JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Proteomic Analysis of High Protein Soybean (*Glycine max*) Accessions Demonstrates the Contribution of Novel Glycinin Subunits

Hari B. Krishnan*,†,‡ and Randall L. Nelson§

†Plant Genetics Research Unit, Agricultural Research Service, U.S. Department of Agriculture, 108 Curtis Hall, Columbia, Missouri 65211, United States

‡Division of Plant Sciences, University of Missouri, Columbia, Missouri 65211, United States

§USDA-Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Department of Crop Sciences, 1101 W. Peabody Drive, University of Illinois, Urbana, Illinois 61801, United States

ABSTRACT: Limited biochemical information is available on soybean accessions that have seed protein content greater than 45% of the seed dry weight. SDS−PAGE analysis of seed proteins from nine soybean accessions revealed significantly higher amount of seed storage proteins in these accessions when compared with that of soybean cultivar Williams 82. High-resolution two-dimensional gel electrophoretic analysis of seed proteins revealed significant differences among several seed storage protein components in these accessions. A total of 51 protein spots were identified using peptide mass fingerprinting (MALDI-TOF MS). The contribution of these proteins to the overall protein content of the accessions was quantified using Delta2D image analysis software. Results showed that among the majority of the nine accessions, the largest difference in higher protein quantity was within the seed 11S storage globulins. The high protein trait from PI407788A was successfully transferred to an experimental line, LG99-469, demonstrating that this trait was transferable and robust.

KEYWORDS: High-protein soybean, glycinin, two-dimensional gel electrophoresis, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry

INTRODUCTION

The bulk of the world's vegetable protein is supplied by the seed of the soybean plant. The majority of that protein is used to formulate rations for domestic livestock and poultry, while the remaining is utilized for human consumption and in industrial applications (http://www.soystats.com/2009/Default-frames.htm). Seed protein concentration of commercial soybean cultivars calculated on a dry weight basis ranges from approximately 36 to 42% depending on the genotype, location, and growth conditions.[1] While this amount of seed protein is high compared to other seed crops, an ongoing concerted research effort has been established to further increase soybean seed total protein concentration. This effort is essential since the demand for quality soy seed protein and its products, has increased steadily for decades and is not expected to decline (http://www.ers.usda.gov/briefing/baseline/trade.htm).

Cultivated soybean (*Glycine max* [L.] Merr.) seed protein consists mainly, sometimes 80% or higher, of storage proteins.[2] Almost 80 storage proteins have been identified in soybean seed,[3,4] and each falls into one of four basic categories: albumins (water-soluble), globulins (salt soluble), prolamins (alcohol soluble), and glutelins (weak acid/weak base soluble). Cultivated soybean seed storage proteins consist primarily of two major storage protein complexes, glycinin and β-conglycinin,[5−7] which fall into the globulin category (soluble in dilute aqueous-salt solutions). The other seed storage proteins only account for a minor portion of the total seed protein content since globulins can be 60 to 80% of the total seed protein.

Several soybean accessions with seed protein concentrations between 45 and 50% have been reported.[8−10] These accessions can be useful sources of high protein, but developing high protein cultivars has been difficult because of the strong negative relationship between seed protein concentration and seed yield.[11] Identifying and characterizing differences in those protein components in additional high protein soybean accessions could facilitate ongoing efforts to improve both the quantity and quality of soybean seed proteins. Hence, the goal of this project was to identify and characterize the protein components of selected high protein accessions.

MATERIALS AND METHODS

Plant Material. Nine high protein soybean accessions were used in this study. All of the accessions analyzed in this work have been used in our research and breeding program to develop experimental lines with high protein concentrations. Some lines were selected on the basis of data collected by Korczak.[12] The protein content of the soybean accessions reported in this study is based on seed dry weight. In replicated tests over two years, Wisconsin Black and PI 82278 averaged 47.4 and 49.0%, respectively. Sioux, PI 340031A, PI 407788A, PI 423948A, PI 437088A, PI 437461, and PI 445845 were selected on the basis of general germplasm evaluations conducted at Urbana, IL or Lamberton, MN.[13−15] These accessions were also evaluated in a replicated, five location experiment that was conducted over two years. In the germplasm evaluation test, the protein concentrations ranged from 48.4 to 52.2%, and in the second experiment, the range was 46.1 to 50.8%.

Received: November 8, 2010
Revised: January 25, 2011
Accepted: February 3, 2011
Published: February 23, 2011

Table 1. Soybean Protein Band Quantification by Delta2D Image Analysis[a]

	seed identifier #																			
	407788A/		548433/		437088A/		548141/		423948A/		82278/		340031A/		445845/		437461/	
band	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)	W82	± (n = 3)
1	0.962	0.028	1.213	0.035	1.651	0.083	1.913	0.071	2.191	0.099	1.202	0.015	1.892	0.042	1.758	0.010	1.325	0.039
2	1.139	0.005	1.164	0.001	1.790	0.023	1.858	0.038	2.045	0.110	1.217	0.020	1.824	0.044	1.604	0.027	1.227	0.044
3	1.317	0.077	1.179	0.063	1.431	0.041	1.275	0.051	1.552	0.065	1.026	0.052	1.481	0.019	1.436	0.043	1.219	0.030
4	1.059	0.019	1.092	0.032	2.140	0.053	2.003	0.030	2.286	0.017	1.538	0.016	2.033	0.096	2.292	0.034	1.042	0.017
5	1.066	0.018	1.644	0.023	1.998	0.028	1.944	0.027	2.604	0.022	1.082	0.049	2.110	0.079	1.962	0.037	1.677	0.025
6	1.331	0.037	1.553	0.079	2.304	0.095	2.538	0.085	0.000	0.000	1.277	0.053	2.755	0.096	2.300	0.173	1.763	0.170
7	1.162	0.009	1.719	0.019	1.865	0.020	1.724	0.010	2.462	0.062	1.194	0.037	2.263	0.042	1.838	0.023	1.667	0.004
8	0.791	0.032	0.925	0.034	1.244	0.100	1.567	0.043	1.518	0.072			1.314	0.031	1.207	0.011	1.067	0.055
9	1.112	0.058	1.822	0.032	2.205	0.022	2.134	0.054	1.932	0.024	1.247	0.020	2.611	0.014	1.943	0.056	1.848	0.060
10	0.953	0.035	1.437	0.048	1.712	0.052	1.632	0.018	2.273	0.044	0.971	0.069	1.703	0.022	1.656	0.049	1.383	0.045
11	1.464	0.027	1.499	0.008	1.881	0.047	2.352	0.182	2.680	0.077	2.281	0.151	2.935	0.110	2.260	0.091	1.720	0.103
12	1.132	0.040	1.849	0.089	2.585	0.084	2.408	0.098			1.615	0.049	2.775	0.090	1.840	0.019	2.243	0.075

