research articles Journal of proteome research

Quantitation of Soybean Allergens Using Tandem Mass Spectrometry

Norma L. Houston,[†] Dong-Gi Lee,[†] Severin E. Stevenson,[†] Gregory S. Ladics,[‡] Gary A. Bannon,[§] Scott McClain,[||] Laura Privalle,[⊥] Nicola Stagg,[#] Corinne Herouet-Guicheney,[▽] Susan C. MacIntosh,[¶] and Jay J. Thelen*,[†]

Department of Biochemistry and Interdisciplinary Plant Group, University of Missouri, Columbia, Missouri 65211, United States, Pioneer Hi-Bred International, Inc., DuPont Agricultural Biotechnology, DuPont Experimental Station, P.O. Box 80353, Wilmington, Delaware 19880, United States, Monsanto, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167, United States, Syngenta Company, 3054 East Cornwallis Drive, Research Triangle Park, North Carolina 27709, United States, BASF Plant Science, LP, 26 Davis Drive, Research Triangle Park, North Carolina 27709, United States, Dow AgroSciences, 9330 Zionsville Road, Indianapolis, Indiana 46077, United States, Bayer SAS, Bayer CropScience, 355 rue Doistoievski, 06903 Sophia, France, and MacIntosh & Associates, Inc., 1203 Hartford Avenue, Saint Paul, Minnesota 55116, United States

Received September 6, 2010

Soybean (*Glycine max*) seed contain some proteins that are allergenic to humans and animals. However, the concentration of these allergens and their expression variability among germplasms is presently unknown. To address this problem, 10 allergens were quantified from 20 nongenetically modified commercial soybean varieties using parallel, label-free mass spectrometry approaches. Relative quantitation was performed by spectral counting and absolute quantitation was performed using multiple reaction monitoring (MRM) with synthetic, isotope-labeled peptides as internal standards. During relative quantitation analysis, 10 target allergens were identified, and five of these allergens showed expression levels higher than technical variation observed for bovine serum albumin (BSA) internal standard (~11%), suggesting expression differences among the varieties. To confirm this observation, absolute quantitation of these allergens from each variety was performed using MRM. Eight of the 10 allergens were quantified for their concentration in seed and ranged from approximately 0.5 to 5.7 μg/mg of soy protein. MRM analysis reduced technical variance of BSA internal standards to approximately 7%, and confirmed differential expression for four allergens across the 20 varieties. This is the first quantitative assessment of all major soybean allergens. The results show the total quantity of allergens measured among the 20 soy varieties was mostly similar.

Keywords: allergen • soybean • quantitative proteomics • multiple reaction monitoring • spectral counting • mass spectrometry • soya • nongenetically modified • endogenous allergens

Introduction

Soybean (*Glycine max*) is a major oilseed crop rich in protein and oil that is often incorporated into products such as processed foods and pharmaceuticals.[1,2] However, soybean is listed as one of the "big eight" allergenic foods in the United States of America, which together are responsible for 90% of all food allergies.[3−5] Soybean is also listed as one of the "big twelve" allergenic foods in Europe.[6] For allergic consumers, avoidance of soy is the only means to prevent an allergic reaction. Soy allergies are a response to at least one of 37 proteins.[7] All these sequences are grouped into 5 known allergens: the seed storage proteins Gly m 6 (glycinin G1, G2, G3, G4, and G5) and Gly m 5 (beta-conglycinin), Gly m TI (Kunitz typsin inhibitor), Gly m Bd 30K (or P34) and Gly m Bd 28K.[7−9] At present, the natural variation of allergen expression in soy varieties is unknown because the technical ability to quantify more than the most prominent seed storage proteins has not been available. Understanding this range of variation would be useful for supporting eventual determinations of exposure levels and an understanding of "tolerance thresholds" for one or more of the major allergens.

Historically, immunoassays, such as IgE-immunoblotting and enzyme-linked immunosorbent assays (ELISA), have been the standard approach for identification and quantification of soy allergens.[10,11] However, proteomic methods have been used

* Corresponding author. Tel., +1 573 884 1325; fax, +1 573 884 9676; e-mail, thelenj@missouri.edu.
[†] University of Missouri.
[‡] Pioneer Hi-Bred International, Inc.
[§] Monsanto.
[||] Syngenta Company.
[⊥] BASF Plant Science, LP.
[#] Dow AgroSciences.
[▽] Bayer SAS.
[¶] MacIntosh & Associates, Inc.

10.1021/pr100913w

Published on Web 11/12/2010

as alternatives to immunoassays, because immunoassays have sera limitations, high variability, and a limited range of sensitivity and specificity, especially among allergen protein families.[11] Traditional gel-based proteomics have been used to identify and profile multiple soy allergens.[12−16] Two-dimensional gel electrophoresis (2-DGE) followed by matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry (MS) was used to identify, but not quantify, soybean Gly m 5 subunits, Gly m 6 subunits, and Kunitz trypsin inihibitors.[13−15,17,18] To investigate an allergen's natural variation in expression, Xu et al. performed 2-DGE on 16 soybean varieties and showed relative variation in allergens, such as Gly m 5 and Gly m Bd 30K, between wild and cultivated soybean varieties; however, it was unclear how much of this variation was due to experimental (technical precision) versus biological (natural) as the use of an internal control was not employed.[16]

As a result of recent advances in proteomics, gel- and label-free approaches for determining relative and absolute protein levels in a complex biological matrix are available.[19,20] These techniques can compare the expression of multiple proteins in biological samples in a more rapid manner than a 2-D gel approach and with better reproducibility.[20] One strategy, spectral counting, determines the relative quantity of a protein by counting the number of MS/MS (MS^2) spectra from a liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis that match to peptides from an in silico-digested database of proteins.[21] Matched MS^2 spectra can be considered surrogates for the intact protein. In 2009, Stevenson et al. analyzed nontransgenic and transgenic peanut RNAi Ara h 2 knock-down lines using spectral counting and showed relative differences in the quantities of eight peanut allergens.[19] In addition to down-regulation of Ara h 2, other allergens levels changed, which was confirmed by quantitative Western blotting. This investigation was the first to demonstrate the utility of spectral counting for quantifying seed allergens.

Other quantitative proteomic techniques, such as those employing synthetic, stable isotope-labeled peptides, also referred to as absolute quantitation (AQUA) peptides, can determine the per mass unit amount of one or more proteins from a biological sample within a complex matrix.[22,23] With the AQUA strategy, samples are spiked with AQUA peptide standards followed by selected reaction monitoring (SRM) for a single peptide and multiple reaction monitoring (MRM) for multiple peptides.[23] During MRM analysis of a complex mixture, the mass spectrometer is tuned to scan for ideal transition ions (precursor-to-product) of target peptides, both labeled and unlabeled in the case of AQUA analysis. Ion current signals are plotted over time and integrated to calculate peak areas. Peak areas for AQUA and native peptides are compared to calculate absolute quantities of native peptide in the sample. Shefcheck et al. used a similar technique to identify 2 ppm of peanut Ara h 1 in a sample of dark chocolate candy.[24] Both relative and absolute methods have been validated and used to quantify a variety of proteins in numerous plant species.[19,24−29] However, both methods have not been used in parallel for a single study, and neither has been employed to determine natural variation in soy allergen expression across multiple germplasms within a single species.

To characterize natural variation of allergens in a major crop, we report a quantitative proteomic analysis of allergens in 20 commercial soy varieties. Total protein extracted from each seed variety was isolated, trypsin digested, and analyzed by LC−MS/MS. In this analysis, we identified over 100 proteins, including 10 soy allergens. We examined the expression profiles of the 10 allergens using spectral counting. The LC−MS/MS data were also used to design 15 AQUA peptides for various soy allergens, of which 13 were used to calculate absolute quantities for eight allergens.

Experimental Procedures

Plant Material. Four companies provided a total of 20 nongenetically modified commercial soybean varieties as listed here: *90A01*, *91M10*, *92M10*, *92M72*, *93M14*, *93B82*, *S-100*, *Mandarin*, *AK*, *CNS*, *Beck*, *Dwight*, *Hutcheson*, *Opal*, *AJB2501KOC*, *M-SOY 8411*, *Stewart 3454*, *EXP125*, *A2442*, and *Maverick*. Soybean varieties were selected to ensure representation of the germplasm currently grown by U.S. farmers.

