

Identification of Glycinin and β -Conglycinin Subunits that Contribute to the Increased Protein Content of High-Protein Soybean Lines

HARI B. KRISHNAN,^{*,†,‡} SAVITHIRY S. NATARAJAN,[§] AHMED A. MAHMOUD,[‡] AND RANDALL L. NELSON^{||}

Plant Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Columbia, Missouri 65211, Plant Science Division, University of Missouri, Columbia, Missouri 65211, Soybean Genomics and Improvement Laboratory, USDA ARS, 10300 Baltimore Avenue, Beltsville, Maryland 20705, and USDA Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801

Seed protein concentration of commercial soybean cultivars calculated on a dry weight basis ranges from approximately 37 to 42% depending on genotype and location. A concerted research effort is ongoing to further increase protein concentration. Several soybean plant introductions (PI) are known to contain greater than 50% protein. These PIs are exploited by breeders to incorporate the high-protein trait into commercial North American cultivars. Currently, limited information is available on the biochemical and genetic mechanisms that regulate high-proteins. In this study, we have carried out proteomic and molecular analysis of seed proteins of LG00–13260 and its parental high-protein lines PI 427138 and BARC-6. Sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis revealed that the high-protein lines accumulated increased amounts of β -conglycinin and glycinins, when compared with Williams 82. High-resolution two-dimensional electrophoresis utilizing pH 4–7 and pH 6–11 ampholytes enabled improved resolution of soybean seed proteins. A total of 38 protein spots, representing the different subunits of β -conglycinin and glycinin, were identified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. High-protein was correlated with an increase in the accumulation of most of the subunits representing β -conglycinin and glycinin. Comparisons of the amino acid profiles of high-protein soybean lines revealed that the concentration of sulfur amino acids, a reflection of protein quality, was not influenced by the protein concentration. Southern blot analysis showed the presence of genotypic variation at the DNA level between PI 427138 and BARC-6 for the genes encoding group1 glycinin, β -conglycinin, Bowman–Birk inhibitor (BBI), and the Kunitz trypsin inhibitor (KTI). LG00–13260 inherited the allelic variants of the parental line PI 427138 for glycinin, β -conglycinin, and KTI, while BBI was inherited from the parental line BARC-6. The results of our study indicate that high-seed protein concentration is attributed to greater accumulation of specific components of β -conglycinin and glycinin subunits presumably mediated by preferential expression of these genes during seed development.

KEYWORDS: High-protein soybeans; protein composition; protein quality; proteomics

INTRODUCTION

A preponderance of the world's vegetable protein is supplied by soybeans, the majority being used to formulate rations for domestic livestock and poultry and the remaining utilized for human consumption. Presently, soybeans grown in the United States have an average protein content of 40%, but shifts in

production to the upper Midwest and continued emphasis on breeding for yield rather than quality threatens to lower this percentage of protein (1, 2). A concerted effort deemed the Better Bean Initiative was inaugurated in 2002 by the United States Soybean Board to facilitate improvements in soybean seed composition (1). One major objective of the Better Bean Initiative is to develop high-protein soybeans that could facilitate cost-efficient production of high-protein meals containing at least 48% protein. Underscoring the importance of soybean protein content is the fact that today, 70% of the crop's value is derived from this component (3).

* Corresponding author. Tel.: (573) 882-8151. Fax: (573) 884-7850. E-mail: KrishnanH@missouri.edu.

[†] Agricultural Research Service.

[‡] University of Missouri.

[§] Soybean Genomics and Improvement Laboratory.

^{||} University of Illinois.

A major impediment to increasing soybean protein through selective breeding lies with the inverse relationship existing between protein content and yield (4–7). In the past 60 years of soybean breeding, the focus has been primarily on yield using elite domestic cultivars as parents. This selection regime has resulted in a narrow genetic base that potentially could limit the concurrent improvement in yield, protein, and oil contents (8–11). To overcome this problem, parental lines from plant introductions (PI) have been employed to expand the genetic base and ostensibly improve the soybean seed composition. Although extensively time-consuming, the use of PI as a parent in backcrossing or in recurrent selection has in some cases been successful in overcoming the negative genetic correlation between protein and yield (6, 12–14). There are high-protein cultivars and experimental lines that can serve as parental lines for further varietal development. Examples of such high-protein lines include Prolina, D76–9070, BARC-6, BARC-7, BARC-8, BARC-9, and S96–2461 (15–19).

The work of Wilcox and Cavins (6) showed definitively that the protein content could be increased without sacrificing yield. Currently, limited information is available on the biochemical and genetic mechanisms that regulate high-seed protein concentrations. In this study, we employed various analytical methods to determine whether the high-protein lines preferentially accumulate specific proteins and to compare the relative accumulation of amino acids in the high-protein lines. Our data provide increased knowledge of the variability of protein and protein subunit accumulation among high-protein cultivars, which will facilitate ongoing efforts to improve both quantity and quality of soybean seed protein.

MATERIALS AND METHODS

Plant Materials. PI 427138 is a maturity group 0 line that was introduced from the Crop Experiment Station at Suweon, South Korea in 1978 with the name Choseng No. 1 (20). PI 427138 averaged 49.8% protein from data collected at four locations over 2 years (unpublished data). BARC-6 (16) was developed from a cross of two high-protein experimental lines CX797–21 (21) and D76–8070 (15). In 19 field tests, BARC-6 averaged 50.6% protein on a dry weight basis. LG00–13260 is an F9 line selected from the cross of PI 427138 by BARC-6. LG00–13260 averaged 53.9% protein from data collected at three locations in 2001. Williams 82, a maturity group III cultivar, averaged 41.4% protein on a dry weight basis. The seeds of Williams 82 and the high-protein soybean lines used in this research were grown in Urbana, IL. Cultural practices were typical of those utilized for soybean production in the Midwest United States.