[a] Percent volume ratio [high protein line]/[W82] generated from 1D % band volume collected using Delta2D image analysis software and the image from Figure 1. Error is presented as the SEM from three separate fractionations and three separate separations analyzed in Delta2D using three separate images, fused using global warping and exact spot matching. Band numbers correspond to those designated in Figure 1.

Reagents. Urea, thiourea, methanol, ammonium acetate, β-mercaptoethanol, glycerol, mineral oil, agarose (low EEO), and buffer reagents were obtained from Fisher Scientific (Pittsburgh, PA, USA). Phenol (Tris-equilibrated), 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 3-(4-heptyl)phenyl-3-hydroxypropyl) dimethylammonio-propanesulfonate (C7BzO), 2-hydroxyethyl disulfide (2-HED), dithiothreitol (DTT), iodoacetamide, and protease inhibitor cocktail (P-9599) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Acrylamide, bis-acrylamide, ammonium persulfate, *N,N,N′,N′*-tetramethylethylenediamine (TEMED) and Coomassie Brilliant Blue G-250 were obtained from BioRad (BioRad Laboratories, Inc., Hercules, CA, USA). SYPRO Ruby and the EZQ protein quantification kit were obtained from Molecular Probes (Invitrogen, Carlsbad, CA, USA). IPG strips, IPG Buffer 4-7 and 3−10, Cy3, and Cy5 were obtained from GE Healthcare (Piscataway, NJ, USA).

One-Dimensional Electrophoresis (1-DE). For 1-DE analysis, mature, dried soybean seeds were ground into a fine powder using a chilled mortar and pestle and 20 mg placed into a tube for extraction using 1 mL of 20 mM Tris-Cl, pH 6.8, containing protease inhibitor cocktail. Protein extraction was carried out for 15 min, with vigorous shaking, at room temperature. The solution was clarified with centrifugation at 16,100*g* for 10 min. The supernatant was removed, and $CaCl_2$ was added to a final concentration of 10 mM. The solution was left at room temperature for 5 min followed by centrifugation at 16,100*g* for 10 min. The resulting pellet was resuspended in 1 mL of SDS−PAGE sample buffer. One-dimensional separation followed the method of Laemmli[16] using 13.5% T gels run using Mini250 (GE Healthcare). Separation was achieved with a constant 20 mA per gel and run time of 1.5 h. Gels were removed from the cassette and placed immediately in Coomassie staining solution (20% ethanol, 8% ammonium sulfate, 1.6% phosphoric acid, and 0.4 g/L Coomassie Brilliant Blue G-250). Typically, 20−40 μg of protein[17] from each sample was loaded per well.

Two-Dimensional Electrophoresis (2-DE). 2-DE analysis of soybean seed proteins were performed as previously described.[19,25,26]

Gel Staining. All 2-DE gels used in this study were stained with SYPRO Ruby unless otherwise stated. For SYPRO Ruby stained gels, 200 μg of each protein sample was run per gel and after electrophoresis, and the gels were fixed in 1:8:1 (ethanol/water/acetic acid) for 1 h, stained overnight using SYPRO Ruby, and processed according to the manufacturer's protocol prior to imaging. 2-DE gels, used for protein identification purposes, were immediately removed and fixed in 5:4:1 (methanol/water/acetic acid) for 1 h, followed by staining in Coomassie G-250 for 24 h.

Differential In-Gel Electrophoresis (DIGE). DIGE was performed in a manner identical to previous 2-D methodology; however, (i) the initial protein isolate was dissolved in 7 M urea, 2 M thiourea, 1% CHAPS, and 2% C7BzO and 30 mM Tris-Cl, pH 8.5 (without reducing agent) prior to dye labeling, and (ii) the exact protein concentration was determined using the EZQ system (Molecular Probes). For DIGE, a linear gradient 13 cm, 3−10 IPG was utilized, and 50 μg of each protein sample was labeled with 200 pmol of Cy3 or Cy5 (separately) following the manufacturer's protocol (GE Healthcare). After combining Cy3 and Cy5 samples, exactly 100 μg was loaded onto each strip using in-gel rehydration. The final strip rehydration volume was brought to 250 μL with 7 M urea, 2 M thiourea, 1% CHAPS, and 2% C7BzO with a final concentration of the following: 5% glycerol, 2.2% 2-HED, and 0.5% 3−10 IPG buffer. Again, the final concentration of DTT in each strip, optimized previously to be used in conjunction with 2-hydroxyethyl disulfide,[18] was adjusted precisely to 60 mM. Rehydration solutions were vortexed with moderate force and incubated on ice for 30 min prior to loading. Strips were then passively rehydrated with the entire rehydration solution containing the protein sample at 22 °C for 15 h protected from light prior to focusing. After two-dimensional separation, CyDye labeled DIGE gels were immediately removed from the cassette and

Table 2. Identification of Soybean Seed Proteins by MALDI-TOF-MS and Their Relative Concentration[a]

