should be compared within, but not between, the panels. The antibody used was polyclonal antibody made to a related proline-rich protein, PRP2, isolated from the hypocotyl hook region (Kleis-San Francisco and Tierney, 1990). Monoclonal antibody to purified PRP1 from the seed coat gave identical results (data not shown).

reverse transcriptase during cDNA synthesis or the existence of closely related and duplicated genes. Thus, the seed coat 35-kD and 34-kD proteins are very similar, if not identical, to the protein predicted by the SbPRP1 hypocotyl cDNA sequence. The origin of the doublet is unknown and could represent degradation or post-translational modifications. For brevity, we refer to the seed coat proteins as PRP1.

PRP1 Polypeptide and mRNA Levels Are Regulated Developmentally and Reduced in Mutant Seed Coats

We compared the levels of saline-soluble PRP1 in Richland and T157 by immunoblotting equal amounts of seed coat proteins from various stages of seed development before the onset of anthocyanin accumulation in the T157 line. We weighed seed individually and separated them into nine weight ranges before dissecting the seed coats. Figure 5 shows that PRP1 levels were very low in the early stages of seed coat development (10 mg to 25 mg), rose to maximal levels in seed that were 75 mg to 150 mg, fresh weight, and declined gradually in the larger seed. Figure 5 also illustrates that the less abundant 34-kD species followed the same basic profile except that it appeared to be proportionally more prevalent in seeds larger than 150 mg. The abundance of both the 35-kD and 34-kD species of PRP1 was significantly lower in T157 seed coats as compared with Richland at each stage of development.

Figure 6A shows that levels of PRP1 mRNA were regulated developmentally and they increased and decreased in a pattern similar to the profile for the PRP1 soluble polypeptides. Figures 6B and 6C show two additional independent experiments comparing equal loadings of mRNAs from several stages of development. PRP1 mRNA was less abundant in T157 seed coats. This difference

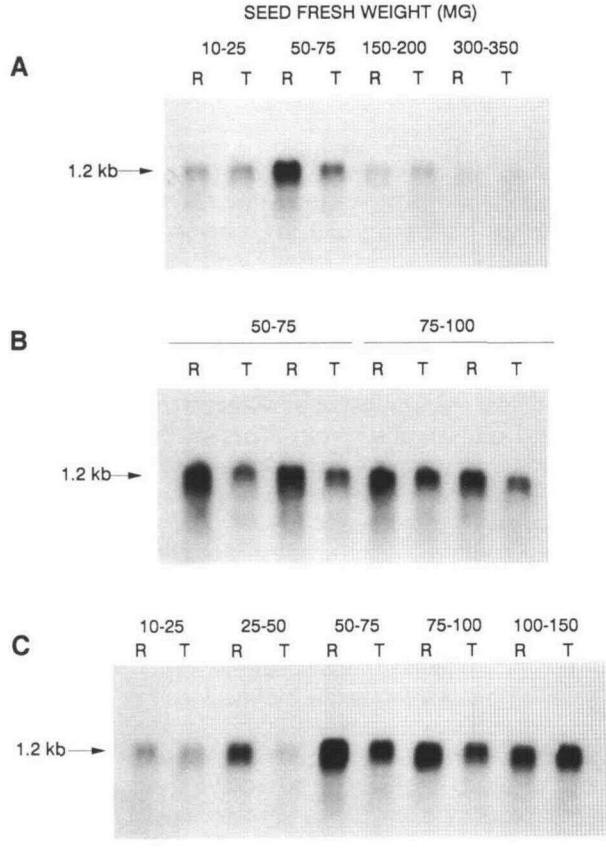

Figure 6. Comparison of PRP1 mRNA Levels in Richland and T157.

(A) Equal loadings of seed coat total RNA (1 μg) from Richland (R) and T157 (T) for each indicated seed fresh weight range (in milligrams) were used and verified by staining intensity. RNA was transferred to nitrocellulose and probed with the seed coat PRP1 cDNA clone (pB1-3).

(B) and **(C)** Additional independent experiments analyzing seed coat RNA obtained in the early weight ranges.