Protein and Proteolytic Peptide Preparations. Total proteins were extracted from three biological replicates (4 seeds) of the soybean varieties using a phenol extraction method.[18,30] Mature soybean seed from each variety were pulverized into a fine powder using a standard blender. Protein pellets were resuspended in 300 μL of resuspension buffer (50 mM Tris-HCl, pH 8.0, 5 M urea) by pipetting and vortexing, and finally centrifuged at 13 000*g* for 10 min at 4 °C. Supernatants were collected and stored at −80 °C. Protein was quantified using the Coomassie dye binding assay (Bio-Rad, Hercules, CA) employing bovine serum albumin (BSA; Fisher Scientific, Houston, TX) as a standard[31] and are the average of six replicate readings.

As an internal standard, BSA was added to give a protein ratio of 1% (w/w) BSA/soybean protein. Protein digestion with trypsin was carried out after reducing and alkylating Cys residues, according to Lee et al.[30]

Soy Food Allergy Sequences. The 10 soy allergens quantified included glycinins G1, G2, G3, and G4 (Genbank accession numbers 121276, 121277, 121278, and 121279), glycinin precursor (75221455), Gly m 5 beta-conglycinin alpha subunit (121281), Kunitz trypsin inhibitor 1 (125722), Kunitz trypsin inhibitor 3 (125020), Gly m Bd 28K (12697782), and Gly m Bd 30K (or P34) (84371705).

Relative Quantitation Using Mass Spectrometry. Relative quantitation of trypsin-digested protein samples was carried out on a linear ion trap tandem mass spectrometer (ProteomeX LTQ-ETD, Thermo-Fisher, San Jose, CA) using liquid chromatography and nanospray ionization. The relative quantitation was performed according to previously described methods.[19] Briefly, LC separation was performed using fused silica nanospray needles, 10 cm length (360 μm outer diameter, 150 μm inner diameter; Polymicro Technologies, Phenix, AZ), that were packed with "magic C18" (100 Å, 5 μm particle; Michrom Bioresources, Auburn, CA) in 100% methanol. Dried peptides were reconstituted in 0.1% (v/v) formic acid to a final concentration of 100 ng/μL. Ten-microliter injections were analyzed on a mass spectrometer. Peptides were separated on the C18 column using a 60 min gradient, 0−90% solution 1 (100% acetonitrile with 0.1% (v/v) formic acid) and 100−10% solution 2 (0.1% formic acid). Samples were analyzed in the data-dependent positive acquisition mode using the normal scan rate for precursor ion analysis, with dynamic exclusion enabled. Following each full scan (400−2000 m/z), data-dependent trigged MS/MS scans were collected for the 10 most intense parent ions.

Uninterpreted LC−MS/MS files were searched against the translated soybean genome (www.phytozome.net) using

SEQUEST (ThermoFinnigan, San Jose, CA; version 2.7) batch search in Bioworks version 3.3.1. Search parameters were as follows: fragment ion mass tolerance, 1.0 amu; peptide tolerance, 1000 ppm; static modification, carboxyamidomethylation (C); variable modification, oxidation (M). Matching peptide were filtered according to cross-correlation score (XCorr at least 1.5, 2.0, and 2.5 for +1, +2 and +3 charged peptides, respectively). Search result files were imported into Scaffold version 2.2.1 (Proteome Software, Portland, OR). Scaffold was used for spectral counting and to validate MS/MS based peptide and protein identification. Peptide and protein identifications were accepted if they could be established at greater than 95% and 99% probability, respectively. For all protein assignments, a minimum of two unique peptides was required. Spectral counts for allergens were normalized by calculating the average percent of total spectral counts. The average percent of total spectal counts was determined by dividing the allergen spectral count by the total spectral counts per sample; the ratio was multiplied by 100 and averaged for the three biological replicates.

Absolute Quantitation Using Mass Spectrometry. Isotope-labeled, synthetic peptide standards (certified ≥95% purity) were purchased from Sigma-Aldrich (St. Louis, MO). Peptides were synthesized with a single C^{13}- and N^{15}-labeled arginine or lysine, depending upon the peptide sequence. Peptide stock solutions were prepared at 250 fmol/μL in a mixture of 0.1% formic acid/acetonitrile (97:3, v/v) and stored at −80 °C.

Absolute quantitation was carried out using an Agilent 6410 triple quadrupole MS system coupled with an 1100 series LC system and HPLC-Chip cube (Agilent Corporation, Santa Clara, CA). LC separation was performed using an Agilent large capacity chip (II) with a 150 mm separation column and a 160-nL enrichment column packed with C18 (300 Å, 5 μm particle). The system was controlled by MassHunter software v. B.02.01 (Agilent Corporation). Mass spectrometry was performed in positive ionization mode with a capillary voltage set at 1900 V and electron multiplier voltage of 400 V. Nitrogen drying gas flow was 4 L/min and gas temperature was 300 °C. For collision-induced dissociation (CID), ultra high-purity nitrogen was used as collision gas. The position of the nano-Chip tip was optimized for low spray flow rate to produce a direct ESI spray under the correct voltage to achieve high MS sensitivity and a stable spray.

Dynamic MRM parameters for each AQUA peptide were optimized with the MassHunter Optimizer Software using default parameters (Agilent Corporation). To develop the LC−MRM assay, it was necessary to qualify the data collection process by "tuning" on each peptide standard and identifying optimal product ions for quantitation. The "tuning" process first required analysis of the AQUA peptides by LC−MS/MS to identify the most abundant charge state. For internal standard optimization, 2 μL of each AQUA peptide stock was injected into the mass spectrometer and analyzed with a gradient of 0−75% solution B (0.1% formic acid and 10% water in acetonitrile; 300 nL/min) and 100−25% solution A (0.1% formic acid and 3% acetonitrile; 300 nL/min) in 9 min. During iterative runs, the optimizer software identified the most abundant precursor and fragment voltage for each precursor ion. It also selected the most abundant product ion and optimized the collision energy (CE) for each product ion (Supporting Information Table S1). Following analyses, four product ions were chosen for acquisition and quantitation during dynamic MRM mode (Table 1). Suitable product ions for GlycininPre-1 were not detected despite repeated efforts to fragment this peptide. Table S1 shows the optimized dynamic MRM parameters for AQUA and native peptides.

After identifying LC−MRM parameters, the linear quantitative range for native and AQUA peptides was determined. Serial dilutions (ranging from 2 to 4000 ng) of soy tryptic peptides were analyzed by LC−MRM. In a separate analysis, serial dilutions (0.01−16 fmol/ng of soy matrix) of AQUA peptides spiked into 63 ng of soy matrix were analyzed by LC−MRM. Peak integration was performed on MRM chromatograms using MassHunter Qualitative Software v B.02.00 (default parameters) and yielded peak areas for each product ion (Agilent Corporation). Data were manually inspected to ensure correct peak detection and accurate integration. Peak areas for biological replicates were averaged and graphed to visually identify linear quantitative ranges for native and AQUA peptides (Figure 3 and Figures S1−S2).

For absolute quantitation of allergens, each soy sample was spiked with 13 AQUA peptides (0.25−4 fmol/ng of soy matrix) in a mixture of 0.1% formic acid/acetonitrile (97:3, v/v), and 4 μL of this sample mix was analyzed (Table S1). Each sample mix was separated with a 20 min gradient of 0−75% of solution B (0.1% formic acid and 10% water in acetonitrile; 300 nL/min) and 100−25% solution A (0.1% formic acid and 3% acetonitrile; 300 nL/min). After the separation phase of the gradient, residual pepetides were washed from the column with 100% solution B for 30 s. The column was rinsed using a 4 min "sawtooth" gradient with 100% solution B. The column was re-equilibrated for 6 min with 100% of solution A before the next sample was loaded. Quantification was performed in dynamic MRM mode. The peak widths of precursor and product ions were maintained at 0.7 amu at half-height in the dynamic MRM mode. All MRM data were analyzed with MassHunter Qualitative Software v. B.02.00 using default parameters for peak integration. All data were manually inspected to ensure correct peak detection and accurate integration. Native peptide peaks were identified based on retention times and fragment ion distribution and intensity closely matching their respective AQUA peptide standard. Product ion ratios were calculated for AQUA and native peptides to verify identical precursors. Product ions with the smallest peak area were designated as "qualifier" ions. The remaining three ions were classified as quantifier ions, and their peak areas were summed (Table 1). The ratio of the summed quantifier ions to the qualifier ions was calculated, averaged, and graphed. Absolute quantities were calculated using the ratio of native to AQUA peak area multiplied by the concentration of the AQUA internal standard.