SID	protein identification	MO	PM	SC	Acc. #	407788A/ W82	*t* test (n = 3)	82278/ W82	*t* test (n = 3)	423948A/ W82	*t* test (n = 3)
1	β-conglycinin (α-subunit)	267	31	53	NCBI: gi|9967357	**0.84**	96.94	**1.00**	4.44	**1.03**	31.63
2	β-conglycinin (α′-subunit)	282	31	48	NCBI: gi|9967361	**0.86**	99.38	**1.06**	75.08	**1.10**	91.15
3	sucrose binding protein	63	13	28	NCBI: gi|548900	**1.28**	94.63	**0.99**	9.54	**1.23**	92.51
4	sucrose binding protein	61	12	28	NCBI: gi|548900	**1.50**	99.93	**1.18**	94.28	**1.26**	98.23
5	sucrose binding protein	109	14	28	NCBI: gi|548900	**1.47**	99.96	**1.25**	99.84	**1.24**	99.33
6	sucrose binding protein	113	16	31	NCBI: gi|548900	**1.01**	22.29	**0.94**	75.03	**0.87**	91.60
7	sucrose binding protein	83	16	30	NCBI: gi|548900	**0.33**	99.99	**0.34**	99.95	**0.23**	99.99
8	LEA-protein (related)	82	33	57	MSDB:Glyma10g07410.1	**1.28**	99.60	**1.11**	96.80	**1.41**	99.70
9	β-conglycinin (β-subunit)	188	19	43	NCBI: gi|63852207	**0.57**	99.97	**0.39**	99.98	**0.95**	43.12
10	β-conglycinin (β-subunit)	194	25	49	NCBI: gi|21465628	**0.37**	99.99	**0.35**	99.99	**0.52**	99.95
11	β-conglycinin (β-subunit)	255	26	52	NCBI: gi|21465628	**0.40**	99.99	**0.45**	99.99	**0.64**	99.98
12	β-conglycinin (β-subunit)	260	29	52	NCBI: gi|21465628	**0.60**	99.99	**0.65**	99.94	**0.87**	99.85
13	β-conglycinin (β-subunit)	*63400*	28	52	NCBI: gi|63852207	**1.00**	9.37	**1.08**	82.73	**1.27**	99.94
14	β-conglycinin (β-subunit)	247	28	49	NCBI: gi|21465628	**1.62**	99.97	**2.00**	99.99	**1.95**	99.86
15	β-conglycinin (α-subunit)	143	19	33	NCBI: gi|9967357	**0.56**	99.56	**0.43**	99.53	**0.80**	84.64
16	β-conglycinin (α-subunit)	142	20	36	NCBI: gi|9967357	**0.84**	91.78	**0.55**	99.67	**0.77**	94.83
17	β-conglycinin (α-subunit)	215	21	40	NCBI: gi|9967357	**0.62**	99.99	**0.59**	99.85	**0.85**	98.92
18	glycinin (A3B4)	76	11	22	NCBI: gi|33357661	**0.41**	99.70	**0.41**	99.91	**1.00**	1.55
19	glycinin (A3B4)	86	10	22	NCBI: gi|33357661	**0.97**	40.08	**1.21**	92.99	**1.20**	98.32
20	glycinin (A3B4)	70	9	18	NCBI: gi|33357661	**1.11**	84.16	**1.33**	98.90	**1.20**	93.79
21	glycinin (A3B4)	70	8	19	NCBI: gi|33357661	**2.85**	99.90	**2.26**	99.80	**2.51**	99.70
22	glycinin	126	10	39	NCBI: gi|6015515	**0.94**	58.22	**0.85**	87.99	**0.21**	99.99
23	glycinin	122	10	39	NCBI: gi|6015515	**1.89**	99.87	**1.56**	99.54	**0.49**	99.98
24	glycinin (A2B1a)	96	10	21	NCBI: gi|121277	**0.65**	99.67	**0.73**	99.52	**0.80**	74.84
25	glycinin (A2B1a)	102	10	21	NCBI: gi|121277	**0.94**	52.49	**0.91**	73.96	**0.98**	28.39
26	glycinin (A2B1a)	72	10	23	NCBI: gi|121277	**0.91**	79.91	**0.99**	6.74	**1.06**	68.84
27	glycinin (A1aBx) precursor	63	8	18	NCBI: gi|121276	**1.09**	69.99	**1.09**	92.74	**1.02**	39.24
28	proglycinin (A1ab1b)	76	9	19	NCBI: gi|15988117	**0.99**	23.84	**1.14**	85.84	**1.00**	8.38
29	proglycinin (A1ab1b)	79	8	19	NCBI: gi|15988117	**1.02**	18.48	**1.25**	95.64	**1.82**	99.71
30	glycinin (A1aBx) precursor	63	9	20	NCBI: gi|121276	**1.44**	99.59	**1.25**	97.46	**0.66**	98.78
31	proglycinin (A1ab1b)	96	10	20	NCBI: gi|15988117	**1.86**	99.98	**1.78**	99.99	**1.78**	99.96
32	glycinin (A1aBx) precursor	204	7	16	NCBI: gi|121276	**4.67**	99.99	**4.66**	99.99	**4.57**	99.93
33	proglycinin (A1ab1b)	83	10	20	NCBI: gi|15988117	**7.73**	99.98	**8.26**	99.99	**8.00**	97.56
34	proglycinin (A1ab1b)	85	10	20	NCBI: gi|15988117	**3.89**	99.99	**4.64**	99.99	**3.77**	98.12
35	35 kDa seed maturation protein	109	16	45	NCBI: gi|4102190	**1.25**	96.40	**1.11**	74.70	**0.30**	99.80
36	glycinin	96	9	39	NCBI: gi|6015515	**0.53**	96.14	**0.43**	97.98	**0.46**	95.01
37	glycinin	*77868*	5	22	NCBI: gi|6015515	**0.49**	82.93	**0.44**	96.39	**0.56**	76.05
38	lectin (SBA), precursor	*608432*	12	35	NCBI: gi|126151	**1.48**	97.40	**0.07**	100	**0.87**	69.80
39	lectin (SBA), precursor	*311000*	11	35	NCBI: gi|126151	**0.75**	87.70	**0.22**	100	**0.75**	86.00
40	lectin (SBA), precursor	*77422*	11	39	NCBI: gi|126151	**1.48**	97.40	**0.22**	100	**0.79**	70.00
41	unidentified protein					**1.23**	99.90	**1.16**	99.60	**1.29**	99.80
42	Kunitz trypsin inhibitor, chain A	78	15	77	NCBI: gi|3318877	**0.72**	100	**0.69**	100	**0.91**	90.80
43	Kunitz trypsin inhibitor, precursor	6358	6	32	NCBI: gi|125722	**0.82**	98.90	**0.95**	42.00	**0.72**	99.30
44	glycinin (A2B1a) precursor	60	4	26	NCBI: gi|169967	**1.10**	48.47	**0.39**	99.99	**0.86**	54.00
45	glycinin (A1aBx) precursor	*99898*	7	17	NCBI: gi|121276	**1.17**	84.61	**0.81**	70.15	**1.21**	88.98
46	glycinin (A2B1a) precursor	60	5	27	NCBI: gi|169967	**1.18**	98.63	**0.54**	97.34	**1.24**	94.43
47	glycinin (A1aBx) precursor	61	7	27	NCBI: gi|121276	**1.24**	99.68	**1.04**	16.09	**1.29**	98.34
48	glycinin (A1aBx) precursor	*53758*	7	13	NCBI: gi|121276	**0.84**	99.28	**0.85**	87.63	**0.85**	97.06
49	seed maturation protein PM31	91	9	46	NCBI: gi|4838149	**0.95**	43.10	**1.12**	85.40	**1.04**	19.20
50	Bowman-Birk proteinase inhibitor	*900*	4	40	UniProtKBP01064	**1.06**	92.70	**1.29**	91.10	**1.21**	97.10
51	glycinin (A5A4B3) precursor	*737*	7	8	NCBI: gi|121279	**1.35**	99.14	**1.55**	99.91	**0.05**	99.99

