J Am Oil Chem Soc (2007) 84:1107–1115
DOI 10.1007/s11746-007-1144-3

ORIGINAL PAPER

Assessing Glycinin (11S) and β-Conglycinin (7S) Fractions of Soybean Storage Protein by Near-Infrared Spectroscopy

Stephen R. Delwiche · Lester O. Pordesimo · Dilip R. Panthee · Vincent R. Pantalone

Received: 7 May 2007 / Revised: 18 August 2007 / Accepted: 18 September 2007 / Published online: 18 October 2007
© AOCS 2007

Abstract Soybean breeding programs underway today are addressing the goal of improving the protein profile to benefit the human diet as well as that of livestock. Glycinin, a globulin storage protein of the meal and designated as the 11S size fraction by ultracentrifugation, is desirable because of its relative abundance of sulfur-containing amino acids, such as methionine and cysteine. The current study examined the feasibility of near-infrared (NIR) measurement of glycinin and the other prevalent protein fraction, β-conglycinin (7S size fraction), as well as the electrophoretically separable sub fractions that comprise these two components. From a population of 101 F_6-derived recombinant inbred lines in a field replicated trial, single whole soybeans were scanned in transmittance (800–1,798 nm, 24 beans/sample × 197 samples total). Additional scanning of the ground meal was performed in reflectance (1,100–2,498 nm). Partial least squares (PLS) calibrations were developed, using the 24-bean average log(1/*T*) spectrum for each sample, as well as the average spectrum from duplicate packs of log(1/*R*) spectra of the meal. The results indicate that NIR prediction of 11S and 7S, as well as the sub fractions thereof, is at best limited to screening purposes in soybean breeding programs for probable reasons of an inherent lack of spectral specificity of the protein fractions and a non-constant proportion of soluble-to-total protein.

Keywords Soybean · Storage proteins · Protein fractions · Glycinin · β-Conglycinin · Near-infrared · Spectroscopy

S. R. Delwiche (✉) · L. O. Pordesimo
Henry A. Wallace Beltsville Agricultural Research Center, Food Safety Laboratory, USDA-ARS, Building 303, BARC-East, 10300, Baltimore Ave., Beltsville, MD 20705-2350, USA
e-mail: Stephen.Delwiche@ars.usda.gov

D. R. Panthee · V. R. Pantalone
Department of Plant Sciences, 252 Ellington Plant Sciences, The University of Tennessee, 2431 Joe Johnson Dr., Knoxville, TN 37996, USA

Introduction

Soybean [*Glycine max* (L.) Merr.] is often used in feedstuffs for its protein, and it is finding increasing application in foods for both its protein content and protein functionality. However, soy protein is not a complete protein because it is deficient in the sulfur-containing amino acids, methionine and cysteine [1]. Consequently, ingredients possessing such components from non-soy sources are often needed in feedstuff/food formulations. Increasing the concentration of these two amino acids would improve the nutritional value of soybean meal as well as enhance the quality of soy protein for food applications such as tofu.

The greater proportion of protein in soybean is storage protein, which provides a source of nitrogen and carbon for the plant. Soybean storage protein has two main fractions: β-conglycinin and glycinin, both of which are globulins. These two fractions account for greater than 65% of the total soybean protein [2]. Based on size separation by ultracentrifugation, these fractions are designated as 7S and 11S, respectively. Glycinin comprises 25–35% of the total seed protein and is the largest single storage protein fraction [3, 4]. Monomeric subunits of glycinin consist of acidic and basic polypeptides while β-conglycinin has three prevalent types of subunits, designated as α′, α, β. Glycinin has three to four times more sulfur containing amino acids than β-conglycinin [5, 6]. In contrast, the β subunit of

β-conglycinin is known to be devoid of methionine and cysteine [7], with only traces of these amino acids in the α' and α subunits. Unlike 7S, it is not clear which of the glycinin polypeptides contains more sulfur-containing amino acids. Therefore, as suggested by Panthee et al. [8], the glycinin fraction may be more desirable than β-conglycinin in improving the amino acid balance in soybeans. Considering that there is a typical inverse relationship between 7S and 11S concentrations, glycinin can be increased at the expense of β-conglycinin. In their study, Panthee et al. actually found glycinin content to be high in their high protein lines irrespective of the β-conglycinin content. Some plant breeders have already taken advantage of this relationship in developing lines with increased glycinin [9, 10]. Apparently, the ratio of 11S to 7S could be a single important value representing this relationship, which may be considered in developing new lines high in sulfur-containing amino acids. Furthermore, Kwanyuen and Wilson [11] stated that 11S:7S may be regarded as a predictive indicator of relative protein nutritional quality and functionality.