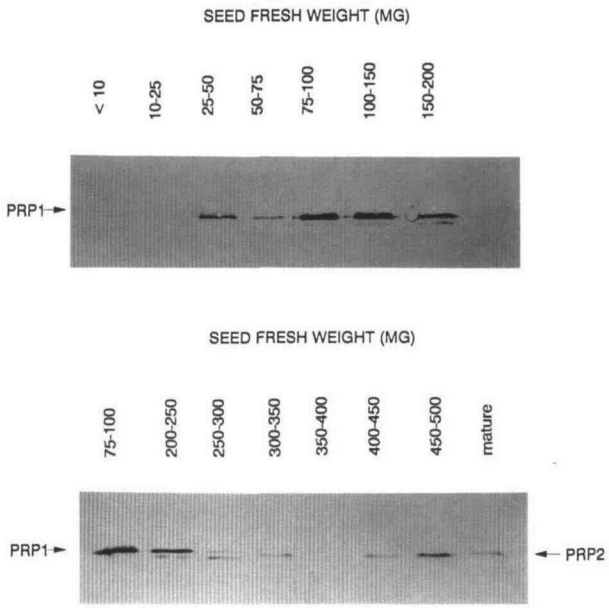

Figure 7. Detection of PRP2 Protein Late in Development of the Seed Coat.

Saline-soluble seed coat proteins were isolated from Richland seed coats taken from a range of seed from less than 10 mg, fresh weight, to mature. Equal protein amounts were fractionated by SDS-PAGE, transferred to nitrocellulose, and probed with monoclonal antibody to purified PRP1. The use of polyclonal antibody to PRP2 results in the same profile (data not shown).

was especially apparent with seed coats from seeds between 50 mg and 100 mg, fresh weight, which was the period of peak mRNA accumulation. The length of the PRP1 mRNA transcripts was approximately 1.2 kb in both cultivars.

PRP2 Polypeptides Are Found in Mature Seed Coats

We extended our examination of seed coat development to late maturation and seed desiccation to determine whether we could detect the related PRP2 protein because its mRNA is found in older seed coats (Hong et al., 1989). Figure 7 shows that the 35-kD PRP1 disappeared at about 300 mg, fresh weight, and only a smaller, less abundant polypeptide persisted to maturity. The smaller protein was most prevalent in the latter stages of maturation when the axis is yellow and the cotyledons have begun to lose color and enter the desiccation stage. This immunoreactive protein generally migrated at about 33 kD, which was slightly smaller than the lower of the PRP1 doublet polypeptides. However, we could not distinguish whether this protein was similar to PRP1 or PRP2 by immunoblotting experiments because antibodies to PRP1 (monoclonal) and PRP2 (polyclonal) gave similar results. Antigenic cross-reactivity between the two proteins is not surprising because of the similar repeat unit motif of both proteins.

PRP1 and PRP2 have different amino-terminal sequences. To identify the 33-kD seed coat protein, we attempted to purify it from older seed coats by affinity isolation on insoluble PVPP. Figure 8 shows that the 33-kD protein did not bind insoluble PVPP using the same procedure that was successful previously for PRP1. A small amount of PRP1 in the older seed coats (400 mg to 500 mg range) was bound by the column (lane 3), but the 33-kD protein passed through the column with all of the other proteins of the crude extract (lane 3). The behavior of the 33-kD protein as compared with PRP1 indicated that it was not a modified form of PRP1 but was PRP2. We then examined mRNA profiles using gene-specific probes. Figure 9 shows that PRP2 mRNA was found in the older seed coats but that PRP1 mRNA was absent or very low. These data confirm that the 33-kD protein in the older seed coats was PRP2. The gene-specific probes also showed the developmental patterns expected for PRP1 and PRP2 synthesis in hypocotyl hook and stem regions (Averyhart-Fullard et al., 1988; Hong et al., 1989).

PRP2 Expression Is Not Affected by the Seed Color Genotype

We found that soluble PRP2 protein was most abundant in older seeds characterized by yellow axes and green cotyledons. Its level was maintained or declined gradually until the cotyledons were completely yellow and almost

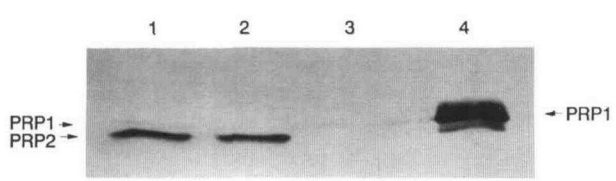

Figure 8. PRP2 Protein Does Not Bind to Insoluble PVPP.

Extracts from older seed coats (400 mg to 500 mg, fresh weight) were passed over a column containing insoluble PVPP in a manner similar to that used to purify PRP1 from young seed coats (Figure 3). Proteins obtained during various steps were fractionated by SDS-PAGE, transferred to nitrocellulose, and probed with antibody to PRP1. Lane 1, saline-soluble crude supernatant fraction; lane 2, unbound proteins of the crude supernatant that have passed through the PVPP column; lane 3, bound proteins eluting with application of 0.5 M acetic acid to the column; lane 4, 0.25 μg of purified PRP1 that is composed of the characteristic doublet 35-kD and 34-kD polypeptides. Note that the PRP2 protein passes through the column (lane 2), whereas the small amount of PRP1 in older seed coats binds to the column (lane 3).