[a] Data presented here is from high protein lines (PI407788A, PI82278, and PI423948A) in comparison with Williams 82. The table denotes the ratio of % spot volumes collected from the image analysis of each compared to Williams 82 using Delta2D software. Ratios greater than 1 indicate a greater % spot volume compared to that of Williams82, whereas those ratios less than 1 indicate less. MO, mascot score; PM, peptide matched; SC, sequence coverage.

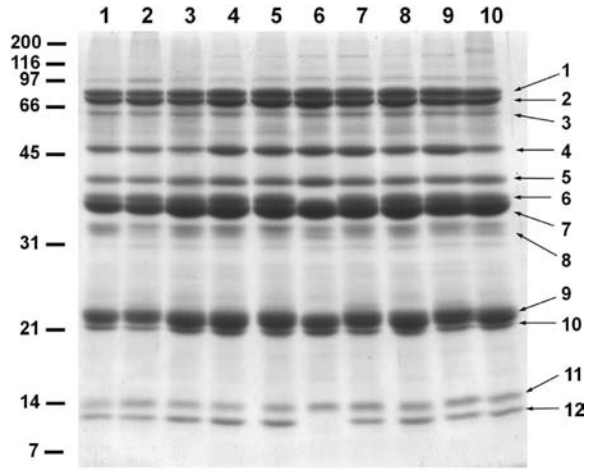

Figure 1. 1-DE separation of soybean seed proteins extracted from nine high protein lines and Williams 82 fractionated with 10 mM calcium. Band numbers shown on the right correspond to those listed in Table 1. Lane 1, Williams 82; lane 2, PI407788A; lane 3, PI548433; lane 4, PI437088A; lane 5, PI548141; lane 6, PI423948A; lane 7, PI82278; lane 8, PI340031A; lane 9, PI445845; lane 10, PI437461. The sizes of protein markers in kilodaltons are shown at the left of the figure.

soaked in 2:8 (methanol/water) for 1 h, protected from direct light, prior to imaging.

Image Acquisition and Analysis. DIGE gels were scanned using an *Ettan*DIGE Imager v1.0, processed using ImageQuant TL, and analyzed for proteome differences using DeCyder Differential Image Analysis v5.01 (GE Healthcare) and Delta2D v3.6 (Decodon, Greifswald, Germany). Gels stained with SYPRO Ruby were scanned using a Fuji FLA5000 at 473 nm excitation with a 510 nm long pass blue filter and optimized for image quality using Fuji Multi Gauge. Images from SYPRO Ruby stained gels were analyzed for proteome differences using Delta2D image analysis software. Coomassie stained gels were destained with multiple changes of ultrapure H_20 to remove the background and scanned using a HP Scanjet 5470c controlled through Adobe Photoshop. DeCyder analysis software was set to default parameters for normalization using maximum volume and a differential detection threshold of 2-fold using the spot volume ratio. DeCyder was primarily used for differential image acquisition (2 channel overlay), while Delta2D provided the % volume data used throughout this article.

Delta2D parameters were set to maximize spot detection (using *global* image warping and *exact* spot matching); hence, very little background subtraction was used throughout. Percent spot volume ratio differences were noted in the spot quantitation table, and cutoffs were determined according to fold changes, keeping nearly all protein spots for calculation/comparison purposes. Here, the relative quantity of the spot, excluding the background, gave a more accurate determination of the % change since the total quantity of all spots on the gel is 100%. Band and spot % volume ratio differences were calculated within each comparison; high protein line (n = 3)/Williams 82 (n = 3). Error is reported as the standard error of the mean in Table 1 and reported as the Student's t test in Table 2.