Chemicals. Chemicals for electrophoresis including acrylamide, bis-acrylamide, SDS, TEMED, ammonium persulfate, thiourea, dithiothreitol, CHAPS, and IPG strips were purchased from GE healthcare (Piscataway, NJ). Urea and ampholytes were purchased from Bio-Rad (Hercules, CA). Tris-HCl, 2-mercaptoethanol (2-ME), and glycerol were purchased from Sigma (St. Louis, MO). α -Cyanohydroxycinnamic acid (CHCA) matrix was purchased from Bruker Daltonics, Billerica, MA. Water from a Millipore Milli-RO4 reverse osmosis system was used for making all solutions.

Extraction of Proteins from Seeds. Seed proteins were extracted by the modified trichloroacetic acid (TCA)/acetone precipitation/urea solubilization protocol (22). For this method, soybean seeds were powdered in liquid nitrogen using a mortar and pestle. One hundred milligrams of the soybean seed powder was homogenized with 2 mL of a solution containing 10% (w/v) TCA in acetone with 0.07% (v/v) 2-mercaptoethanol. Total protein was precipitated for 1 h at -20°C . The extract was centrifuged at 20 800g for 20 min at 4°C . The pellet was washed 2–3 times with acetone containing 0.07% (v/v) 2-mercaptoethanol. Then, the pellet was dried under vacuum for 30 min, and the acetone dry powder was resuspended in 1 mL of lysis buffer [9 M urea, 1% CHAPS, 1% [w/v] ampholytes (pH 4–7 or 6–11), and

1% DTT] followed by sonication for 30 min. Non-proteinaceous material was removed by centrifugation at 20 800g for 20 min at 4°C , and the supernatant was used in 2-D-PAGE analysis. The concentration of proteins extracted by the modified TCA/acetone method was determined using the Bradford method (23) and using a commercial dye reagent from Bio-Rad (Hercules, CA).

2-D-PAGE Analysis. The first dimension, IEF, was performed using 13 cm pH 4–7 and 6–11 linear IPG strips in the IPGphor apparatus (GE Healthcare, Piscataway, NJ), according to the manufacturer's recommendations. For the second dimension, the IPG strips were incubated with 50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, and 1% DTT for 15 min and acetylated with iodoacetamide and subsequently placed onto 12% polyacrylamide gels as described by Laemmli (24). The electrophoresis was performed using a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ) according to the manufacturer's recommendations. The 2-D-PAGE gels were visualized by staining with Colloidal Coomassie Blue G-250 as described by Newsholme (25). The gels were stored in 20% ammonium sulfate solution and scanned using laser densitometry (GE Healthcare, Piscataway, NJ). Triplicate samples were used for soybean seed protein extraction and 2-D-PAGE analysis.

Various isoelectric forms of glycinin and β -conglycinin were identified by comparing the matrix-assisted laser desorption ionization time-of-flight mass fingerprinting data against the National Center of Biotechnology Information nonredundant database as described earlier (26). Image analysis of 2-D-PAGE spots was performed as described by Hajduch et al. (27) with ImageMaster 2-D Platinum software version 5.0 (Amersham Biosciences, Uppsala, Sweden). This software performs 2-D spot analysis, background subtraction, and spot matching among different gels. Quantification of protein spots from high-protein lines and Williams 82 were determined from three independent gels. Drawing tools were used to encircle detectable spots to specify regions for analysis and to subtract background signals. Automatic algorithms calculated the quantitative statistics including spot volume, pixel area, mean pixel intensity, standard deviation, and background values, and the data were tabulated in a spreadsheet format.

Amino Acid Analysis. Four replicates of dry seed powder of Williams 82 and high-protein lines were hydrolyzed under nitrogen at 155°C for 16 h in 6 N HCl. The content of sulfur amino acids and tryptophan was determined as described earlier (26). Amino acids were fractionated on a high-performance cation-exchange resin column installed in a Beckman 6300 Amino Acid Analyzer (Beckman Instruments, Fullerton, CA).

Southern Hybridization. Total genomic DNA was isolated from leaf tissue of the soybean cultivars using the cetyltrimethyl ammonium bromide (CTAB) method (28). Genomic DNA (8 μg) was digested overnight at 37°C with different restriction enzymes. Southern hybridization was performed as described earlier (2). Briefly, restricted DNA was electrophoretically separated on a 0.8% agarose gel and then transferred to a nylon membrane by capillary action using 0.04 M NaOH as the carrier. DNA fragments corresponding to the coding regions of the glycinin *Gy2* and *Gy4*, β -conglycinin, BBI, Soy CII, and KTI were used as probes for Southern hybridization. The probe DNA was labeled with ^{32}P -dCTP using the Ladderman kit (Takara Bio Inc. Otsu, Shiga, Japan). A minimum of a 10 h prehybridization followed by overnight hybridization were both conducted at 65°C using $6\times$ SSPE buffer, ($1\times$ SSPE is 0.1 M NaCl, 0.01 M Na_2PO_4 , and 0.001 M EDTA), $5\times$ Denhardt's solution, 0.5% SDS, and 50 $\mu\text{g}/\text{mL}$ sheared and denatured salmon sperm DNA. After hybridization, the membranes were washed 3 times in a solution of $2\times$ SSPE and 0.5% SDS for 10 min at room temperature and then twice in a solution of $0.1\times$ SSPE and 0.1% SDS at 65°C for 30 min. Following washing, the nylon membrane was exposed overnight at -80°C to X-ray film.