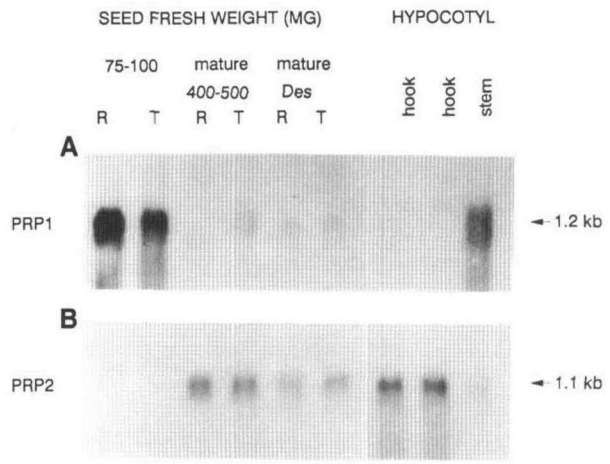

Figure 9. Comparison of PRP2 mRNA Levels in Richland and T157 Seed Coats.

(A) RNA was extracted from Richland (R) and T157 (T) seed coats dissected from seed of the following stages: 75 mg to 100 mg, 400 mg to 500 mg (with yellow axes), and desiccating (Des) seed that have both yellow axes and yellow cotyledons. RNA was also obtained from 5-day-old Richland hypocotyl hooks or mature stems. Equal loadings of RNA (1 μg of seed coat and 5 μg of seedling RNAs) were transferred to nitrocellulose and probed with a labeled gene-specific probe for PRP1.
(B) The same RNA samples were used as in **(A)** but were probed with a gene-specific probe for PRP2.

fully desiccated. Figures 9 and 10 compare the levels of PRP2 mRNA and protein from Richland and T157 seed coats. There appeared to be no significant difference in the PRP2 mRNA and protein levels in the mature seed coats at a time when anthocyanin accumulation was visible in the T157 seed coats. This contrasted markedly with the situation for PRP1, where both mRNA and protein levels were reduced in T157 at a time period that precedes visible anthocyanin accumulation.

DISCUSSION

A Protein That Corresponds to PRP1 Is Isolated from Seed Coats

A significant aspect of our work is the identification and isolation of PRP1 (SbPRP1), a protein whose existence had been predicted from cDNAs isolated from tissue culture cells and germinating seedlings (Hong et al., 1987; Averyhart-Fullard et al., 1988). PRP1 is a member of a proline-rich gene family in soybean consisting of at least three different members: PRP1, PRP2, and PRP3 (Hong et al., 1990). Each family member shows organ-specific and stage-specific developmental expression as determined with RNA dot blots using gene-specific probes (Hong et al., 1989). The presence of their mRNAs has been examined in detail in the hypocotyl hook and stem region of germinating seedlings. PRP2 is known to be a cell wall protein containing hydroxyproline residues (Averyhart-Fullard et al., 1988; Datta et al., 1989; Kleis-San Francisco and Tierney, 1990). It is clear from our amino acid composition data that approximately 50% of the proline residues in PRP1 are also hydroxylated and that these are most likely the second residues of the repeat unit motif (Pro-Hyp-Val-Tyr-Lys) as found in the N-terminal sequence (Figure 4).

Although PRP2 has been purified and characterized from cell cultures and hypocotyl hooks (Averyhart-Fullard et al., 1988; Datta et al., 1989; Kleis-San Francisco and Tierney, 1990), PRP1 has not been isolated previously from any soybean tissue although its mRNA is present in mature hypocotyl stems. It may be that PRP1 is bound to the hypocotyl cell wall at a faster rate than PRP2 during seedling growth and cannot be extracted. Likewise, PRP3 mRNA is synthesized in hypocotyls and seed coats (Hong et al., 1989), but no protein has been detected. In contrast, the seed coat is an extremely good source of PRP1 that can be identified by Coomassie Blue staining of total saline-soluble protein extracts (Figure 2). Other cell wall proteins cannot be detected as easily in crude extracts.

The seed coat PRP1 has a predicted molecular mass of 29 kD (excluding signal sequence) based on cDNA structure and consists of 48 repeat units. Its apparent higher molecular mass of 35 kD using SDS-PAGE is not unusual for proteins that are rich in proline and hydroxyproline. Similarly, the apparent molecular mass for PRP2 is 33 kD, yet its sequence predicts a 24-kD protein (Datta et al.,

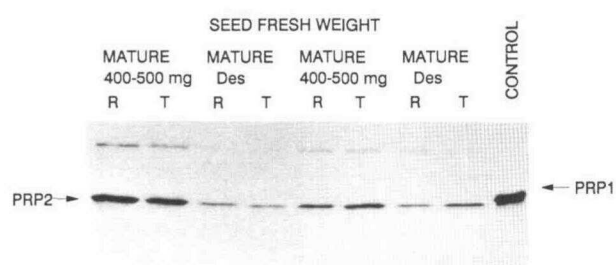

Figure 10. Comparison of PRP2 Protein Levels in Richland and T157 Seed Coats.