Results

Protein Yield of the 20 Soy Varieties. Twenty varieties of mature soybean were provided by four seed companies, Pioneer Hi-Bred International, Inc. (Johnston, IA), BASF (Research Triangle, NC), Dow Agrosciences (Indianapolis, IN), and Monsanto (St. Louis, MO). Total seed proteins were extracted from three biological replicates of each variety using the phenol extraction method.[18,30] Figure 1 shows the average concentration of extracted protein from each soy variety. The soy protein concentrations among the 20 varieties showed variations in concentration ranging from approximately 5 to 8 μg/μL.

Relative Comparison of Allergens in Soybean Varieties Using Spectral Counting. Tryptic peptides from soy protein extracts were analyzed by LC−MS/MS. Mining these data

Table 1. Precursor and Product Ions Analyzed in Absolute Quantitation of Soy Allergens[a]

allergenic proteins	peptide name	peptide sequence[b]	Z^c	native peptide[d] precursor ion (m/z)	native peptide[d] product ion (m/z)[f]		AQUA peptide[e] precursor ion (m/z)	AQUA peptide[e] product ion (m/z)	
Glycinin G1	GlyG1-1	LSAEFGSLR	2	490.3	579.3	L	495.3	876.5	L
					779.4	T		789.4	T
					866.5	T		589.3	T
					432.3	T		442.3	T
	GlyG1-2	VLIVPQNFVVAAR	2	713.4	425.3	L	718.4	425.3	L
					501.3	T		506.3	T
					1001.6	T		1011.6	T
					326.2	T		326.2	T
Glycinin G2	GlyG2-1	LSAQYGSLR	2	497.8	881.5	L	502.8	891.5	L
					794.4	T		804.4	T
					595.3	T		605.3	T
					432.3	T		442.3	T
	GlyG2-2	NLQGENEEEDSGAIVTVK	3	644.6	899.5	L	647.3	897.5	L
					687.5	T		695.5	T
					446.3	T		454.3	T
					347.2	T		355.2	T
Glycinin G3	GlyG3-1	LSAQFGSLR	2	489.8	547.3	L	494.8	547.3	L
					778.4	T		788.4	T
					707.4	T		717.4	T
					865.5	T		875.5	T
	GlyG3-2	FYLAGNQEQEFLQYQPQK	3	744.3	791.4	L	747.0	799.4	L
					663.4	T		671.4	T
					372.2	T		380.2	T
					311.1	T		311.1	T
Glycinin G4	GlyG4	VESEGGLIQTWNSQHPELK	3	718.0	854.5	L	720.7	858.5	L
					962.5	T		966.5	T
					1027.0	T		1031.0	T
					486.3	T		494.3	T
Glycinin precursor	GlyPre-2	NGLHLPSYSPYPR	3	501.0	535.3	L	504.3	535.3	L
					619.3	T		542.3	T
					532.3	T		629.3	T
					485.2	T		485.2	T
Beta-conglycinin, alpha subunit	Bcon	LITLAIPVNKPGR	3	464.7	532.8	L	468.0	537.8	L
					583.4	T		588.4	T
					767.5	T		777.5	T
					476.3	T		481.3	T
Kunitz trypsin inhibitor 3	KTI3-1	FIAEGHPLSLK	3	404.6	751.5	L	407.3	759.5	L
					557.4	T		565.4	T
					532.8	T		536.8	T
					476.3	T		480.3	T
	KTI3-2	VSDDEFNNYK	2	615.8	538.3	L	619.8	546.3	L
					1044.4	T		1052.4	T
					685.3	T		693.3	T
					1131.5	T		1139.5	T
Kunitz trypsin inhibitor 1	KTI1	DTVDGWFNIER	2	676.3	531.3	L	681.3	541.3	L
					921.5	T		931.5	T
					678.4	T		688.4	T
					1036.5	T		1046.5	T
Gly m Bd 28K	AllGly28	DGPLEFFGFSTSAR	2	765.9	1148.6	L	770.9	1158.6	L
					1019.5	T		1029.5	T
					872.4	T		882.4	T
					725.4	T		735.4	T

[a] Fifteen AQUA peptides were synthesized from 10 soy allergens. LC−MS/MS was used to determine the most abundant precursor and product ions. Four product ions were chosen to verify and quantify each precursor ion. [b] AQUA peptides synthesized for Gly m Bd 30K and Glycinin Precursor (GlycininPre-1) were excluded from the analysis due to low abundance and poor ionization, respectively. [c] Most abundant charge state determined by LC−MS/MS. [d] Soy matrix peptide. [e] Isotope-labeled internal standard. [f] L-qualifier, T-quantifier.

against the translated soybean genome resulted in the assignment of approximately 110 different proteins, including 10 known allergenic proteins: Glycinin precursor, glycinin G1, glycinin G2, glycinin G3, glycinin G4, beta-conglycinin alpha-subunit, Kunitz trypsin inhibitor 1 (KTI1), Kunitz trypsin inhibitor 3 (KTI3), Gly m Bd 28K, and Gly m Bd 30K. Spectral counts for these 10 allergens were averaged from biological replicates and normalized with total spectral counts (assigned spectra). Figure 2 shows percent spectral counts for each allergen. The BSA internal standard revealed approximately 11% coefficient of variance (CV) or technical variation with the spectral counting strategy. Several allergen profiles showed expression variation greater than 11% CV, suggestive of biological variation. On the basis of percent spectral counts and variation among the soy samples, the 10 allergens were grouped into three classes. Class 1 proteins (glycinin G1 and G2) showed

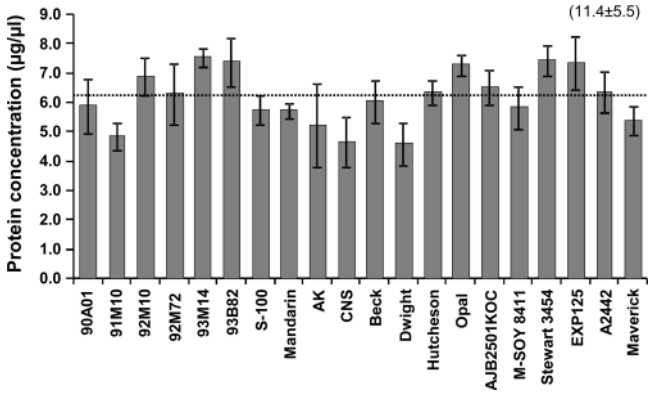

Figure 1. Extracted protein concentrations from 20 soybean varieties. Total protein was isolated from three biological replicates of each 20 commercial soy varieties. Isolated proteins were quantified using the Coomassie dye binding assay employing BSA as standard[31] and are the average of six replicate readings. Protein concentrations (μg/μL) for biological replicates were averaged and graphed. The median of the protein concentration is shown by a dotted line. The overall percent coefficient of variance (CV) with standard deviation is shown in parentheses.

lower or similar variation (8−11% CV) to the BSA internal standard as well as the highest aggregate spectral counts, approximately 20% each, reflecting the high abundance of these two proteins in mature soybean seed.[15] Class 2 proteins (glycinin G4, Gly m 6 precursor, and Gly m 5) showed slightly higher variation than BSA and 5−12% of the overall spectral counts. Class 3 proteins (Gly m Bd 30K, Gly m Bd 28K, KTI1, KTI3, glycinin G3) exhibited high variation (18−38% CV) and low overall percentages of spectral counts (<2% each). Protein classes 2 and 3 are potentially examples of either biological variation (seed-to-seed variability) or allergen expression differences.

Identification of Synthetic Isotope-Labeled Peptides and Development of MRM Mass Spectrometry Assay. Absolute quantitation of soy allergens requires isotope-labeled peptide synthesis followed by MRM assay development. To produce internal standards for the 10 allergens, tryptic peptides representing those allergens identified in the relative quantitation analysis were further analyzed as candidates for synthesis. Tryptic peptides for the 10 allergens were analyzed based upon trypsin cleavage rules using *in silico* predictions: amino acid length, amino acid composition, and sequence redundancy. Peptides with sequences less than 8 or greater than 19 amino acids were excluded. Also, peptides that contained modified amino acids, such as cysteine and methionine, were excluded because of the potential for post-translational modification. Peptides that contained internal trypsin cleavage sites, lysine and arginine, were also excluded. Finally, peptides were searched against the translated soybean genome. If peptides matched more than one sequence, they were considered redundant and were removed from the candidate list. Synthetic peptides were required to be unique to ensure protein-specific quantitation. The final list of tryptic peptides adhered to all of these criteria. Although we initially sought to produce two different peptides for each protein (i.e., 20 peptides), the stringent criteria resulted in 15 ideal peptides for synthesis (Table 1). Two AQUA peptides were synthesized for five allergens while one peptide was synthesized for the remaining five allergens. AQUA peptides were named for each protein, and allergens with two AQUA peptides were sequentially numbered (e.g., KTI3-1 and KTI3-2). Stable isotope-labeled lysine or arginine residues were incorporated into each AQUA peptide. The labeled peptides produce a mass shift in the MS spectrum of +8 or +10 Da when compared to its corresponding soy peptide (i.e., native peptides).