RESULTS

SDS-PAGE Analysis of Seed Protein Composition of High-Protein Lines. We initiated our studies by comparing the protein composition of high-protein soybean lines PI 427138 (49.8% protein), LG00–13260 (53.9% protein), and BARC-6 (50.6%

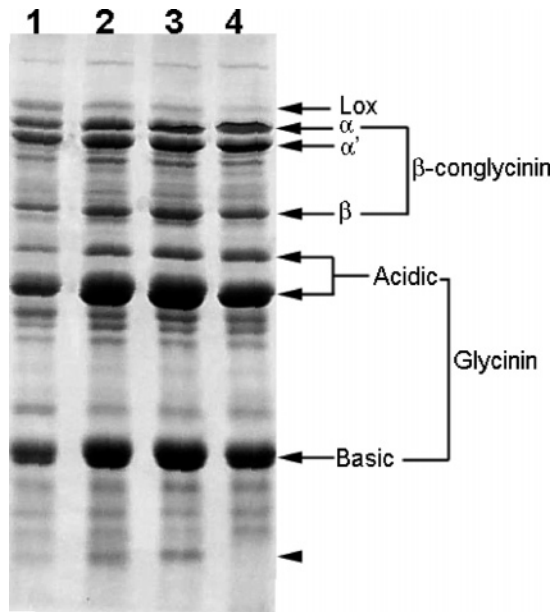


Figure 1. Protein profiles of high-protein soybean lines. Total seed proteins from Williams 82 (lane 1), and high-protein lines PI 427138 (lane 2), LG00-13260 (lane 3), and BARC-6 (lane 4) obtained on the basis of equal seed dry weight were fractionated by SDS-PAGE on a 13.5% gel and stained with Coomassie Blue. Abundant seed storage proteins are identified. The arrowhead points to a 14 kDa protein that is absent in BARC-6. Lox, lipoxygenase.

protein) to that of a normal seed protein line, Williams 82 (41.4% protein). A visual examination of the total protein profile of seed proteins resolved by SDS-PAGE gel clearly shows that the overall protein was elevated in all the high-protein lines when compared with that of Williams 82 (**Figure 1**). Most of the storage proteins, including the α' -, α -, and β -subunits of β -conglycinin (7S storage proteins), and the acidic and basic

subunits of glycinins (11S storage proteins) accumulated at higher amounts in these high-protein lines when compared to Williams 82 (**Figure 1**). In contrast, the accumulation of lipoxygenase (94 kDa) was higher in Williams 82 when compared to the high-protein lines (**Figure 1**). Overall, the seed protein composition of all soybean lines was strikingly similar with the exception of a 14 kDa protein, which was absent in BARC-6 (**Figure 1**).

High-Resolution 2-D Gel Analysis of Seed Protein Composition of High-Protein Lines. The two predominant storage proteins of soybean, the glycinins and β -conglycinin, are encoded by multigene families (29–32). These proteins are composed of different subunits and exhibit heterogeneity as evidenced by differences in their isoelectric points and molecular weights (33, 34). The use of one-dimensional gels, although useful in detecting obvious differences in the quantities of protein present among the high-protein lines and Williams 82, does not provide information on the accumulation of the different subunits of glycinins and β -conglycinin. To investigate the contribution of individual protein components to the high-protein trait, 2-D polyacrylamide gel electrophoresis was performed. Since the different subunits of glycinins and β -conglycinin are not adequately resolved in pH 3–10 linear IPG strips, analysis under two different pH ranges was performed. The various subunits of β -conglycinin and the acidic subunits of glycinins were distinctly resolved utilizing pH 4–7 linear IPG strips (**Figure 2**) and the basic subunits of glycinins with pH 6–11 linear IPG strips (**Figure 3**). Comparison of 2-D electrophoresis patterns revealed distinct changes among high-protein soybean lines and Williams 82 (**Figures 2 and 3**). Both subtle and obvious differences in the number and the abundance of various spots representing the acidic and basic subunits of glycinins were detected. Even though the number of protein spots representing α' - and α -subunits was similar among the lines, the abundance of the α -subunit was higher in the high-protein soybean lines when compared to that of Williams 82