Saline-soluble proteins were extracted from mature seed coats obtained from seed of the indicated stages (see Figure 9 legend). Proteins (10 μg) were fractionated by SDS-PAGE, transferred to nitrocellulose, and probed with antibody to PRP1. Two independent extractions are shown for each weight range. The control lane contains 0.25 μg of purified PRP1.

1989; Kleis-San Francisco and Tierney, 1990). The repeats unit of PRP2 is basically a decamer of Pro-Hyp-Val-Tyr-Lys-Pro-Hyp-Val-Glu-Lys. It is lower in tyrosine and higher in glutamine than PRP1 because some of the tyrosines are replaced with glutamic acid residues.

PRP1, but Not PRP2, Binds to Insoluble PVPP

A novel procedure developed to purify and identify PRP1 from the seed coats was based on our observation that the abundant 35-kD protein appeared to bind to insoluble PVPP. The procedure is very simple and permits large amounts of the protein to be purified for structural studies. We do not know the reason why PRP1 binds to PVPP whereas PRP2, a very similar protein, does not bind (Figure 8). Both proteins have basic isoelectric points of 10.7 and 10.4 for PRP1 and PRP2, respectively. Perhaps an explanation for the different affinities is found in the higher tyrosine content of PRP1 (16%) versus PRP2 (11%). PVPP is commonly used to remove phenolics from plant tissues because it binds to the phenolic hydroxyl residues (Loomis, 1974), and it also has been used to purify flavonoids and anthocyanins (Van Teeling et al., 1971). If the very high tyrosine content enables PRP1 to bind PVPP through the tyrosyl hydroxyl groups, then PRP2 might not bind because its tyrosine content is lower. Alternatively, it is possible that the PRP1 protein is complexed with flavonoid or other phenolic compounds and binds because of their affinity for PVPP.

Developmental Regulation of Soluble PRP1 and PRP2 Polypeptides in Seed Coats

The patterns of soluble PRP1 and PRP2 accumulation follow the translation of their respective mRNAs in the seed coat with exceptional fidelity. In essence, they form a slightly overlapping bimodal curve with PRP1 synthesis earlier and higher than that of PRP2. Initial accumulation of soluble PRP1 in young seed coats appears to be coincident with that for soybean extensin, a hydroxyproline-rich protein found in thickened walls of the palisade and hourglass cells (Cassab et al., 1985; Cassab and Varner, 1987, 1988). PRP1 is synthesized as the seeds and seed coats expand rapidly during early maturation, whereas PRP2 begins to accumulate later and peaks in the fully expanded seeds coincident with the loss of chlorophyll and accumulation of anthocyanin pigments. The decrease in soluble PRP1 protein indicates either that PRP1 proteins are degraded or they are bound to the wall by cross-linking (Fry, 1986) and are no longer extractable when the seed matures. Direct immunolocalization in the wall and detailed turnover studies will be required to determine whether PRP1 persists in the mature seed coat or is degraded. However, there are currently no antibodies that discriminate between PRP1 and PRP2. The existing monoclonal antibody to PRP1 and the polyclonal antibodies to PRP2 recognize both proteins (Figures 5 and 7). The monoclonal antibody probably recognizes a portion of the consensus repeat unit that is prevalent in both proteins.

The Recessive *i* Mutation Affects Protein and mRNA Levels for PRP1 but Not PRP2

We identified PRP1 after finding that a specific 35-kD seed coat protein was affected by seed color genotype (Figure 2). Richland is homozygous for the dominant *I* gene that inhibits production of anthocyanin pigments and results in a yellow seed coat. T157 is a Richland isoline containing a homozygous spontaneous mutation to the recessive *i* allele that conditions a dark-pigmented seed coat. Our investigations show that PRP1 protein is lower at each stage of development of T157 as compared with developing Richland seed coats (Figure 5). Maximal PRP1 extraction from T157 seed coats requires a slightly higher NaCl concentration than do Richland seed coats (data not shown); however, in all extraction conditions the amount of PRP1 protein was at least 50% lower in T157 as compared with Richland. PRP1 mRNA abundance in T157 is also less than that of Richland (Figure 6). Interestingly, the mRNA difference is most significant at the time of peak mRNA accumulation. It appears that the developmentally regulated boost in PRP1 mRNA production that occurs in both varieties is higher in Richland than in the recessive *i/i* genotype of T157. This effect could be a consequence of increased transcription of the PRP1 gene in Richland at this developmental stage or a decrease in processing/stability of the mRNA in T157. Further experiments will be needed to determine whether the observed difference in PRP1 cytoplasmic mRNA levels is the only factor affecting the amount of PRP1 protein translated in vivo.