Absolute Quantitation of Allergens in Soybean Varieties Using Synthetic Isotope-Labeled Peptides. In preliminary analyses, the linear quantitative range of native and AQUA peptides was measured for accurate and reproducible quantitation. Serial dilutions (2−4000 ng) of tryptic peptides (v. *Maverick*) were analyzed by LC−MRM analysis. Peak area for biological replicates were averaged and graphed to identify the linear quantitative range for each soy product ion (Figure S1). A linear range was determined for each allergen, except for the Gly m Bd 30K, possibly due to low abundance. Although all four transition ions for the Gly m Bd 28K peptide showed a linear relationship for the dilution series, the y-intercept suggests the absolute amount of this peptide could be underestimated. The linear quantitative range for 13 AQUA peptides was individually narrow and collectively broad, from 2 to 500 ng. The only dilution of soy matrix that overlapped with the linear range of all 13 peptides was 63 ng (Figure 3A). As a result, serial dilutions of each AQUA peptide were spiked into 63 ng of soy matrix and analyzed by LC−MRM. Peak areas for AQUA peptides were average and graphed (Figure S2). Quantitative regions spanned from 0.01 to 16 fmol AQUA peptide/ng of soy matrix (Figure 3B).

Using this collective information, a single multiplexed assay was performed to quantify protein allergens from soybean seed. Tryptic peptides prepared from the 20 soybean samples (with a BSA internal standard) were spiked with 13 AQUA peptides and analyzed by LC−MRM. Product ions with the smallest peak area served as the qualifier ion while the remaining three ions served as quantifier ions (Table 1). The qualifier ion was chosen to validate or "qualify" the peptide identified. The ratio of the product ion levels is reproducible for each precursor ion (peptide) fragmented under similar conditions; therefore, the ratio can be used to validate peptide identity. To verify that AQUA and native precursors were identical, product ion ratios (quantifier/qualifier ratios) were calculated, averaged, and graphed. Product ion ratios were maintained for both the AQUA and native peptides which suggests the product ions originated from identical precursors (Figure 4). Absolute quantities were calculated using the ratio of native to AQUA quantifiers for the 20 soybean varieties. Absolute quantities (in μg of peptide/mg of protein) were averaged from three biological replicates and graphed for each soybean variety (Figure 5). The BSA internal standard showed approximately 7% technical variation which was 36% lower than the technical variation observed for relative quantitative analysis (spectral counting). Similarly, allergen concentrations ranged from 5% to 12% technical variation among biological replicates, considerably lower than the variation observed for relative quantitative analysis.

Allergens were grouped into classes based upon abundance. As observed with spectral counting, glycinin G1 and G2 were by far the two most abundant allergens with concentrations of approximately 4.5 and 5.7 μg of peptide/mg of protein, respectively. The moderately abundant class included KTI1, KTI3, and glycinin G4 at approximately 1.6 μg of peptide/mg of protein. KTI1 showed the highest biological variation (11% CV) within the moderately abundant allergenic proteins. The low abundant class included four allergens (glycinin G3, Gly

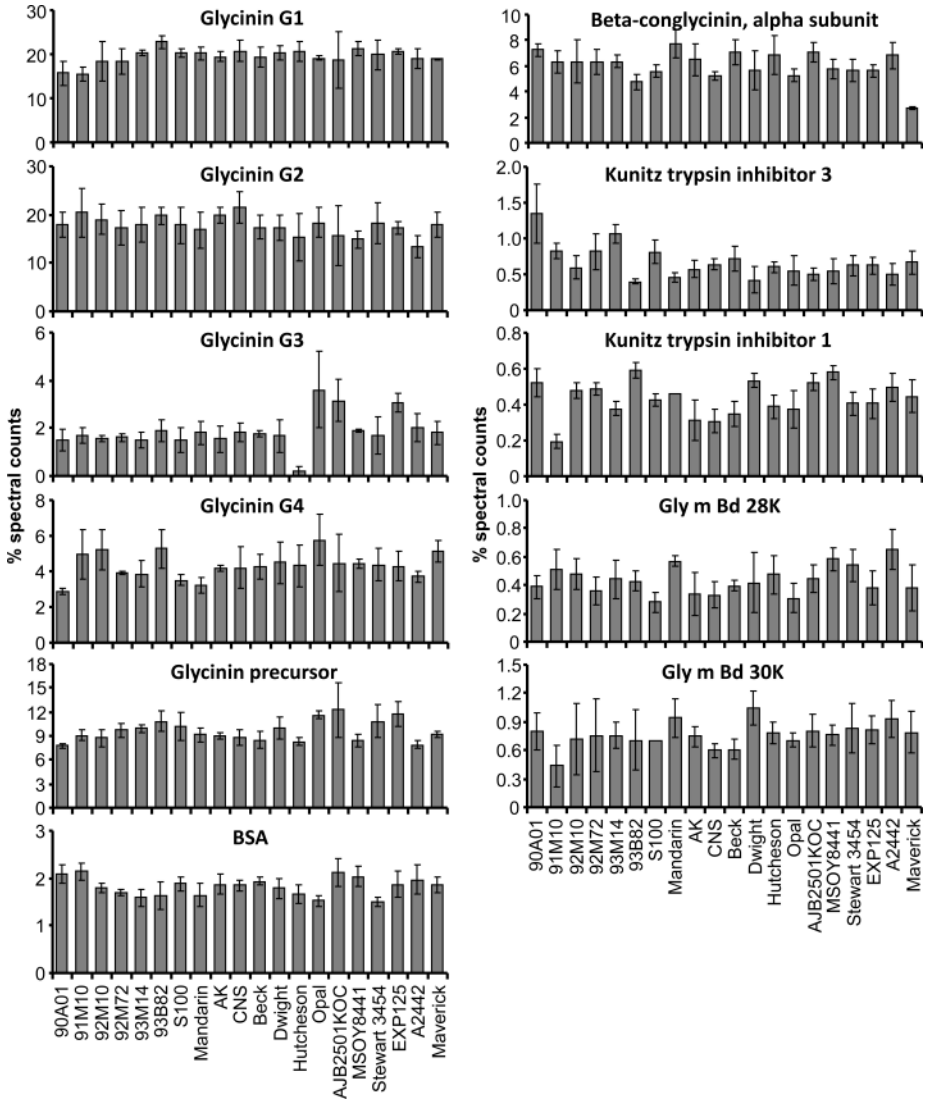

Figure 2. Quantitation of 10 proteins and BSA internal standard by spectral counting in 20 soybean varieties. Trypsin-digested protein (1 μg) was analyzed by LC−MS/MS on a linear ion trap mass spectrometer using a Top 10 (1 MS + 9 MS2 scans) scanning approach to monitor most abundant ions. Over 100 proteins were identified including the 10 allergens shown here. Histogram plots show average percent of total spectral counts and standard deviation from three biological replicates for each soybean variety.

m Bd 28k, Gly m 5, and Glycinin precursor) expressed at approximately 140−700 ng of peptide/mg of protein each. Glycinin G3 showed the highest biological variation (12% CV) of all allergens.

Figure 5 also shows variation in concentration among peptides from the same protein sequence. For example, peptides glycinin G1-1 and G1-2 showed an average concentration of 0.5 and 4.5 μg of peptide/mg of protein, respectively. Variation among peptides from the same allergens was most likely due to peripheral alternative trypsin cleavage sites or "ragged ends". Ragged ends are tandem cleavage sites found at the end of a peptide (LSAEFGSLR.K). Because cleavage can occur at either site, monitoring one form of the peptide underrepresents protein abundance. Unfortunately, five of the AQUA peptides used in this analysis have ragged ends (Table 2) and were therefore not used to determine the absolute quantification of allergens. As a result, only eight proteins were quantified.

Of those quantified, four allergens (glycinin G3, glycinin G4, Gly m 5, and KTI1) showed at least a 2-fold change in abundance among the soy varieties (Figure 5). For example, *92M10* showed an approximately 5-fold increase in KTI1 when compared to *A2442*. Also, glycinin G3 levels in *Opal* showed a 7-fold increase over *Dwight* and *AK* varieties. Glycinin G3 quantities were the lowest in the *Hutcheson* variety where the quantity was not within the range of detection. However, *Hutcheson*, along with *S100* showed the highest levels of glycinin G4. *Hutcheson* showed an approximately 3-fold increase in glycinin G4 compared to *Beck*. Similarly the *Mandarin* variety, showed a 3-fold increase in Gly m 5 over *Dwight*, the variety with the lowest expression. When compared to *Dwight*, *Opal*, *CNS*, *Beck*, and *Maverick* varieties, Gly m 5 was nearly tripled in *Mandarin*.