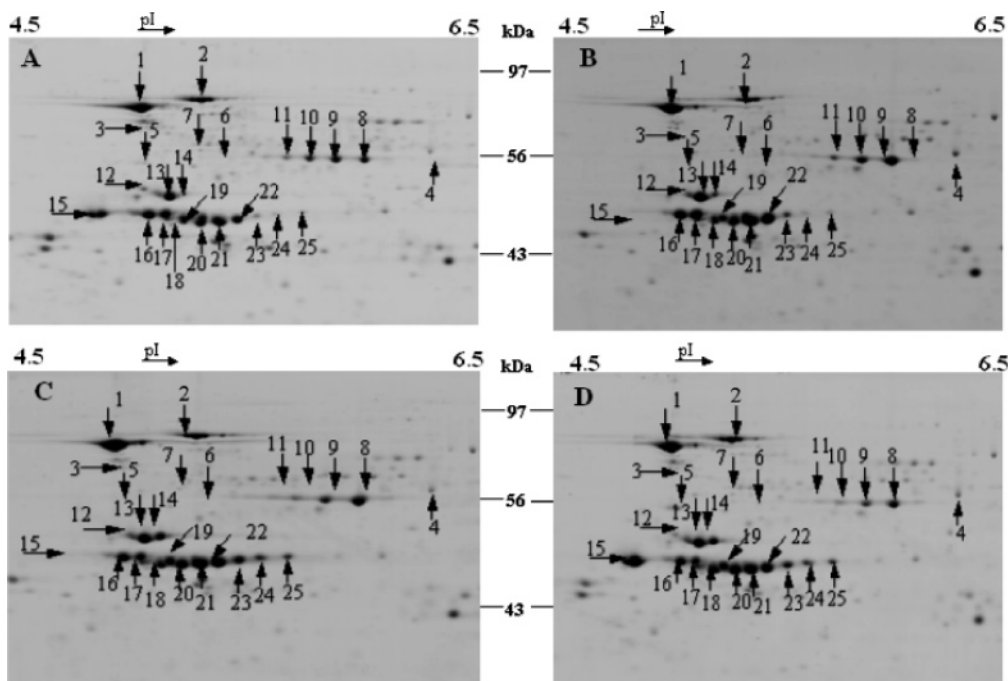


Figure 2. Comparison of the storage proteins of Williams 82 and high-protein soybean lines by 2-D gel electrophoresis. Proteins obtained on the basis of equal seed dry weight were first separated by isoelectric focusing using a pH gradient from 4 to 7 and then by SDS-PAGE on a 12% gel. The gels were stained with Colloidal Coomassie Blue G-250. The arrows point to different proteins that were identified by MALDI-TOF-MS. (A) Williams 82; (B) PI 427138; (C) LG00-13260; and (D) BARC-6.

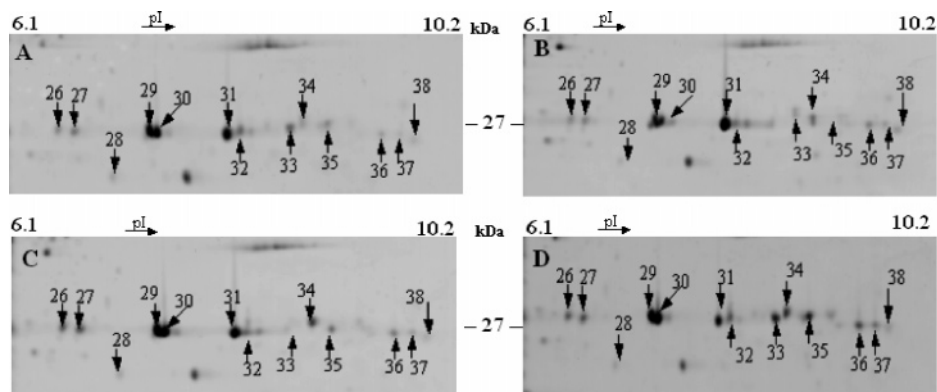


Figure 3. Comparison of the storage proteins of Williams 82 and high-protein soybean lines by 2-D gel electrophoresis. Proteins obtained on the basis of equal seed dry weight were first separated by isoelectric focusing using a pH gradient from 6 to 11 and then by SDS-PAGE on a 12% gel. The gels were stained with Colloidal Coomassie Blue G-250. The arrows point to different proteins that were identified by MALDI-TOF-MS. (A) Williams 82; (B) PI 427138; (C) LG00-13260; and (D) BARC-6.

(**Figure 2**). To elucidate fully the role of individual protein spots to the overall protein content, a total of 38 previously identified protein spots (26) representing the different isoelectric forms of glycinins and β -conglycinin were selected for further analysis. A comparison of the abundance of these protein spots was performed with the aid of ImageMaster 2-D Platinum software version 5.0 (Amersham Biosciences, Uppsala, Sweden). The abundance of the protein spots was expressed as a relative volume to compensate for differences in sample loading and variation in gel staining. An examination of the data provides evidence for quantitative increases in the accumulation of several of the subunits of both glycinins and β -conglycinin in the high-protein soybean lines when compared to that of Williams 82. When the relative volumes of each spot associated with different subunits of glycinin and β -conglycinin subunits were combined, significant differences in their concentration was evident between Williams 82 and each of the high-protein lines. The concentration of the α -subunit of β -conglycinin in the high-protein lines was 20–40% higher than Williams 82. For the α' -subunit, the changes ranged from 5 to 59%, and for the β -subunit, the differences varied from -3 to $+85\%$. There were also large differences for glycinin subunits.

Comparison of Amino Acid Composition. Even though soybeans are an excellent source of protein, their quality can be further improved by increasing the concentration of sulfur-containing amino acids. To examine if there is any alteration in sulfur amino acid content as a consequence of increasing the overall protein content, the amino acid composition of soybean seed proteins was performed. Subtle differences were seen in the amino acid content among the four tested soybean lines. Pairwise comparisons revealed that the high-protein soybean lines, LG00–13260 and BARC-6, had a similar amount of methionine and cysteine as Williams 82. In contrast, the methionine and cysteine content of PI 427138 was 9 and 19% higher when compared to other high-protein soybean lines, but the actual differences were still quite small. The concentration of tryptophan decreased with increasing protein concentration. In contrast, the arginine content increased with increasing protein concentration.