In contrast to the effect of *I* genotype on PRP1 protein and mRNA levels in young seed coats, PRP2 expression in the older seed coats is not affected by the *I* genotype (Figures 9 and 10). If the *I* gene is regulatory, then its transacting effect appears to be specific for PRP1, possibly because of differences in upstream sequences of PRP1 and PRP2 genes. Alternatively, expression of the *I* product could be limited to the early stages of seed coat development when PRP2 is not expressed.

An Unexpected Connection Exists between Anthocyanin Mutations and Cell Wall Proteins in Soybean Seed Coats

Quantitative variation in the level of an abundant cell wall protein due to seed color genotype was an unanticipated result. Most likely, the *I* gene does not encode PRP1;

rather, it affects its abundance. There are a number of isogenic pairs of soybean lines that are due to spontaneous mutations at the *I* locus giving a recessive *i/i* genotype. We have examined several of these mutations in other cultivars (Harsoy, Beeson, and Mandarin, for example) and found that PRP1 protein was reduced in plants carrying recessive *i/i* mutations (data not shown) similar to the situation we have described with the Richland/T157 isolines.

The molecular mechanism of how the *I* locus inhibits pigmentation is unknown. Anthocyanin genes have been investigated more thoroughly in other plant species, including maize, snapdragon, and petunia. In these plants, there is no evidence of anthocyanin genes that have either a direct or pleiotropic effect on cell wall proteins or on the structural integrity of pericarp, aleurone, or flower tissues. However, the precursors for anthocyanin biosynthesis derive from phenylalanine by way of the general phenylpropanoid pathway that also provides precursors for biosynthesis of lignin found in secondary thickenings of the cell wall (Stafford, 1990). Sclereid cells of the soybean seed coat are lignified (Cassab and Varner, 1988), but it is not known whether there is a difference in the amount of lignin or other wall components in seed coats with *I/I* versus *i/i* genotypes.

Our results describing the quantitative effect of *I* mutations on PRP1 mRNA and protein levels raise the possibility that *I* may be a regulatory gene with multiple effects. For example, the dominant *I* allele may act to increase certain mRNAs, as PRP1, during early seed coat development and decrease or suppress simultaneously the accumulation of cytoplasmic RNAs for key structural genes needed in the anthocyanin pathway. Alternatively, it is possible that inhibition of the anthocyanin pathway could be a direct result of increased PRP1 protein if phenolic precursors of anthocyanin biosynthesis preferentially bind to cytoplasmic PRP1 because of its high proline and rodlike structure. Thus, increased levels of PRP1 in seed coats with the *I* allele might titrate the precursors or shuttle them to the cell wall, leading to a shutdown of the anthocyanin pathway by a feedback mechanism. Phenolic compounds are known to have a high affinity for proline-rich proteins in animal saliva. The proline-rich proteins are induced in parotid glands of mice fed diets high in polyphenolic tannins (Butler et al., 1986), presumably as a defense mechanism to bind the phenolic compounds.

In retrospect, the defective seed coat phenotype characteristic of soybean lines like T157 indicates there must be a connection between cell walls and seed coat pigmentation at some level (Figure 1). Whether the lower amount of PRP1 in *i/i* genotypes leads directly to reduced strength or lack of flexibility of seed coat cell walls remains to be tested. However, it is clear that a defect occurs in the wall composition or structure leading to the extensive cracks. The defective seed coat character does not occur in all pigmented soybeans but only in those that are homozygous recessive for both the *i* and *t* genes. The *T* locus is another gene involved in the flavonoid pathway and is possibly a 3′ hydroxylase that modifies the β-ring of flavonoid compounds (Buzzell et al., 1987).

In addition to the effect of the *I* locus on PRP1, we have found recently that another soybean gene affects both seed coat pigmentation and the proline-rich proteins. Surprisingly, near isogenic lines containing the dominant *Im* gene (inhibitor of pigment mottling) have modified forms of both PRP1 and PRP2 proteins in the seed coats (L.O. Vodkin, J.T. Lindstrom, and C. Nicholas, unpublished results). Each protein is approximately 1 kD smaller in SDS-PAGE, although the nature of the difference is unknown and could represent a proteolytic cleavage or other epigenetic modification such as a change in glycosylation or hydroxylation levels of the proline-rich proteins.

In summary, there are a number of intriguing mutations in soybean that have unusual effects on seed coat pigmentation and structure. Undoubtedly, the connection between genotype and phenotype is complex, but our initial studies demonstrate that some of these mutations affect a specific class of hydroxyproline-rich cell wall proteins.