Protein Identification. A small gel piece of each protein for identification was excised with a 1.5 mm Spot Picker (The Gel Company, San Francisco, CA, USA) from a Coomassie G-250 stained gel, washed briefly in distilled water, and then destained completely in a 50% (v/v) solution of acetonitrile containing 25 mM ammonium bicarbonate. After a 100% acetonitrile wash, the protein contained in the acrylamide gel was subjected to digestion using 20 μL (10 μg/mL) of modified porcine trypsin (Promega, Madison, WI, USA) in 25 mM ammonium bicarbonate. Peptides resulting from the tryptic digestion were analyzed using a Voyager DE-STR (Applied Biosystems, Framingham, MA, USA) matrix-assisted-laser-desorption-time-of-flight mass spectrometer (MALDI-TOF). The peptides were cocrystallized with *R*-cyano-4-hydroxycinnamic acid matrix on a MALDI-TOF MS/MS plate, briefly dried, and ionized using a 337 nm nitrogen laser operating at 20 Hz. Trypsin autolysis peaks of charge/mass ratios 842.51, 1045.56, and 2211.10 served for internal calibration. Peptide mass searches were performed via Mascot (Matrix Sciences, http://www.matrixscience.com) and/or Protein Prospector (University of California-San Francisco, http://prospector.ucsf.edu) primarily using the NCBI nonredundant *Glycine max* database and UniProtKB *Glycine max* database, both with a peptide mass tolerance of 20 ppm.

RESULTS AND DISCUSSION

Previously we had devised a simple and rapid procedure to enrich storage proteins from soybean seeds using calcium precipitation.[19] We exploited this simple procedure to monitor changes in the storage protein composition of all nine accessions compared to that of Williams 82, a very common U.S. cultivar with about 40% protein content (dry weight). 1-DE separation of soybean seed proteins, based on per volume of equal extract, from all nine high protein lines and Williams 82, showed that all nine PI had more visible amounts of many of the fractionated proteins (Figure 1). The largest differences among them were within the seed storage globulins. Band quantification (% volume) data collected using Delta2D image analysis software confirmed this, and those values were used to calculate the ratio of [high protein line]/[W82] for comparison (Table 1). The data showed that many of those selected proteins were 1.5- to nearly 3-fold higher when compared with the control cultivar, Williams 82. While some proteins were relatively unchanged, the majority of this subset of proteins showed a significant increase in most of the accessions.

The 12 major 1D bands found in soybean extract after fractionation with calcium are listed in Table 1 and correspond to those shown on the right of Figure 1. We had previously determined that of these 12 major bands, the overwhelming majority included the seed storage globulins, glycinin and β-conglycinin. Glycinin, accounting for roughly 40 to 60% of the total seed protein, is a hexameric protein composed of the G1, G2, G3, G4, and G5 subunits (approximately 56 kDa, 54 kDa, 54 kDa, 64 kDa, and 58 kDa, respectively), all of which consist of one acidic and one basic chain (approximately 37 to 44 kDa and 17 to 22 kDa, respectively).[20,21] β-Conglycinin, accounting for roughly 30 to 40% of the total seed protein, is a trimeric glycoprotein composed of three subunits, α, α', and β (approximately 76 kDa, 72 kDa, and 53 kDa, respectively).[7,22−24] Approximately 90% of the globulins can be fractionated using 10 mM calcium,[19] simplifying a comparison among the seed storage globulins.

Further analysis of all nine accessions, along with Williams 82, using 2-DE separation of the extracted total seed proteins demonstrated that the seed storage globulins were primarily responsible for the increase in the total amount of protein among the high protein accessions. For this analysis, an equal amount of protein from each high protein line was separated and compared with an equal amount of protein from Williams 82. Looking at the fusion of all nine gel images overlaid with that of Williams 82, it became evident that many protein spots are significantly higher in amount. However, the picture obtained from the fusion of all nine images was complex. Therefore, we focused our attention