Southern Blot Analysis of Seed Protein Genes. To determine the allelic composition at the DNA level of the genes encoding seed proteins in the parental lines, BARC-6 and PI 427138, as compared to the progeny LG00–13260, Southern blot analysis was carried out on different restriction digests of genomic DNA from these genotypes as well as from Williams 82 for comparison. A DNA fragment corresponding to exon 3

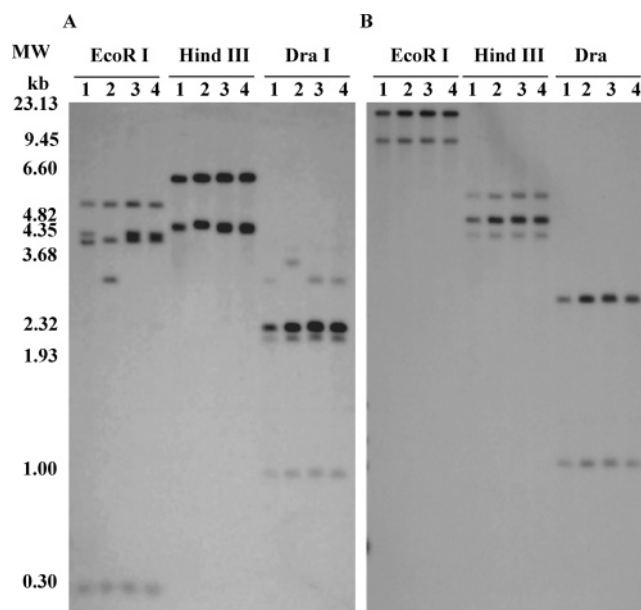


Figure 4. Southern blot analysis of soybean glycinin. Genomic DNA from leaf tissue of Williams 82 (lane 1), BARC-6 (lane 2), LG00-13260 (lane 3), and PI 427138 was individually digested with either *EcoRI*, *HindIII*, or *DraI*, fractionated on an 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with ^{32}P labeled *Gy2* (A) and *Gy4* (B) glycinin probes and then subjected to autoradiography. Molecular weight markers in kilobases are shown on the left side of the figure.

and exon 4 of the *Gy2* glycinin gene was used as a probe for Southern hybridization to detect DNA differences in group 1 glycinin genes, *Gy1*, *Gy2*, and *Gy3*. The Southern hybridization profile presented in **Figure 4A** showed that the two parental lines were different in their hybridization patterns, with LG00–13260 showing the same pattern as that of the parent PI 427138, indicating that LG00–13260 inherited the group 1 glycinin alleles of PI 427138 rather than those of BARC-6. Both PI 427138 and LG00–13260 were very similar in their hybridization to that of Williams 82. For group 2 glycinin genes *Gy4* and *Gy5*, a probe corresponding to exon 3 of *Gy4* was used for southern hybridization analysis. Unlike group 1 glycinin genes, group 2 glycinin showed no detectable differences in their hybridization patterns in the parental lines and LG00–13260 as well as in Williams 82 (**Figure 4B**). These results indicated that group 2 glycinin genes are generally more conserved or under more natural selection pressure than the group 1 genes.

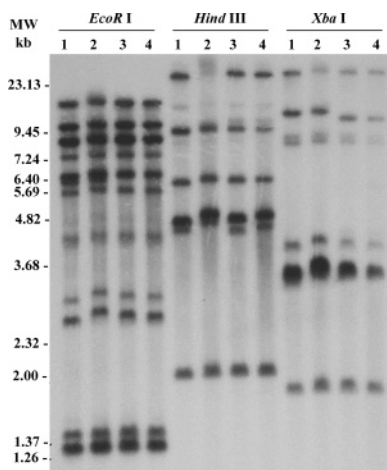


Figure 5. Southern blot analysis of soybean β -conglycinin. Genomic DNA from leaf tissue of Williams 82 (lane 1), BARC-6 (lane 2), LG00-13260 (lane 3), and PI 427138 was individually digested with either *EcoRI*, *HindIII*, or *DraI*, fractionated on a 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with a ^{32}P labeled β -subunit of the β -conglycinin gene and then subjected to autoradiography. Molecular weight markers in kilobases are shown on the left side of the figure.

The β -conglycinin storage proteins are encoded by a multi-gene family of at least 15 highly homologous genes (30). When a probe corresponding to one of β -conglycinin genes was used to hybridize restriction digests of genomic DNA from the investigated genotypes, a total of 8–12 different fragments, depending on the restriction enzyme, showed hybridization to the probe (Figure 5). Some of the strongly hybridized fragments might actually contain more than one fragment that originated from different genes and closely comigrated on the gel due to very similar molecular sizes. Such a case is usually expected when gene copies are organized in tandem arrays, which has been reported to be the case for some of the β -conglycinin genes (30). A close examination of the hybridization patterns of β -conglycinin genes showed that although the majority of the hybridized fragments was common in all the tested genotypes, few fragments were still polymorphic, indicating the presence of a DNA sequence variation in some of the β -conglycinin genes among the genotypes. The two parental lines could be distinguished from each other based on these polymorphic fragments, with LG00-13260 showing a hybridization pattern similar to the parental line PI 427138 (Figure 5). The parental line BARC-6 was more similar than PI 427138 in its hybridization pattern to Williams 82. Therefore, it appears that LG00-13260 has inherited the β -conglycinin allelic variants of the parental line PI 427138 rather than those of BARC-6.