It would help breeders tremendously if there was a rapid method of assessing glycinin or β-conglycinin fractions of total soybean protein for screening purposes at the early stages of plant breeding. The fastest method for determining total seed crude protein and oil contents is near-infrared (NIR) spectroscopy. Pazdernick et al.'s [12] attempts at developing glycinin calibration equations using ground samples and NIR reflectance were a promising first step. Preferably, such determinations without the need to grind samples would be more desirable, considering a significant savings in time from not grinding and the avoidance of cleaning out between grinds. Therefore, the objective of this study was to evaluate whether NIR transmission through single beans would provide enough information for screening samples for glycinin and β-conglycinin fractions of soybean protein. Whole grain NIR spectroscopy is additionally advantageous because of sample preservation [13]. Plant breeders could use a non-destructive, single seed NIR spectroscopy [14–17] to screen genotypes without sacrificing the limited amount of seed.

Materials and Methods

Sample Material

Selections from a population studied by Panthee [18] comprised the samples analyzed in this study. The population consisted of a total of 101 F_6-derived recombinant inbred lines (RIL) developed from a cross of N87-984-16 × TN93-99. The N87-984-16 line is one of two sister lines, whose blend constitutes the high protein commercial cultivar "Prolina" [19]. The line TN93-99 is a high yielding and well adapted to Tennessee breeding line and is currently registered as a germplasm [20]. These RIL were planted in a randomized complete block design (RCBD) with three replications at the East Tennessee Research and Education Center in Knoxville, Tennessee in May 2003. Each line was planted in a four-row plot of 6-m length with 75 cm between rows. Seeds were harvested with a two-row combine, dried to 13% moisture, and then stored at below 10 °C until analyzed. Selections for this study were made from residual samples from the 2003 planting. A total of 197 samples were obtained. The selections included replications of some of the lines.

Protein Extraction

Ten grams of soybean seeds were ground in a water-cooled (20 °C) Knifetec[1] 1,095 Sample Mill (FOSS Tecator S-26321, Höganäs, Sweden) for 20 s. This setting produced soybean flour with relatively uniform particle size. Soluble protein was extracted for 1 h at room temperature while stirring from one gram of full-fat soybean flour in a 1:15 (w/v) ratio with 0.2 M Tris–HCl buffer, pH 8.0, that contained 0.1 M β-mercaptoethanol. The mixture was centrifuged (10,000×g) for 10 min at 4 °C. Upon removal of the fat layer, an aliquot (1 mL) of crude protein extract or supernatant was taken from each sample. Storage proteins and their polypeptides were dissociated in the crude extract by adding an equal volume of both 5% SDS and 0.1 M β-mercaptoethanol solution to each sample and boiling in a water bath for 10 min. Glycerol and bromphenol blue were added to each sample to achieve a final concentration of 10 and 0.025%, respectively. Additional details are provided in Panthee et al. [8].

Separation of Protein Polypeptides and Subunits

Polyacrylamide Gel Electrophoresis

Proteins and their polypeptides were separated by SDS-PAGE Protean II vertical slab gel apparatus (Bio-Rad, Richmond, CA) according to the procedure of Chua [21], with the following modifications. Each sample, containing ~80–100 µg protein (~10 µL of protein sample), was loaded onto the gel. Proteins and polypeptides were separated using a linear gradient of 10–20% polyacrylamide gel. The dimensions of the separating gel were

[1] Trade names are provided for the purpose of description only and does not imply endorsement by the U.S. Department of Agriculture.

$14 \times 16 \times 0.15$ cm with 15 sample wells in the stacking gel. Blank sample wells were left between loaded samples to prevent protein cross-contamination during electrophoresis and to facilitate accurate quantitation by scanning densitometry after electrophoresis. Because more than one gel was required for the protein samples, and because it was necessary to obtain the same polyacrylamide composition throughout the entire experiment, identical gradient gels were cast from the same polyacrylamide solution and from the same gradient-producing condition. Electrophoresis of each protein sample was carried out in duplicate at a constant current of 10 mA/gel at room temperature until the bromphenol blue tracking dye reached the bottom of the gel.