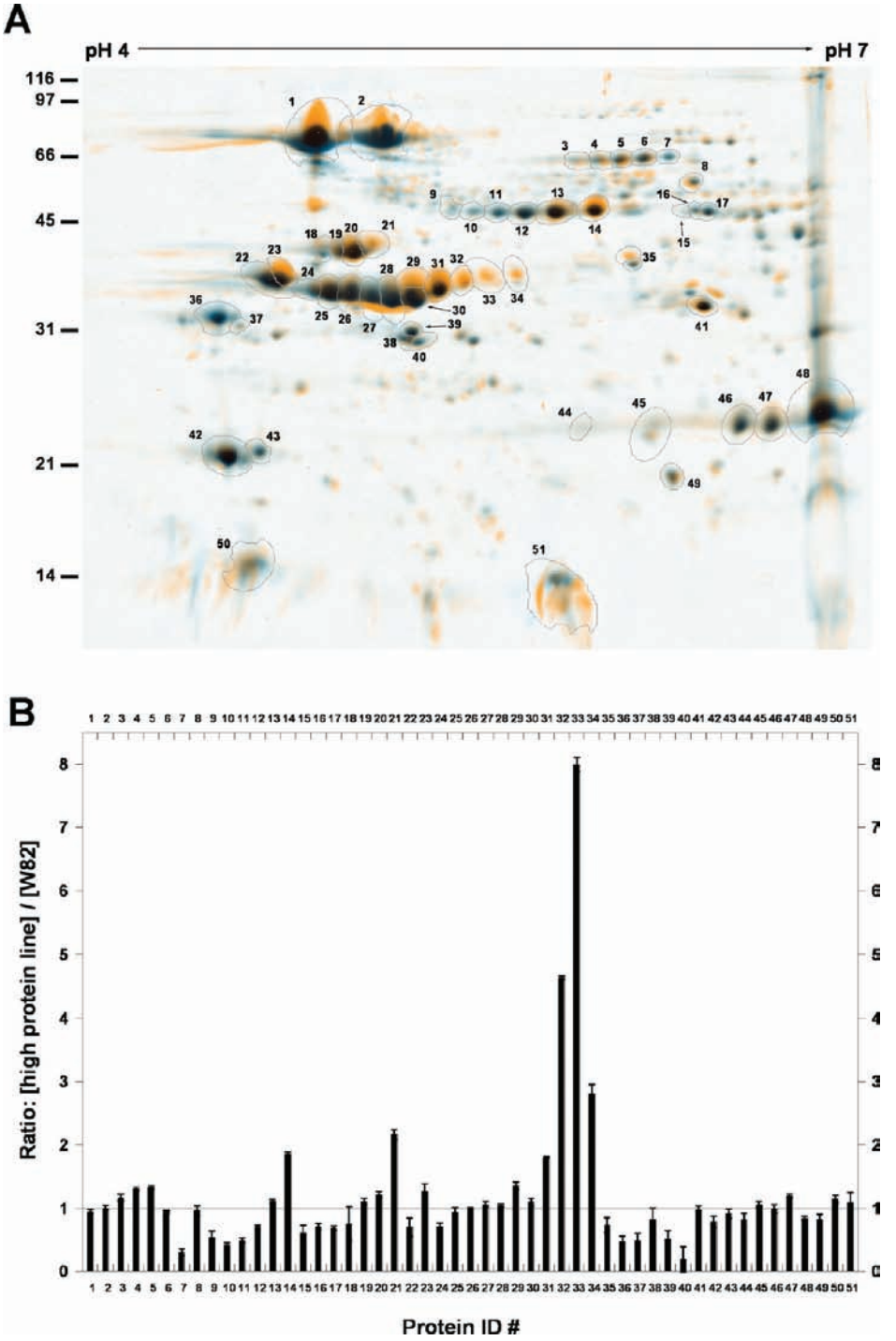

Figure 2. 2-DE separation of seed proteins extracted from three high protein soybean lines and Williams 82. Equal volumes of proteins from each high protein line were separated using 2-DE and compared with an equal volume of W82 protein. Panel A shows the fusion of all nine images (three high protein lines, run in triplicate) compared with that of W82 (run in triplicate). Orange color denotes those protein spots that were found to be higher in % spot volume compared with W82, whereas the blue color denotes those proteins that were found to be lower in % spot volume. IEF separation was from pH 4−7, and molecular weight markers are designated in kDa. Panel B shows the ratios of % volume [high protein line]/[W82] of those 51 spots generated from % spot volume data collected using Delta2D image analysis software and the three high protein lines combined.

on three (PI 407788A, PI 82278, and PI 423948A) of those nine PI that were representative of the group and did a more complete analysis using 2-DE in conjunction with high-resolution image analysis software. Information regarding these accessions, general narratives, pedigree narratives, collection site descriptions, and source histories was obtained from USDA Germplasm Resources Information Network (http://www.ars-grin.gov/cgi-bin/npgs/acc/). PI 407788A originated from South Korea, has a determinate growth habit, belongs to maturity group IV, has yellow seed coat color, and yields 2.06 megagrams per hectare (Mg/ha) at 13% seed moisture. PI 82278 also originated from South Korea but unlike PI 407788A has an indeterminate growth

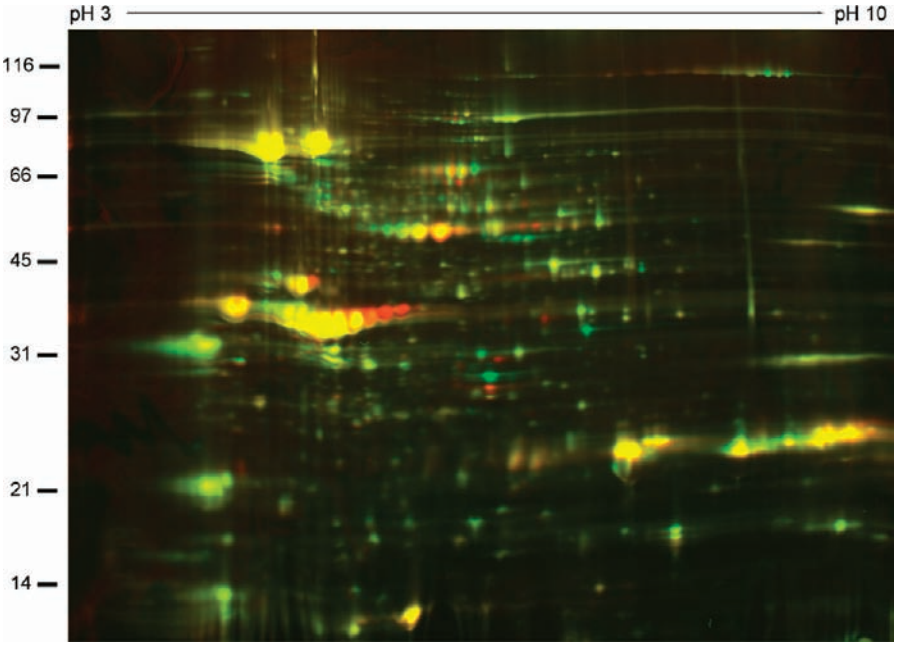

Figure 3. DIGE analysis of LG99-469 as compared to Williams 82. Image shown is the overlay image of LG99-469 (labeled with fluorescent Cy3; red) and Williams 82 (labeled with fluorescent Cy5; green). Those proteins present in LG99-469 that are not equal in % volume to W82 show up bright red, and those proteins not present in LG99-469 that are not equal in % volume to W82 show up bright green. Yellow spots denote those proteins equal in % spot volume. IEF separation was from pH 3−10, and the molecular weight markers are designated in kDa.