METHODS

Plant Material

The soybean (*Glycine max*) cultivars Richland (genotype *I/I R/R t/t*) and spontaneous mutant T157 (*i/i R/R t/t*) were obtained from the U.S. Department of Agriculture germplasm collection (USDA/ARS, Department of Agronomy, University of Illinois). Plants were grown under standard greenhouse conditions. Supplemental lighting was applied to extend daylength to 14 hr. Additional material was obtained using plants grown under field conditions. Immature pods were harvested from both Richland and T157 and the seed separated by fresh weight into 13 classes: 10 mg to 25 mg, 25 mg to 50 mg, 50 mg to 75 mg, 75 mg to 100 mg, 100 mg to 150 mg, 150 mg to 200 mg, 200 mg to 250 mg, 250 mg to 300 mg, 300 mg to 350 mg, 350 mg to 400 mg, 400 mg to 450 mg, 450 mg to 500 mg, and mature. Seed coats were dissected from immature seed, frozen in liquid nitrogen, lyophilized, and stored at −20°C until needed. Both Richland and T157 are Maturity Group II soybean lines. Under our growing conditions, seed 27 DAF weigh approximately 75 mg to 100 mg and correspond to Stage K of soybean seed development, as described by Meinke et al. (1981) for cultivar Provar (also Maturity Group II). The first appearance of anthocyanins in the seed coat occurs in seed greater than 300 mg in fresh weight (Figure 1).

Protein Purification

Purification of PRP1 from seed coats was achieved by chromatography on insoluble PVPP (Sigma catalog No. P-6755). The PVPP was hydrated in sterile distilled water at least 1 hr before use and fines were removed. The hydrated PVPP was poured

into a Bio-Rad Econo column and allowed to settle, and 2 bed volumes of PBS (200 mM NaCl, 9 mM phosphate, pH to 7.2) were then passed through the resin.

For each protein extraction, 150 mg of lyophilized soybean seed coats, collected from seed weighing between either 75 mg to 100 mg or 100 mg to 150 mg, was used. The seed coats were ground in 3 mL of PBS. The crude extract was poured into a 15-mL Corex tube and centrifuged for 10 min at 7500 rpm, 4°C, in a Beckman JA-17 rotor. The supernatant was transferred to a new tube on ice and the pellet discarded. A 100-μL sample of the crude extract was reserved for protein determination, and the rest was applied to the PVPP column and allowed to pass through by gravity at room temperature. Effluent from the column was collected and passed over the column a second time. Proteins not binding to the PVPP were removed by washing the column with 2 bed volumes of PBS. Bound protein was eluted from the column using 2 bed volumes of 0.5 M acetic acid. Effluent pH was monitored using pH paper; when pH dropped below 6.5, effluent was collected into microcentrifuge tubes. Fractions collected after the pH change were pooled and immediately neutralized by the addition of 0.3 mL of 1 M NaOH per milliliter of effluent.

Protein samples were dialyzed overnight in 200 mM NaCl and then overnight in distilled water with one change. Samples were collected from dialysis tubing, placed in microcentrifuge tubes, frozen in liquid nitrogen, and lyophilized. The white pellet was resuspended in 10 μL of sterile water and samples were pooled. Protein concentration of the pooled sample and the crude extract was determined using a 5-μL aliquot and the Bio-Rad protein assay (Bradford, 1976). Determination of purity was made using SDS-PAGE. The protein is stable when stored at 4°C in water.

N-Terminal Amino Acid Sequencing and Amino Acid Composition

Samples of protein were submitted for N-terminal amino acid sequencing either blotted onto Immobilon (Millipore) or dissolved in water. Both N-terminal analysis and amino acid composition were performed by the University of Illinois Biotechnology Center. The National Biomedical Research Foundation protein data bases were searched for similarity to the N-terminal sequences using the FASTA algorithm of GCG software (Devereux et al., 1984).

Antibody Production

Antibodies to the purified 35-kD seed coat protein, PRP1, were produced in mice at the University of Illinois Hybridoma Facility. A monoclonal line was selected based on its reaction to purified PRP1 that was bound to nitrocellulose after electrophoretic transfer as described below. Polyclonal antibodies to soybean hypocotyl hook protein, PRP2, were obtained from Dr. Mary Tierney of the Ohio State University.