To evaluate all allergens in each soy variety, absolute quantities of each allergen were summed and graphed (Figure 6). Peptides with ragged ends were not used to calculate the summed values. Despite varied quantities of each allergen, summed values show low variation (9% CV) among the 20 soy varieties. The average total sum of allergens from soy varieties was approximately 16 μg of peptide/mg of protein (approximately 487 μg of full-length allergen/mg of protein).

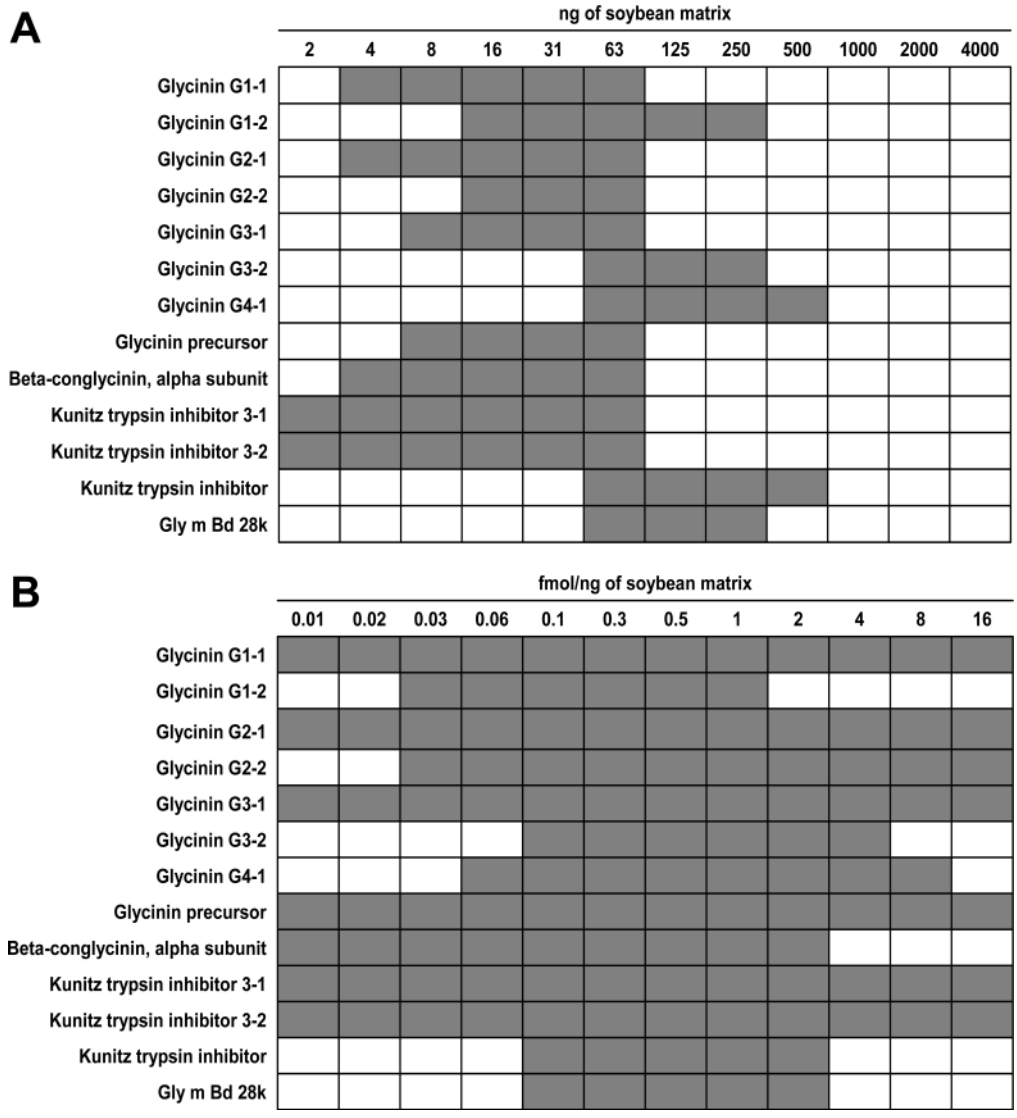

Figure 3. Linear range of the 13 native peptides in soy variety *Maverick*. (A) Trypsin-digested protein (2−4000 ng) was analyzed by LC−MS/MS on a triple quadruple mass spectrometer using multiple reaction monitoring (MRM) scanning to monitor product ions. Gray regions represent the linear range of native product ions. (B) Linear range of the 13 AQUA peptides in *Maverick* soy variety. AQUA peptides (0.01−16 fmol/ng) were analyzed by LC−MS/MS on a triple quadruple mass spectrometer using MRM scanning approach to monitor product ions. Gray regions represent the linear range of AQUA product ions.

Discussion

Comparison of Relative and Absolute Quantitative Analyses. This parallel study of spectral counting and a MRM strategy using AQUA peptides for absolute quantitation reveals both methods are suitable to detect and compare soybean allergen expression levels. The results from this study provide insight into advantages and disadvantages of relative and absolute quantitative proteomic methods. Both proteomic methods showed similar expression profiles for most allergens among the 20 soy varieties (Figures 2 and 5). Also, both proteomic methods produced technical variation less than 12%; however, the technical variation with MRM quantitation (7% CV) was lower than that of the spectral counting analysis (11% CV). Increasing the number of replicates may reduce the CV value associated with the BSA internal standard, which would serve to further refine biological variability of individual allergens. Further analysis of the MRM technical variation determined that manual "pipetting" was likely a contributing source of experimental error (data not shown). Other sources of variation, such as less than complete tryptic digestion cannot be fully discounted, although we have verified by SDS-PAGE the absence of proteins after digestion (data not shown). Despite having higher technical variation, the spectral counting approach successfully monitored relative expression for 10 proteins while the MRM technique successfully quantified only eight proteins. The remaining two allergens were not quantified due to low abundance and lack of adequate AQUA peptides. Absolute quantitation by the MRM-AQUA strategy is dependent upon the identification and production of an ionizable, unmodified, and unique synthetic peptide. Each AQUA peptide used in this study was screened manually because there are few software programs available. The development of inexpensive but effective software to identify AQUA peptides would decrease the time needed to perform this process. In this study, we began with a limited list of empirically determined tryptic peptides instead of tryptic peptides determined by *in silico*

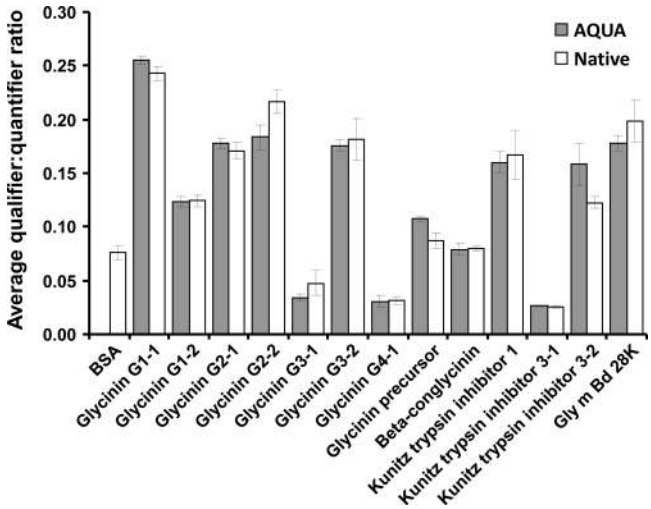

Figure 4. Product ion distribution between AQUA and native peptides. Product ions with the smallest peak area served as qualifier ions, and the remaining product ions (quantifier ions) were summed. The ratio of the qualifier peak area and summed quantifier peak areas were averaged and graphed. Histogram plots show average ratio and standard deviation from three biological replicates for 20 soybean varieties.

analysis, to ensure that each was ionizable. As a result of the limited list, five test proteins had only one AQUA peptide. The screening also excluded AQUA peptide candidates with internal tryptic cleavage sites, but neglected peripheral alternative cleavages. The inclusion of AQUA peptides with peripheral alternative cleavages caused differences in final absolute quantiation and inaccurate quantification of Gly m Bd 28K. In the future, peptides with ragged ends within three amino acids of the tryptic peptide will be excluded. AQUA peptide candidates were also screened for redundancy in the soy translated genome. A unique peptide is necessary to ensure that a single allergen is being measured; however, determining if a peptide is unique is dependent upon the quality of the database being searched.