Proteinase inhibitors in soybean seeds, which can make up to 6% of the total seed proteins, are highly diverse in their molecular properties and biological functions (31, 32). The two major types of proteinase inhibitors in soybean seeds are the Bowman-Birk inhibitors (BBI) and the Kunitz trypsin inhibitors (KTI). The BBI-A and C-II (previously known as soy C-II) are two major protease isoinhibitors in the BBI family. Southern blot analysis was carried out using DNA probes specific to the genes encoding these isoinhibitors to determine the level of variation in their DNA sequences in the parental lines and their progeny. The BBI-A probe showed hybridization to three to four main fragments, depending on the restriction enzyme, in all the tested genotypes (Figure 6A). However, the parental line PI 427138 showed clear differences in the molecular size of the hybridized fragments as compared to

BARC-6, LG00-13260, and Williams 82, which all showed very similar molecular sizes to the hybridized fragments. This indicates the presence of DNA sequence variation between the two parental lines for the BBI-A protease inhibitor and that LG00-13260 inherited the allelic variants of BBI-A present in the parental line BARC-6 rather than those in PI 427138 (Figure 6A). Similarly, the Southern hybridization results of BBI C-II showed a unique hybridization pattern for the parental line PI 427138 as compared to BARC-6, LG00-13260, and Williams 82, which all showed very similar hybridization patterns (Figure 6B). To examine the organization of KTI inhibitors in the investigated genotypes, a probe corresponding to the KTI 3 gene, which is the major gene in the KTI gene family, was used for the Southern blot analysis (Figure 6C). The *EcoRI* digest was more informative than that of *HindIII* in revealing the level of variation between the genotypes for KTI genes. The two parental lines and Williams 82 each showed a distinct hybridization pattern from one another, indicating the presence of sequence variation for KTI genes among these genotypes (Figure 6C). The hybridization pattern of the hybrid was similar to the parental line PI 427138, indicating that LG00-13260 acquired the allelic variants of the parental line PI 427138 rather than those of BARC-6.

DISCUSSION

During the last 60 years, the primary focus of soybean breeders has been on yield improvement with minimal emphasis on protein composition. However, both of these parameters must be considered for the continued success of the soybean industry. Presently, an estimated 70% of the value of the soybean is derived from the protein component of the seed (3). Consequently, the generation of high-protein soybean cultivars has been emphasized and remains as an important goal of the Better Bean Initiative. The well-documented negative correlation between protein content and yield has hampered the development of agronomically viable high-protein cultivars. However, soybean breeders have made significant progress in overcoming this negative correlation by employing advanced breeding techniques resulting in the development of agronomically viable high-protein cultivars (15–19). Development of such high-protein soybean lines is facilitated by the high heritability of the trait, availability of high-protein germplasm, and minimal number of genes controlling the trait (35, 36).

In addition to developing high-protein commercial lines, emphasis should also be placed on the issue of protein digestibility and functional properties desirable to industry. Merely increasing the overall seed protein concentration, however, does not address the problem associated with soybean protein quality. A major concern with respect to quality is that the amino acid composition of soybean protein does not meet the nutritional demands of young children or the monogastric animals such as swine and poultry. Without supplementation of synthetic amino acids, feed efficiency is compromised, cost of production is increased, and excess nitrogen excretion due to higher overall protein consumption has a detrimental affect on the environment. The quality of the soybean seed protein as measured by the balance of amino acids required for human and animal nutrition has not changed substantially through the years of selective breeding (2). Thus, there is an impetus to generate soybeans that have not only a higher protein concentration but also a desirable amino acid profile. Prolina soybean produces meal with 54% crude protein and contains a higher concentration of lysine, arginine, and leucine than normal cultivars with 48% crude protein (37). Broiler performance on

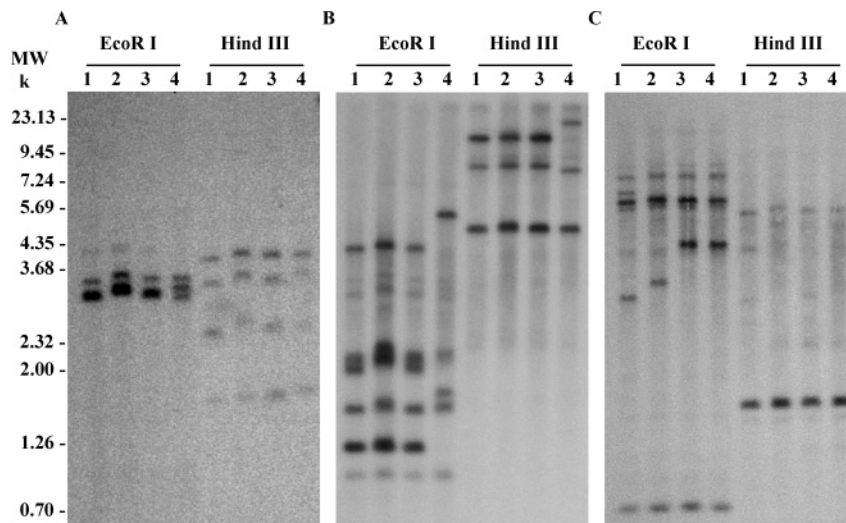


Figure 6. Southern blot analysis of soybean Bowman-Birk-type proteinase inhibitors. Genomic DNA from leaf tissue of Williams 82 (lane 1), BARC-6 (lane 2), LG00-13260 (lane 3), and PI 427138 was individually digested with either *EcoRI* or *HindIII*, fractionated on an 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with a ^{32}P labeled BBI A (A) or KTI (B) and BBI C-II (C) and then subjected to autoradiography. Molecular weight markers in kilobases are shown on the left side of the figure.

diets containing Prolinea soybeans was superior presumably due to a higher concentration of these essential amino acids (37). In this study, we have determined that the amino acid composition of the high-protein soybean lines is not significantly different from that of Williams 82. Additionally, the methionine content of high-protein lines, a quality indicator, was comparable to that of Williams 82, demonstrating a lack of relationship between protein and methionine content that corroborates the earlier findings (38–40). Even though subtle changes in the concentration of a few essential amino acids were observed, it remains to be seen if these changes could improve the performance of monogastric animals.