habit, belongs to maturity group III, and has a black seed coat color. It yields 1.20 Mg/ha, which is significantly lower than PI 407788A. In contrast to these two soybean accessions which originated from South Korea, PI 423948A was collected from Japan. It exhibits an indeterminate growth habit, yellow seed coat color, belongs to maturity group II, and yields 1.94 Mg/ha. An equal volume of protein extract from each of these three high protein lines was separated using 2-DE and compared with an equal volume of Williams 82 protein extract. The fusion of all 12 images (three high protein lines, run in triplicate) compared with that of Williams 82 (run in triplicate) demonstrates clearly that several of the seed storage globulins are primarily responsible for the increase in amount of total protein (Figure 2A). The orange color denotes those protein spots that were found to be higher in % spot volume compared with that of Williams 82, whereas the blue color denotes those proteins that were found to be lower in % spot volume. The numbered protein spots designate those 51 proteins that are known to include the many subunits of seed storage globulins and those proteins that when combined with the globulins make up roughly 50% of the total separated seed protein. Figure 2B shows the calculated ratios of % volume [high protein lines]/[Williams 82] of those 51 spots generated from the % spot volume data collected using Delta2D image analysis software, when the data from all three high protein lines are combined. Ratios are the mean of all seed lines used and those greater than 1 indicate those % spot volumes that were higher (cumulatively) in all three high protein lines. Ratios less than 1 indicate those that were lower, however very few ratios deviate substantially from 1, indicating those protein spots are very similar in % volume with that of Williams 82. To identify these 51 spots, we excised a small gel piece corresponding to each spot and subjected it to trypsin digestion. The identity of the protein spots was examined by MALDI-TOF MS (Table 2). Protein spot numbers shown in Figure 2A also correspond to those presented in Table 2, where the identity of each protein is listed. Ratios of the % spot volumes generated from these separately analyzed images (Table 2), in conjunction with high-percentile *t* test values, clearly demonstrate the statistical significance of the results.

The graph of the ratios of % volume [high protein lines]/[Williams 82] of those 51 spots generated from % spot volume data collected (Figure 2B) showed us precisely which proteins were significantly responsible for the increase in total protein amount (Figure 2B). Clearly, the data indicates that protein numbers 21 (glycinin A3B4), 32 (glycinin A1aBx precursor), 34 (proglycinin (A1ab1b), and especially 33 (proglycinin A1ab1b) are substantially higher in % spot volume. Hence, it can be concluded that these four proteins are primarily responsible for the higher amount of protein in these three soybean accessions. Previous studies have also shown that some of the high-protein soybean lines contain more glycinin than normal-protein soybean lines.[10,25,26]

Having these results, we wanted to see how transferable these increases in those specific proteins were when a high protein line was crossed with a line targeted for protein increase. Using DIGE (differential gel electrophoresis) analysis, we took soybean seed from LG99-469, a breeding cross of one of the high protein lines used in this study (PI 407788A) with an F5, and compared it to Williams 82. The image shown (Figure 3) is the digital overlay image of LG99-469 (labeled with fluorescent Cy3; red) and Williams 82 (labeled with fluorescent Cy5; green), combined and separated together using the same 2-DE gel. Those proteins present in LG99-469 that are not equal in % volume to Williams 82 are bright red, those proteins not present in LG99-469 that are not equal in % volume to Williams 82 are bright green, and those proteins equal in % spot volume between the two are yellow. Clearly, when compared to the image in Figure 2A, one can see that the same proteins higher in quantity in PI 407788A are the

Table 3. Identification of Protein Spots 33 and 34 as Glycinin by MALDI-TOF MS

SID	proteins identified (*Glycine max*)	accession number NCBInr	Mascot score (25 ppm)	sequence coverage	theoretical/ experimental MW (Da)	theoretical/ experimental pI	peptides matched
33	chain A, crystal structure of soybean proglycinin A1ab1b homotrimer	gi|15988117	83	20%	54,047/39,500	5.78/5.65	REQPQQNECQIQK KLNALKPDNRI RPSYTNGPQEIYIQQGKG RGQSSRPQDRH RRFYLAGNQEQEFLKY RFYLAGNQEQEFLKY KYQQEQGGHQSQKG KNLQGENEGEDKGAIVTVKG KDKHCQRPRG KHCQRPRG
34	chain A, crystal structure of soybean proglycinin A1ab1b homotrimer	gi|15988117	85	20%	54,047/40,000	5.78/5.72	REQPQQNECQIQK KLNALKPDNRI RPSYTNGPQEIYIQQGKG RGQSSRPQDRH RRFYLAGNQEQEFLKY RFYLAGNQEQEFLKY KYQQEQGGHQSQKG KNLQGENEGEDKGAIVTVKG KDKHCQRPRG KHCQRPRG

same proteins higher in quantity in LG99-469. This demonstrates that the trait is transferable and that the trait is robust, as indicated by the intensity of the red spots. We also isolated protein spots 33 and 34 from LG99-469 and subjected them to MALDI-TOF MS analysis. Using Mascot, the empirically determined mass-to-charge ratios of peptides were compared to peptides of known proteins stored in the National Center for Biotechnology Information nonredundant database. The result of this analysis (Table 3) confirmed the significant homology of protein spots 33 and 34 to the soybean proglycinin A1ab1b homotrimer. Our results demonstrate that the high protein content is mainly attributable to greater accumulation of specific components (spots 21, 32, 33, and 34) of the 11S glycinin.

Several of the high protein soybean accessions have been exploited in breeding programs to transfer the high protein trait to high-yielding lines.[12] There exists a negative correlation between seed protein and oil content[27] and seed protein and yield,[28] which has hampered the development of both high protein and high yielding soybean cultivars. Soybean seed protein content is influenced by the environment and its interaction with various genetic loci that control the high protein trait. Previous studies have identified several quantitative trait loci (QTLs) that govern seed protein content, and their genomic positions are also documented (http://soybase.org/resources/QTL.php). Our study demonstrates that the higher content of protein in all nine high protein soybean accessions is mainly due to preferential accumulation of 7S and 11S seed storage protein components. We have also demonstrated that the high protein trait can be transferred by crossing PI 407788A with Williams 82, a soybean cultivar with a conventional protein content. Attempts are also being made to identify QTLs associated with high protein and oil utilizing backcross populations derived from mating Williams 82 × PI 407788A.[29] Identification and introduction of high protein QTL alleles from these high protein soybean accessions into high-yielding commercial soybeans will improve both yield and protein content.