Differentially Expressed Allergens in Commercial Soy Varieties. Expression levels of some soy allergens have been estimated based upon partial purification from mature seed. For example, depending on the soy variety, Gly m 6 and Gly m 5, KTI, and Gly m Bd 30K proteins represent as much as 70, 15, and 2−3% of total seed protein.[32−35] However, these percentages are only estimates because in many cases the isolated protein fractions were impure, containing several proteins and/or protein isoforms (e.g., Gly m 6 and Gly m 5). More recently, the use of high-resolution two-dimensional gel electrophoresis and peptide mass fingerprinting resulted in more accurate and specific estimations of many soybean allergens,[17] although accurate determination of gene isoforms and absolute expression levels still remained elusive. Using parallel mass spectrometry approaches, we have determined both the relative and absolute quantities of 10 and eight allergens, respectively, including individual members that compose Gly m 6 allergen.

Gly m 6 (Glycinin) is a hexameric protein that is assembled by five subunits G1, G2, G3, G4, and G5 where each subunit is composed of an acidic and a basic chain.[36] Two other glycinin genes (G6 and G7) have been identified in the soybean genome; however, G6 is a pseudogene and G7 protein has not been detected, most likely due to low expression.[37] Glycinin subunits have diverged into groups. Group-I includes G1, G2, G3 while group-II is composed of G4 and G5. In this analysis, both proteomic methods identified members of group-1, G1 and G2, as the most abundant allergens in soybean seeds which corroborate results from previous studies.[38,39] By using subunit-specific peptides, we were able to specifically and accurately quantify four of the five glycinin subunits. The results show not only different overall expression levels, but also that some glycinin subunits are more variable in expression than others. For example, glycinins G1 and G2 have similar expression level among the 20 soy varieties while glycinin G3 was approximately 10-fold less abundant than G1 and G2 and was much more variable in expression among the 20 soybean varieties (Figure 5). One variety, *Hutcheson*, expressed little to no G3 protein while the *Opal* variety had the highest protein abundance (1.3 μg of peptide/mg of protein). These results were corroborated by spectral counting and suggest the possibility of a gene lesion for this specific isoform, which is not unprecedented. Cho et al. identified a soybean cultivar, *Forrest*, that showed reduced G3 mRNA levels due to a possible sequence inversion.[40] The G4 glycinin was the most abundant subunit, but like G3 was variable in expression; however, that expression trend was different than G3. Overall, these results suggest a lack of coordinated control for the glycinin gene family. Upon the basis of this observation, future discussions of glycinin variability must account for individual subunits, and employ appropriate methods to resolve and quantify them.

The other major seed storage protein in soybean seed is Gly m 5 (beta-conglycinin).[41] Like Gly m 6, Gly m 5 is a predominant storage protein composed of subunits that are allergenic.[36,41] In this study, the average total sum of the glycinin-related subunits and Gly m 5 was 13 μg of peptide/mg of protein (approximately 449 μg of full-length allergen/mg of protein). Gly m 5 alpha subunit was shown to have low abundance (avg. 0.5 μg of peptide/mg of protein) and high variation among the soy varieties analyzed. Further data is needed to determine if variation in Gly m 5 subunit expression affects the production of the complex and allergenicity. The Kunitz soybean trypsin inhibitor is known to be composed of three proteins (KTI1, KTI2 KTI3).[42] Natarajan et al. identified KTI1 and KTI3 among 16 different soybean genotypes by using 2-DGE followed by MALDI-TOF; of the KTIs identified, KTI3 spots showed the highest intensity.[13] Our data showed that KTI1 and KTI3 have a similar average absolute quantitation level of 1.2 and 1.0 μg of peptide/mg of protein, respectively. However, spectral counting and MRM analysis differed, with spectral count analysis showing high variation for the KTI1 and KTI3 (23% and 34%). These differences in expression may be the result of unavailable proteolytic cleavage sites due to post-translation modification, as previously shown in chickpea KTIs.[43] Depending on the soy variety, unavailable proteolytic cleavage sites could modify the number of unique peptides quantified by LC−MS/MS methods.

Gly m Bd 30K and Gly m Bd 28K are allergenic glycoproteins with similar sugar compositions.[44,45] In this study, both proteins were successfully quantified only by spectral counting due to low abundance or an AQUA peptide with ragged ends.[46−49] In the soy varieties analyzed, spectral counting data for Gly m Bd 30K and Gly m Bd 28K suggested both proteins have a low abundance (less than 2% spectral counts) and high variation (17% and 22% CV). Similarly, Tsuji et al. and Natarajan et al. used 2-DGE techniques to find very low concentrations of Gly m Bd 28K in soy flakes and soybean seeds, respectively.[45,50]

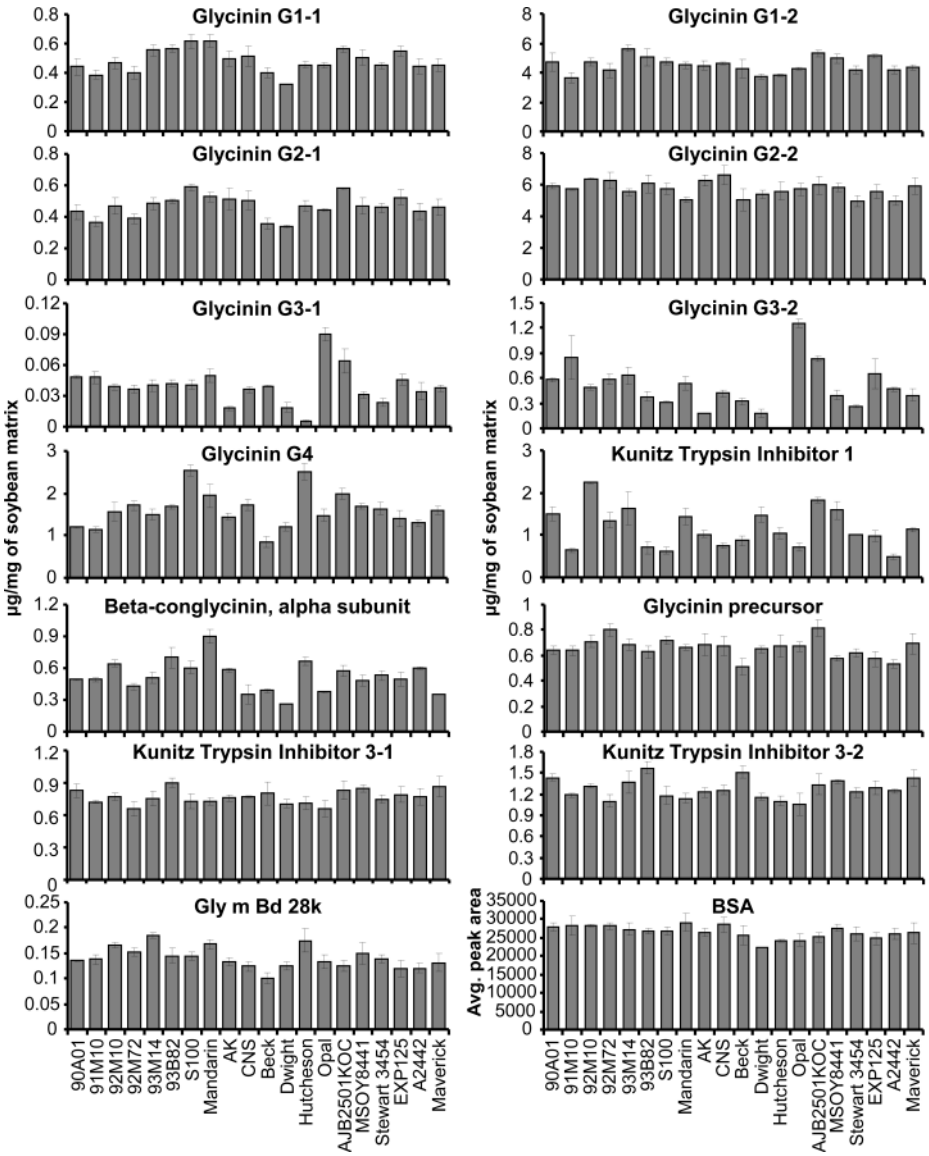

Figure 5. Absolute quantitation of nine proteins in 20 soybean varieties using AQUA peptide standards. Trypsin-digested protein spiked with AQUA peptides was analyzed by LC−MS/MS on a triple quadrupole mass spectrometer using multiple reaction monitoring. Histogram plots show the average microgram of peptide per milligram of soy protein and standard deviation from three biological replicates for each soybean variety. BSA histogram shows average peak area only.