Glycinin and β -conglycinin comprise 70% of the seed storage protein and as such account for both quantity and quality of the seed protein (31, 32). Although neither protein is particularly abundant in the sulfur amino acids, the 11S glycinin is superior in its content of cysteine and methionine (33, 34). Therefore, it will be desirable to develop high-protein soybean lines that preferentially accumulate 11S proteins. An earlier study has shown that the high-protein cultivars accumulate higher amounts of both glycinins and β -conglycinins (41). Our study confirms this observation and provides additional information on the identity of individual proteins.

Essentially two avenues of improving soybean protein are currently being utilized. The first is through traditional breeding using high-protein germplasm and the second is the use of biotechnology. Quantitative trait loci (QTL) have been tentatively associated with total protein accumulation, seed storage proteins 11S glycinin and 7S β -conglycinin, and cysteine and methionine content (36, 42–44). In-depth information on QTLs controlling these quantitative traits will accelerate the process of enhancing seed protein content and quality. Genetic engineering also holds a promise to improve the quality of soybean seed proteins. Heterologous proteins rich in limiting amino acids such as methionine and cysteine have been expressed in soybeans to increase the nutritive value of the seed (45). However, the accumulation of the sulfur-rich proteins often occurs at the expense of endogenous sulfur-rich proteins in soybeans, indicating a limitation in the availability of sulfur amino acids in developing seeds. Although the metabolic pathways of major seed components, protein, oil, and carbohydrates, have been individually characterized, the mechanism by which these

compounds are partitioned remains mostly unknown. Additionally, knowledge of the mechanisms of nutrient assimilation, amino acid synthesis, protein accumulation, and carbon partitioning will facilitate the production of agronomically viable plants yielding high-protein seed with desirable amino acid compositions.

ACKNOWLEDGMENT

We thank John Bennett for critically reading the manuscript.

LITERATURE CITED

- (1) Durham, D. The United Soybean Board's Better Bean Initiative: building United States soybean competitiveness from the inside out. *Agric. Biol. Forum* **2003**, *6*, 23–26.
- (2) Mahmoud, A. A.; Natarajan, S. S.; Bennett, J. O.; Mawhinney, T. P.; Wiebold, W. J.; Krishnan, H. B. Effect of six decades of selective breeding on soybean protein composition and quality: A biochemical and molecular analysis. *J. Agric. Food Chem.* **2006**, *54*, 3916–3922.
- (3) Nelson, R. L.; Guzman, P. S.; Neece, D. J. *QTL associated with seed protein concentration from three sources of high protein*. VII World Soybean Research Conference: Foz Do Iguassu, Brazil, 2004; pp 56.
- (4) Burton, J. W. Breeding soybeans for improved protein quantity and quality. In *Proceedings of the World Soybean Research Conference III, Ames, IA*; Shibles, R., Ed.; Westview Press: Inc. Boulder, CO, 1984; pp 361–367.
- (5) Burton, J. W. Quantitative genetics: results relevant to soybean breeding. In *Soybeans: Improvement, Production, and Uses*; Wilcox, J. R., Ed.; American Society of Agronomy: Madison, WI, 1987; pp 211–247.
- (6) Wilcox, J. R.; Cavins, J. F. Backcrossing high-seed protein to a soybean cultivar. *Crop Sci.* **1995**, *35*, 1036–1041.
- (7) Helms, T. C.; Orf, J. H. Protein, oil, and yield of soybean lines selected for increased protein. *Crop Sci.* **1998**, *38*, 707–711.
- (8) Gizlice, Z.; Carter, T. E.; Burton, J. W. Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* **1994**, *34*, 1143–1151.
- (9) Gizlice, Z.; Carter, T. E., Jr.; Gerig, T. M.; Burton, J. W. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. *Crop Sci.* **1996**, *36*, 753–765.
- (10) Sneller, C. H. Pedigree analysis of elite soybean lines. *Crop Sci.* **1994**, *34*, 1515–1522.