Protein Extractions and Immunoblotting

Seed coats were extracted in PBS that was at least 200 mM NaCl. In general, individual freeze-dried seed coats were extracted in microtubes using 150 μL of buffer for genetic studies. For immunoblotting, pooled samples of five seed coats were extracted with 500 μL of buffer after grinding with nitrogen in a mortar with pestle. Supernatants were subjected to SDS-PAGE (Laemmli, 1970), stained with 0.1% Coomassie Brilliant Blue R (Sigma), or transferred to Immobilon or nitrocellulose using 10 mM Caps (3-[cyclohexylamino]-1-propanesulfonic acid), pH 11.00, 0.1% SDS, 10% methanol. Membranes were blocked 1 hr in Blotto (5% nonfat dry milk in Tris-buffered saline [TBS]) and then incubated with a 1:1000 dilution of PRP1 antibody in TTBS (0.15% Tween 20 in TBS) for 1 hr. Membranes were washed four times for 15 min each wash in Blotto. Horse anti-mouse IgG alkaline phosphatase conjugate (Vector Laboratories, Burlingame, CA) was used in a 1:500 dilution in TTBS as the secondary antibody. Membranes were incubated in the secondary antibody for 1 hr, washed three times for 15 min each in TTBS, and developed using BCIP/NBT substrate (Sambrook et al., 1989).

Tests showed that transfer of PRP1 to the membrane support was most effective using 10 mM Caps, pH 11, 0.1% SDS, as the transfer buffer. With a lower buffer, as with Tris-glycine, pH 8.3, 0.1% SDS, much PRP1 remained in the gel even after electrophoretic transfer. This is consistent with the calculated pI of 10.7 for PRP1 based on its sequence.

RNA Isolation

To obtain total RNA, seed coats (10 to 15) or dark-grown hypocotyl regions were ground under liquid nitrogen to a fine powder, extracted with phenol-chloroform, and precipitated with lithium chloride, essentially as described by McCarty (1986).

Seed Coat mRNA Library Construction and Sequencing

A cDNA library was constructed from seed coat (75 mg to 100 mg, fresh weight) polyA$^+$ mRNA isolated by chromatography on oligo(dT)-cellulose (Maniatis et al., 1982). Eight micrograms of polyA$^+$ mRNA was used to construct the cDNA library in λgt10 as per instructions of the kit manufacturer (Bethesda Research Laboratories) using EcoRI methylase and EcoRI linkers. The resulting cDNA library was probed with p1A10-1, the cloned repeat unit from PRP1 (Averyhart-Fullard et al., 1988). Positive clones were isolated and phage DNA extracted (Maniatis et al., 1982). A 993-bp insert was subcloned from phage DNA into pGEM7 (Promega). Single-stranded DNA was obtained from this clone, pB1-3, and sequenced using dideoxy nucleotide analogs (Sanger et al., 1977). Additional sequence data were obtained by constructing a series of nested deletions using an ExoIII-mung bean nuclease deletion kit (Stratagene).

Gene-Specific Probes and RNA Gel Blots

A PRP1 sequence-specific probe was prepared from pB1-3 by digesting the plasmid DNA with BstNI and EcoRI, yielding a 94-bp fragment containing the 3′ untranslated region. The fragment was isolated using low-melt agarose gels and labeled directly with ^{32}P-dATP using random primers (Feinberg and Vogelstein, 1983). A PRP2 gene-specific probe was prepared from p1A10-3 (Datta et al., 1989) by digesting the plasmid with AccI. The 400-bp AccI fragment was separated on an agarose gel, purified using Geneclean II (Bio 101, La Jolla, CA), and radiolabeled with

^{32}P-dATP using random primers. For RNA gel blots, 1 μg of total seed coat RNA was electrophoresed on a 1.2% formaldehyde agarose gel (Thomas, 1980; Maniatis et al., 1982). Gels were rinsed briefly in distilled water, transferred overnight to nitrocellulose, cross-linked to the membrane with UV light using a Stratalinker (Stratagene), and hybridized with gene-specific probes. Equal amounts of RNA were applied to each lane and accuracy of loading was checked by including 400 ng of ethidium bromide in each RNA sample (K.M. Rosen and L. Villa-Romaroff, 1990. An alternative method for the visualization of RNA in formaldehyde agarose gels. Focus, Vol. 12, 23–24, Bethesda Research Laboratories). Gels were photographed before transfer.

Note on Terminology for Soybean Proline-Rich Proteins and Sequences

We have followed the terminology of Hong et al. (1987, 1989, 1990) for the SbPRP1, SbPRP2, and SbPRP3 sequences and predicted protein products except that the qualifier Sb (soybean) is dropped for brevity in most cases. Other abbreviations in the literature are RPRP2 (28-kD repetitive proline-rich protein) and RPRP3 (33-kD repetitive proline-rich protein), which have been used to designate the two forms of cell wall polypeptides that are products of the PRP2 gene as determined by N-terminal protein sequence and amino acid composition (Averyhart-Fullard et al., 1988; Datta et al., 1989; Kleis-San Francisco and Tierney, 1990).