Despite low abundances, both Gly m Bd 30K and Gly m Bd 28K are frequently recognized by soybean-sensitive individuals.[44]

Conclusion

In the past decade, significant technical advances have been made to identify the natural variation of allergen expression in major crop seeds using mass spectrometry. Natural variation is generally referred to as the differences in expression of any individual proteins that might be expected among different varieties due to plasticity in a plant's genome as well as variation due to environmental effects or effects from agronomic practices. Compared to immunoassays, proteomic methods measure allergen quantities with increased specificity, sensitivity, and capacity to measure multiple proteins simultaneously. We used two quantitative proteomic methods, spectral counting and the AQUA strategy, to evaluate the natural variation of 10 allergens (that represent all five food soybean allergens, which have been identified so far) in 20 soy varieties. Both quantitative techniques produced similar allergen expression profiles for most of the allergens analyzed. Although the absolute quantities of the studied allergens spanned over a 10-fold range, the total allergen expression level showed little variation among the soy varieties, which suggest individual allergenic protein variation may be as useful as the combined data. This proteomic analysis of 20 soy varieties showed that mass spectrometry can effectively profile the natural variation of several allergens, and, by extrapolation, the major soybean food allergens that have been identified and sequenced. In the future, quantification of each allergen using mass spectrometry may provide accurate information for soy-product labels. Also, our data collection methods along with additional proteomic studies of soy natural variation support a platform that can provide high-throughput and reliable data. Importantly, the data collection methods described herein can support experimental designs that may help to elucidate how

Table 2. AQUA Peptide with "Ragged Ends"[a]

allergenic proteins	peptide name	peptide sequence
Glycinin G1	GlyG1-1	LR.LSAEFGSLR.K
	GlyG1-2	GR.VLIVPQNFVVAAR.S
Glycinin G2	GlyG2-1	LK.LSAQYGSLR.K
	GlyG2-2	VR.NLQGENEEEDSGAIVTVK.G
Glycinin G3	GlyG3-1	LK.LSAQFGSLR.K
	GlyG3-2	PR.FYLAGNQEQEFLQYQPQK.Q
Glycinin G4	GlyG4	HR.VESEGGLIQTWNSQHPELK.C
Beta-conglycinin, alpha subunit	Bcon	LR.LITLAIPVNKPGR.F
Kunitz trypsin inhibitor 3	KTI3-1	IR.FIAEGHPLSLK.F
	KTI3-2	ER.VSDDEFNNYK.L
Kunitz trypsin inhibitor 1	KTI1	AR.DTVDGWFNIER.V
Gly m Bd 28K	AllGly28	SR.DGPLEFFGFSTSAR.K
Glycinin precursor	GlyPre-1[b]	KR.EQDQDQDEDEDEDEDQPR.K
	GlyPre-2	NR.NGLHLPSYSPYPR.M
Gly m Bd 30K	34MatureSeed-1[c]	TK.EESETLVSAR.V

[a] Five AQUA peptides sequences showed peripheral alternative cleavage sites or ragged ends, i.e., adjacent Lys or Arg residues. AQUA peptide sequences are delimited by dots, which represent cleavage sites. Ragged ends are shown underlined. [b] AQUA peptide was excluded from the analysis due to poor ionization. [c] AQUA peptide was excluded from the analysis due to low abundance.

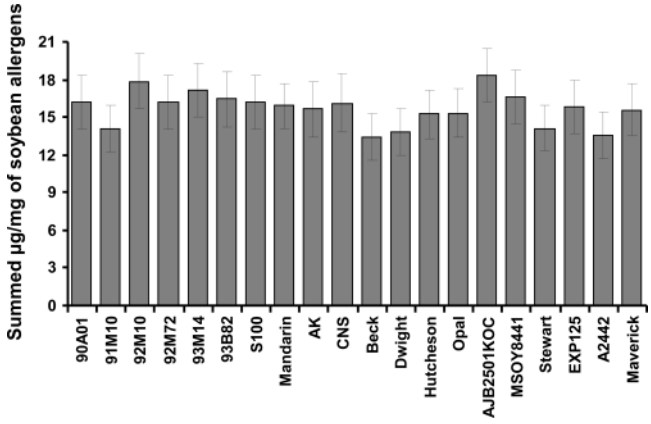

Figure 6. Summed studied food allergens for each soybean variety. Absolute quantities (μg of peptide/mg of soy protein) of each protein were summed and graphed for each soybean variety. Quantities calculated from peptides with alternative cleavage sites were excluded.

individual crop varieties compare with other varieties in their allergen content within the context of the overall range of natural variation. Also, knowledge of soybean's natural allergen content and variation may someday support a better understanding of allergen exposure thresholds.

Acknowledgment. This work was supported by a grant to J.J.T. from the International Life Science Institute-Health Environmental Science Institute, Protein Allergenicity Technical Committee, 1156 Fifteenth St., NW, Suite 200, Washington, DC 20005.

Supporting Information Available: Figure S1, linear range of four product ions for the 13 native peptides in soybean seed variety Maverick. Figure S2, linear range of four product ions for 13 AQUA peptides in soybean seed variety Maverick. Table S1, dynamic MRM parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Thelen, J. J. Proteomics tools and resources for investigating protein allergens in oilseeds. *Regul. Toxicol. Phamacol.* **2009**, *54*, S41–S45.
(2) Friedman, M.; Brandon, D. L. Nutritional and health benefits of soy proteins. *J. Agric. Food Chem.* **2010**, *49*, 1069–1086.
(3) Burks, A. W.; Brooks, J. R.; Sampson, H. A. Allergenicity of major component proteins of soybean determined by enzyme linked immunosorbent assay (ELISA) and immunoblotting in children with atopic dermatitis and positive soy challenges. *J. Allergy Clin. Immunol.* **1988**, *81*, 1135–1142.
(4) Ogawa, T.; Bando, N.; Tsuji, H.; Okajima, H.; Nishikawa, K.; Sasoka, K. Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of soybean-sensitive patients with atopic dermatitis. *J. Nutr. Sci. Vitaminol.* **1991**, *37*, 555–565.
(5) Food and Drug Administration (FDA). Food Allergen Labeling and Consumer Protection Act (FALCPA) of 2004. Congressional record v. 150 http://www.fda.gov/downloads/Food/LabelingNutrition/FoodAllergensLabeling/GuidanceComplianceRegulatoryInformation/UCM179394.pdf.
(6) European Food Safety Authority (EFSA). Opinion of the Scientific Panel on Dietetic Products, Nutrition and Allergies on a request from the Commission related to a notification from FEDIOL and IMACE on fully refined soybean oil and fat pursuant to Article 6, paragraph 11 of Directive 2000/13/EC- for permanent exemption from labeling; 2006. Request N EFSA-Q-2007-002. Adopted on 15 October 2007 by written procedure.
(7) Food Allergy Research and Resource Program (FARRP) version 10. Searchable protein database:soybean; 2010. Available at www.allergenonline.org.
(8) Ogawa, T.; Samoto, M.; Takahashi, K. Soybean allergens and hypoallergenic soybean products. *J. Nutr. Sci. Vitaminol.* **2000**, *46*, 271–279.
(9) L'Hocine, L.; Boye, J. I. Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Crit. Rev. Food Sci. Nutr.* **2007**, *47*, 127–143.
(10) Williams, P. B. Usefulness of specific IgE antibody tests: a progress report. *Ann. Allergy Asthma Immunol.* **2003**, *91*, 518–524.
(11) Poms, R. E.; Klein, C. L.; Anklam, E. Methods for allergen analysis in food: a review. *Food Addit. Contam.* **2004**, *21*, 1–31.
(12) Batista, R.; Martins, I.; Jeno, P.; Ricardo, C. P.; Oliveira, M. M. A proteomic study to identify soya allergens - the human response to transgenic versus non-transgenic soya samples. *Int. Arch. Allergy Immunol.* **2007**, *144*, 29–38.
(13) Natarajan, S.; Xu, C.; Bae, H.; Bailey, B. A. Proteomic and genomic characterization of Kunitz trypsin inhibitors in wild and cultivated soybean genotypes. *J. Plant Physiol.* **2007**, *164*, 756–763.
(14) Natarajan, S.; Xu, C.; Bae, H.; Bailey, B. A.; Cregan, P.; Caperna, T. J.; Garrett, W. M.; Luthria, D. Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes. *Plant Physiol. Biochem.* **2007**, *45*, 436–444.
(15) Krishnan, H. B.; Kim, W.-S.; Jang, S.; Kerley, M. S. All three subunits of soybean β-conglycinin are potential food allergens. *J. Agric. Food Chem.* **2009**, *57*, 938–943.
(16) Xu, C.; Caperna, T. J.; Garrett, W. M.; Cregan, P.; Bae, H.; Luthria, D. L.; Natarajan, S. Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. *J. Sci. Food Agric.* **2007**, *87*, 2511–2518.
(17) Mooney, B. P.; 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.
(18) 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.
(19) Stevenson, S. E.; Chu, Y.; Ozias-Akins, P.; Thelen, J. J. Validation of gel-free, label-free quantitative proteomics approaches: Applications for seed allergen profiling. *J. Proteomics* **2009**, *72*, 555–566.
(20) Stevenson, S. E.; Houston, N. L.; Thelen, J. J. Evolution of seed allergen quantification—From antibodies to mass spectrometry. *Regul. Toxicol. Phamacol.* **2010**, *58* (3), S36–S41.
(21) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–247.