- (11) Kisha, T. J.; Diers, B. W.; Hoyt, J. M.; Sneller, C. H. Genetic diversity among soybean plant introductions and North American germplasm. *Crop Sci.* **1998**, *38*, 1669–1680.
- (12) Brim, C. A.; Burton, J. W. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. *Crop Sci.* **1979**, *19*, 494–498.
- (13) Wehrman, V. K.; Fehr, W. R.; Cianzio, S. R.; Cavins, J. F. Transfer of high-seed protein to high-yielding soybean cultivars. *Crop Sci.* **1987**, *27*, 927–931.
- (14) Holbrook, C. C.; Burton, J. W.; Carter, T. E. Evaluation of recurrent restricted index selection for increasing yield while holding seed protein constant in soybean. *Crop Sci.* **1989**, *29*, 324–329.
- (15) Hartwig, E. E. Registration of soybean high-protein germplasm line D76-8070. *Crop Sci.* **1990**, *30*, 764–765.
- (16) Leffel, R. C. Registration of high-protein soybean germplasm lines BARC-6, BARC-7, BARC-8, and BARC-9. *Crop Sci.* **1992**, *32*, 502.
- (17) Burton, J. W.; Carter, T. E.; Wilson, R. F. Registration of Prolina soybean. *Crop Sci.* **1999**, *32*, 294–295.
- (18) Cober, E. R.; Voldeng, H. D. Developing high-protein, high-yield soybean populations and lines. *Crop Sci.* **2000**, *40*, 39–42.
- (19) Anand, S. C.; Shannon, J. G.; Wrather, J. A.; Arelli, P. R.; Sleper, D. A.; Young, L. D. Registration of S97-1688. Soybean germplasm line high in protein content and resistant to soybean cyst nematode. *Crop Sci.* **2004**, *44*, 698–699.
- (20) Bernard, R. L.; Juvik, G. A.; Nelson, R. L. *USDA Soybean Germplasm Collection Inventory*; International Agricultural Publications: 1989; Vol. 2, INTSOY Series No. 31.
- (21) Wilcox, J. R. Soybean protein and oil quality. In *Proceedings of the 4th World Soybean Research Conference, Buenos Aires, Argentina*; Pascale, A. J., Ed.; Assoc. Argentina de la Soja (AASOJA): Buenos Aires, 1989; pp 28–39.
- (22) Natarajan, S.; Xu, C.; Caperna, T. J.; Garrett, W. M. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal. Biochem.* **2005**, *342*, 214–220.
- (23) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.
- (24) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685.
- (25) Newsholme, S. J.; Maleeft, B. F.; Steiner, S.; Anderson, N. L.; Schwart, L. W. Two-dimensional electrophoresis of liver proteins: Characterization of a drug-induced hepatomegaly in rats. *Anal. Biochem.* **2000**, *21*, 2122–2128.
- (26) Krishnan, H. B.; Natarajan, S. S.; Mahmoud, A. A.; Bennett, J. O.; Krishnan, A. H.; Prasad, B. J. Assessment of indigenous Nepalese soybean as a potential germplasm resource for improvement of protein in North American cultivars. *J. Agric. Food Chem.* **2006**, *54*, 5489–5497.
- (27) Hajduch, M.; Ganapathy, A.; Stein, J. W.; Thelen, J. J. A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiol.* **2005**, *137*, 1397–1419.
- (28) Saghai Maroof, M. A.; Soliman, K. M.; Jorgensen, R. A.; Allard, R. W. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 8014–8019.
- (29) Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, B. H.; Scallan, B. J.; Fischer, R. L.; Sims, T. L.; Drews, G. N.; Goldberg, R. B. Characterization of the glycinin gene family. *Plant Cell* **1989**, *1*, 313–328.
- (30) Harada, J. J.; Baker, S. J.; Goldberg, R. B. Soybean β -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and post-transcriptional processes. *Plant Cell* **1989**, *1*, 415–425.
- (31) Nielsen, N. C. Soybean seed composition. In *Soybean: Genetics, Molecular Biology, and Biotechnology*; Verma, D. P. S., Shoemaker, R. C., Eds.; CAB: Tucson, 1996; pp 127–163.
- (32) Krishnan, H. B. Biochemistry and molecular biology of soybean seed storage proteins. *J. New Seeds* **2000**, *2*, 1–25.
- (33) Thanh, V. H.; Shibasaki, K. Major proteins of soybean seeds. Reconstitution of β -conglycinin from its subunits. *J. Agric. Food Chem.* **1978**, *26*, 695–698.
- (34) Thanh, V. H.; Shibasaki, K. Major proteins of soybean seeds. Subunit structure of β -conglycinin. *J. Agric. Food Chem.* **1978**, *26*, 692–695.
- (35) Orf, J. H.; Chase, K.; Jarvik, T.; Mansur, L. M.; Cregan, P. B.; Adler, F. R.; Lark, K. G. Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. *Crop Sci.* **1999**, *39*, 1642–1651.
- (36) Chung, J.; Babka, H. L.; Graef, G. L.; Staswick, P. E.; Lee, D. J.; Cregan, P. B.; Shoemaker, R. C.; Specht, J. E. The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* **2003**, *43*, 1053–1067.
- (37) Wilson, R. F. Developing agronomic high-protein soybeans. *Am. Soybean* **2001**, *September*, 14–15.
- (38) Burton, J. W.; Purcell, A. E.; Walter, W. M., Jr. Methionine concentration in soybean protein from populations selected for increased percent protein. *Crop Sci.* **1982**, *22*, 430–432.
- (39) Serretti, C.; Schapaugh, W. T., Jr.; Leffel, R. C. Amino acid profile of high-seed protein soybean. *Crop Sci.* **1994**, *34*, 207–209.
- (40) Wilcox, J. R.; Shibles, R. M. Interrelationships among seed quality attributes in soybean. *Crop Sci.* **2001**, *41*, 11–14.
- (41) Yaklich, R. W. β -Conglycinin and glycinin in high-protein soybean seeds. *J. Agric. Food Chem.* **2001**, *49*, 729–735.
- (42) Brummer, E. C.; Greaf, G. L.; Orf, J.; Wilcox, J. R.; Shoemaker, R. C. Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Sci.* **1997**, *37*, 370–378.
- (43) Panthee, D. R.; Pantalone, V. R.; Sams, C. E.; Saxton, A. M.; West, D. R.; Orf, J. H.; Killam, A. S. Quantitative trait loci controlling sulfur containing amino acids, methionine, and cysteine in soybean seeds. *Theor. Appl. Genet.* **2006**, *112*, 546–553.
- (44) Nichols, D. M.; Glover, K. D.; Carlson, S. R.; Specht, J. E.; Diers, B. W. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Sci.* **2006**, *46*, 834–839.
- (45) Krishnan, H. B. Engineering soybean for enhanced sulfur amino acid content. *Crop Sci.* **2005**, *45*, 454–461.

Received for review August 31, 2006. Revised manuscript received November 16, 2006. Accepted December 1, 2006. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may be suitable.

JF062497N