ACKNOWLEDGMENTS

We thank Dr. Mary Tierney (Ohio State University) for the generous gift of polyclonal antibodies to PRP2 and Dr. Abraham Marcus (Fox Chase Cancer Center, Philadelphia) for the PRP1 partial cDNA (p1A10-1) and the PRP2 AccI subclone (p1A10-3). We thank Steve Miklasz of the University of Illinois Hybridoma Facility for producing the monoclonal antibody to PRP1 and the Biotechnology Center for the N-terminal sequence analysis. This research was supported by grants from the Sohio Center for Crop Molecular Genetics and by U.S. Department of Agriculture Competitive Grant 90-37231-5441.

Received February 8, 1991; accepted April 9, 1991.

REFERENCES

Averyhart-Fullard, V., Datta, K., and Marcus, A. (1988). A hydroxyproline-rich protein in the soybean cell wall. Proc. Natl. Acad. Sci. USA **85,** 1082–1085.

Bernard, R.L., and Weiss, M.G. (1973). Qualitative genetics. In Soybeans: Improvement, Production and Uses, 1st ed, B.E. Caldwell, ed (Madison, WI: American Society of Agronomy), pp. 117–149.

Bradford, M.M. (1976). 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.

Butler, L.G., Rogler, J.C., Mehansho, H., and Carlson, D.M. (1986). Dietary effects of tannins. In Plant Flavonoids in Biology and Medicine, V. Cody, E. Middleton, Jr., and J.B. Harborne, eds (New York: Alan R. Liss), pp. 141–157.

Buzzell, R.I., Buttery, B.R., and MacTavish, D.C. (1987). Biochemical genetics of black pigmentation of soybean seed. J. Hered. **78,** 53–54.

Carlson, J.B., and Lersten, N.R. (1987). Reproductive morphology. In Soybeans: Improvement, Production, and Uses, 2nd ed., J.R. Wilcox, ed (Madison, WI: Agronomy Society of America), pp. 95–134.

Cassab, G.I., and Varner, J.E. (1987). Immunocytolocalization of extensin in developing soybean seed coats by immunogold-silver staining and by tissue printing on nitrocellulose paper. J. Cell Biol. **105,** 2581–2588.

Cassab, G.I., and Varner, J.E. (1988). Cell wall proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. **39,** 321–353.

Cassab, G.I., Nieto-Sotelo, J., Cooper, J.B., van Holst, G.-J., and Varner, J.E. (1985). A developmentally regulated hydroxyproline-rich glycoprotein from the cell walls of soybean seed coats. Plant Physiol. **77,** 532–535.

Datta, K., Schmidt, A., and Marcus, A. (1989). Characterization of two soybean repetitive proline-rich proteins and a cognate cDNA from germinated axes. Plant Cell **1,** 945–952.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12,** 387–395.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. **132,** 6–13.

Fry, S.C. (1986). Cross-linking of matrix polymers in growing cell walls of angiosperms. Annu. Rev. Plant Physiol. **37,** 165–186.

Hong, J.C., Nagao, R.T., and Key, J.L. (1987). Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. J. Biol. Chem. **262,** 8367–8376.

Hong, J.C., Nagao, R.T., and Key, J.L. (1989). Developmentally regulated expression of soybean proline-rich cell wall protein genes. Plant Cell **1,** 937–943.

Hong, J.C., Nagao, R.T., and Key, J.L. (1990). Characterization of a proline-rich cell wall protein gene family of soybean. J. Biol. Chem. **265,** 2470–2475.

Kleis-San Francisco, S.M., and Tierney, M.L. (1990). Isolation and characterization of a proline-rich cell wall protein from soybean seedlings. Plant Physiol. **94,** 1897–1902.

Laemmli, U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature **227,** 680–685.

Lamport, D.T.A., and Northcote, D.H. (1960). Hydroxyproline in primary cell walls of higher plants. Nature **118,** 665–666.

Loomis, W.D. (1974). Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. Methods Enzymol. **31,** 528–545.

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

McCarty, D.R. (1986). A simple method for extraction of RNA from maize tissue. Maize Genet. Coop. Newsl. **60,** 61.

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

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463–5467.

Stafford, H.A. (1990). Flavonoid Metabolism. (Boca Raton, FL: CRC Press), pp. 15–20.

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

Van Teeling, C.G., Cansfield, P.E., and Gallop, R.A. (1971). Chromatography of anthocyanins on columns of insoluble polyvinylpyrrolidone. J. Chromatogr. Sci. **9,** 505–509.

Wilcox, J.R. (1988). Performance and use of seedcoat mutants in soybean. Crop Sci. **28,** 30–32.