(22) Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6940–6945.
(23) Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translation modifications. *Method* **2005**, *35*, 265–273.
(24) Shefcheck, K. J.; Callahan, J. H.; Musser, S. M. Confirmation of peanut protein using peptide markers in dark chocolate using liquid chromatography-tandem mass spectrometry (LC-MS/MS). *J. Agric. Food Chem.* **2006**, *54*, 7953–7959.
(25) Abbatiello, S.; Pan, Y.-X.; Zhou, M.; Wayne, A. S.; Veenstra, T. D.; Hunger, S. P.; Kiberg, M. S.; Eyler, J. R.; Richards, N. G. J.; Conrads, T. P. Mass spectrometric quantification of asparagine synthetase in circulating leukemia cells from acute lymphoblastic leukemia patients. *J. Proteomics* **2008**, *71*, 61–70.
(26) Langenfeld, E.; Zanger, U. M.; Jung, K.; Meyer, H. E.; Marcus, K. Mass spectrometry-based absolute quantification of microsomal cytochrome P450 2D6 in human liver. *Proteomics* **2009**, *9*, 2313–2323.
(27) Prokai, L., Jr.; Rauniyar, N.; Nguyen, V. Rapid label-free identification of estrogen-induced differential protein expression in vivo from mouse brain and uterine tissue. *J. Proteome Res.* **2009**, *8*, 3862–3872.
(28) Ferro, M.; Brugière, S.; Salvi, D.; Seigneurin-Berny, D.; Court, M.; Moyet, L.; Ramus, C.; Miras, S.; Mellal, M.; Le Gall, S.; Kieffer-Jaquinod, S.; Bruley, C.; Garin, J.; Joyard, J.; Masselon, C.; Rolland, N. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics* **2001**, *9*, 1063–1084.
(29) Schmidt, C.; Lenz, C.; Grote, M.; Lv̈hrmann, R.; Urlaub, H. Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. *Anal. Chem.* **2010**, *82*, 2784–2796.
(30) Lee, D. G.; Houston, N. L.; Stevenson, S. E.; Ladics, G. S.; McClain, S.; Privalle, L.; Thelen, J. J. Mass spectrometry analysis of soybean seed proteins: optimization of gel-free quantitative workflow. *Anal. Methods* **2010**, *2*, 1577–1583.
(31) 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.
(32) Internet Symposium on Food Allergens. Searchable allergen composition and publication database: soybean; 1999. Available at: http://www.food-allergens.de/symposium-vol1(2)/data/soy/soy-composition.htm.
(33) Odani, S.; Koide, T.; Ono, T.; Seto, Y.; Tanaka, T. Soybean hydrophobic protein. Isolation, partial characterization and the complete primary structure. *Eur. J. Biochem.* **1987**, *162*, 485–491.
(34) Kalinski, A.; Weisemann, J. M.; Matthews, B. F.; Herman, E. M. Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. *J. Biol. Chem.* **1990**, *265*, 13843–13848.
(35) Samoto, M.; Akasaka, T.; Mori, H.; Manabe, M.; Ookura, T.; Kawamura, Y. Simple and efficient procedure for removing the 34 kDa allergenic soybean protein, Gly m I, from defatted soy milk. *Biosci. Biotechnol. Biochem.* **1994**, *58*, 2123–2135.
(36) Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, V. H.; Scallon, V. J.; Fischer, R. L.; Sims, T. L.; Drews, G. N.; Goldberg, R. B. Characterization of the glycinin gene family in soybean. *Plant Cell* **1989**, *1*, 313–328.
(37) Beilinson, V.; Chen, Z.; Shoemaker, R. C.; Fischer, R. L.; Goldberg, R. B.; Nielsen, N. C. Genomic organization of glycinin genes in soybean. *Theor. Appl. Genet.* **2002**, *104*, 1132–1140.
(38) Beardsless, T. A.; Zeece, M. G.; Sarath, G.; Markwell, J. P. Soybean glycinin G1 acidic chain shares IgE epitopes with peanut allergens Ara h 3. *Int. Arch. Allergy Immunol.* **2000**, *123*, 299–307.
(39) Helm, R. M.; Cockrell, G.; Connaughton, C.; Sampson, H. A.; Bannon, G. A.; Beilinson, V.; Livingstone, D.; Nielsen, N. C.; Burks, A. W. A soybean G2 glycinin allergen: 1 Identification and characterization. *Int. Arch. Allergy Immunol.* **2000**, *123*, 205–212.
(40) Cho, T.-J.; Davies, C. S.; Fischer, R. L.; Tumer, N. E.; Goldberg, R. B.; Nielsen, N. C. Molecular characterization of an aberrant allele for the *Gy3* glycinin gene: A chromosomal rearrangement. *Plant Cell* **1989**, *1*, 339–350.
(41) Holzhauser, T.; Wackermann, O.; Ballmer-Weber, B. K.; Bindslev-Jensen, C.; Scibilia, J.; Perono-Garoffo, L.; Utsumi, S.; Poulsen, L. K.; Vieths, S. Soybean (*Glycine max*) allergy in Europe: Gly m 5 (β-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for sever allergic reactions to soy. *J. Allergy Clin. Immunol.* **2009**, *123*, 452–458.
(42) Jofuku, K. D.; Goldberg, R. B. Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. *Plant Cell* **1989**, *1*, 1079–1093.
(43) Srinivasan, A.; Giri, A. P.; Harsulkar, A. M.; Gatehouse, J. A.; Guptta, V. S. Kunitz trypsin inhibitor from chickpea (*Cicer arietinum*) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae. *Plant Mol. Biol.* **2005**, *57*, 359–374.
(44) Bando, N.; Tsuji, H.; Yamanishi, R.; Nio, N.; Ogawa, T. Identification of the glycosylation site of a major soybean allergen, Gly m Bd 30K. *Biosci. Biotech. Biochem.* **1996**, *60*, 347–348.
(45) Tsuji, H.; Bando, N.; Hiemori, M.; Yamanishi, R.; Kimoto, M.; Nishikaw, K.; Ogawa, T. Purification and characterization of soybean allergen Gly m Bd 28K. *Biosci. Biotech. Biochem.* **1997**, *61*, 942–947.
(46) Ogawa, T.; Tsuji, H.; Bando, N.; Kitamura, K.; Zhu, Y.-L.; Hirano, H.; Nishikawa, K. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci. Biothch. Biochem.* **1993**, *57*, 1030–1033.
(47) Helm, R. M.; Cockrell, G.; Herman, E.; Burks, A. W.; Sampson, H. A.; Bannon, G. A. Cellular and molecular characterization of a major soybean allergen. *Int. Arch. Allergy Immunol.* **1998**, *117*, 29–37.
(48) Hiemori, M.; Ito, H.; Kimoto, M.; Yamashita, H.; Nishizawa, K.; Maruyama, N.; Utsumi, S.; Tsuji, H. Identification of the 23-kDa peptide derived from the precursor of Gly m Bd 28K, a major soybean allergen, as a new allergen. *Biochim. Biophys. Acta* **2004**, *1675*, 174–183.
(49) Joseph, L. M.; Hymowitz, T.; Schmidt, M. A.; Herman, E. M. Evaluation of Glycine germplasm for nulls of the immunodominent allergen P34/Gly m Bd 30k. *Crop Sci.* **2006**, *46*, 1755–1763.
(50) Natarajan, S. S.; Chenping, X.; Hanhong, B.; Caperna, T. J.; Garret, W. Determination of optimal protein quantity required to identify abundant and less abundant soybean seed protein by 2D-PAGE and MS. *Plant Mol. Biol. Rep.* **2007c**, *25*, 55–62.

PR100913W