AUTHOR INFORMATION

Corresponding Author
*E-mail: Hari.Krishnan@ARS.USDA.GOV.

DISCLOSURE

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.

ACKNOWLEDGMENT

We are grateful to Nathan Oehrle, Sungchan Jang, and Mark Johnston for their excellent technical assistance in data collection and analysis. We acknowledge the Charles W. Gherke Proteomics Center at the University of Missouri for use of their laser scanners for image acquisition and for the use of their image analysis software.

REFERENCES

(1) Fehr, W. R. Soybean. In *Oil Crops of the World*; Robbelen, G., Downey, R. K., Ashri, A., Eds.; McGraw-Hill: New York, 1989.

(2) Koshiyama, I. Storage Proteins of Soybean. In *Seed Proteins Biochemistry, Genetics, Nutritive Value (Advances in Agricultural Biotechnology)*; Gottschalk, W., Muller, H. P., Eds.; Kluwer Academic Publishers: Hingham, MA, 1983.

(3) Mooney, B. P.; Krishnan, H. B.; Thelen, J. J. High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry* **2004**, *65*, 1733–1744.

(4) Natarajan, S. S.; Xu, C.; Bae, H.; Caperna, T. J.; Garrett, W. M. Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J. Agric. Food Chem.* **2006**, *54*, 3114–3120.

(5) Nielsen, N. C. Soybean Seed Composition. In *Soybean: Genetics, Molecular Biology and Biotechnology*; Verma, D. P. S., Shoemaker, R. C., Eds.; CABI:Wallingford, U.K., 1996.

(6) Krishnan, H. B. Biochemistry and molecular biology of soybean seed storage proteins. *J. New Seeds* **2000**, *2*, 1–25.

(7) Thanh, V. H.; Shibasaki, K. Major proteins of soybean seeds. A straight forward fractionation and their characterization. *J. Agric. Food Chem.* **1976**, *24*, 1117–1121.

(8) Hartwig, E. E. Registration of soybean high-protein germplasm line D76-8070. *Crop Sci.* **1990**, *30*, 764–765.

(9) Leffel, R. C. Registration of high-protein soybean germplasm lines BARC-6, BARC-7, BARC-8, and BARC-9. *Crop Sci.* **1992**, *32*, 502.

(10) Yaklich, R. W. β-conglycinin and glycinin in high-protein soybean seeds. *J. Agric. Food Chem.* **2001**, *49*, 729–735.

(11) 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.

(12) Korczak, J. F. Using Geographically Diverse High-Protein Germplasm to Simultaneously Improve Yield and Protein Percentage in Soybeans. Ph.D. Dissertation, University of Illinois, Urbana-Champaign, IL, 1984.

(13) Nelson, R. L.: Amdor, P. J.; Orf, J. H.; Lambert, J. W.; Cavins, J. F.; Kleiman, R.; Laviolette, F. A.; Athow, K. A. Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (PI 273.483 to PI 427.107). *USDA Technical Bulletin No. 1718*, 1987.

(14) Nelson, R. L.; Amdor, P. J.; Orf, J. H.; Cavins, J. F. Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (PI 427.136 to PI 445.845). *USDA Technical Bulletin No. 1726*, 1988.

(15) Bernard, R. L.; Cremeens, C. R.; Cooper, R. L.; Collins, F. I.; Krober, O. A.; Athow, K. L.; Laviolette, F. A.; Coble, C. J.; Nelson, R. L. Evaluation of the USDA Soybean Germplasm Collection: Maturity Groups 000 to IV (FC 01.547 to PI 266.807). *U.S. Department of Agriculture Technical Bulletin No. 1844*, 1998.

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

(17) 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.

(18) Sarma, A. D.; Oehrle, N. W.; Emerich, D. W. Plant protein isolation and stabilization for enhanced resolution of two-dimensional polyacrylamide gel electrophoresis. *Anal. Biochem.* **2008**, *379*, 192–195.

(19) Krishnan, H. B.; Oehrle, N. W.; Natarajan, S. S. A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: A case study using *Glycine max*. *Proteomics* **2009**, *9*, 3174–3188.

(20) Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, B. H.; Scallon, 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.

(21) Beilinson, V.; Chen, Z.; Shoemaker, R.; Fischer, R.; Goldberg, R.; Nielsen, N. Genomic organization of glycinin genes in soybean. *Theor. Appl. Genet.* **2002**, *104*, 1132–1140.

(22) Sebastiani, F. L.; Farrell, L. B.; Schuler, M. A.; Beachy, R. N. Complete sequence of a cDNA of the a subunit of soybean β-conglycinin. *Plant Mol. Biol.* **1990**, *15*, 197–201.

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

(24) 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.

(25) Krishnan, H. B.; Natarajan, S. S.; Mahmoud, A. A.; Bennett, J. O.; Hari Krishnan, A.; Prasad, B. N. 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.

(26) Krishnan, H. B.; Natarajan, S. S.; Mahmoud, A. A.; Nelson, R. L. Identification of glycinin and β-conglycinin subunits that contribute to the increased protein content of high-protein soybean lines. *J. Agric. Food Chem.* **2007**, *55*, 1839–1845.

(27) Brim, C. A.; Burton, J. W. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. *Crop Sci.* **1979**, *19*, 494–498.

(28) Wehrmann, 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.

(29) Schultz, S. J.; Nelson, R. L. Seed protein and oil QTL mapping in three backcross soybean populations. In 2006, Agronomy Abstracts [CD-ROM computer file]; ASA: Madison, WI.