Staining and Destaining

Gels were fixed in 40% (v/v) methanol and 10% (v/v) acetic acid for at least half an hour before staining. They were stained for at least 8 h in freshly prepared dye containing 0.25% (w/v) Coomassie Brilliant Blue, 40% (v/v) methanol, and 10% (v/v) acetic acid on an orbit shaker with fixed speed. Gels were then destained in 40% (v/v) methanol and 10% (v/v) acetic acid on the same orbital shaker. The destaining solution was changed every 2 h for a minimum of four changes. Destaining was terminated when the background gels were almost visibly clear of the dye. The residual dye on the gels did not interfere with the analysis of polypeptide bands since local average background subtraction was considered in the analyses. Destained gels were soaked in deionized water for at least 5 min. Each gel was placed between two cellophane sheets and dried in a Bio-Rad GelAir dryer.

Scanning Densitometry

Dried gels were scanned as described by Kwanyuen and Wilson [11] with a Molecular Dynamics Personal Densitometer SI (Molecular Dynamics, Sunnyvale, CA) equipped with a He–Ne laser light source. ImageQuant (Molecular Dynamics) software for volume integration was used in data analysis to determine total absorbance of entire protein bands. The absorbance of protein on gels was adjusted to ensure that the most intense protein bands were within the optimal and linear response range of the detector. Apparent absorbance of each protein was obtained by subtracting the background absorbance from the total absorbance of the protein bands within the same gel volume. The relative amount of each protein fraction was expressed as a percentage of total soluble protein in the same gel lane.

The β-conglycinin (7S) content was calculated as the summation of the original scanned values for α', α, and β while the glycinin (11S) content was calculated as the summation of the acidic and basic components. These polypeptide subunits were also expressed as a percentage of the total protein represented in the gel lane. Thus, each gel lane represented the total soluble protein. A simplification was made that treated all samples as having the same fraction of soluble to total protein, thus allowing for the conversion of densitometry protein fraction values to protein concentration through the application of each sample's total protein value from combustion analysis. The protein fraction percentages were converted to mass units (g/kg soy meal, as-is ambient room moisture basis) using the total protein analyses determined by combustion. Although total soluble protein concentration of each sample was actually determined following the procedure described by Bradford [22] as the reference for properly loading the gels, this data was not retained. Alternatively, the measured crude protein content was used in the calculation of protein fraction concentrations. It was recognized that this would result in values lower than reported in the literature. For combustion protein content analysis ($N \times 6.25$), 150 mg was drawn from the 24-bean grind of each sample, using a Perten 3100 hammer mill (Huddinge, Sweden) equipped with a 2-mm screen. A Leco TruSpec carbon–nitrogen analyzer (St Joseph, MI), calibrated with EDTA and checked with company-certified ground soybean meal, recorded the protein content of each sample. The ratio of 11S to 7S (11S:7S, glycinin to β-conglycinin) was calculated in turn.

Spectral Measurement

Spectra of each sample were collected from both ground whole beans and single intact beans. A NIR Systems Model 6500 Vis/NIR scanning monochromator (Silver Spring, MD) equipped with appropriate attachments (described below) was used for spectral measurement.

Twenty-four randomly selected beans of each sample were scanned. Each of the 24 beans of a sample was scanned in transmission. In this application, a pharmaceutical tablet attachment was coupled to the spectrometer in place of the spinning cup module. The individual beans were mounted onto a tablet cartridge having an aperture of 6.5 mm. To minimize stray light, gaps between the bean and aperture rim were filled with black moldable material (Silly Putty, Crayola Co., Easton, PA), thus leaving an approximate opening of about 6.4 mm in diameter. Each bean was mounted such that the plane separating its two cotyledons was oriented perpendicular to the incident radiation beam and its azimuthal angle was held constant.

Kernel transmittance readings were referenced to energy readings from two stacked panels of 1.5-mm-thick Teflon sheeting, with spectra stored as log(1/*T*) over a scan range of 600–1,898 nm at 2 nm increments. Additional details of this apparatus are found in Delwiche et al. [23]. The purpose of making single bean scans was twofold: first, to determine whether it is possible to develop reasonable NIR calibrations from spectra that are formed as the average individual bean spectra; and if this is successful, second, to determine the possibility of reasonable calibrations that are based on single bean spectra. Lastly, mass of each scanned bean was recorded to the nearest 0.01 mg.

A spinning cup module was used for the scanning of the ground meal. The meal from the grind of all 24 beans was scanned (32 scans per spectrum, referenced to ceramic) over a wavelength range of 1,100–2,498 nm at 2-nm increments. Two spectra were recorded for each sample, with the meal unloaded and reloaded between spectra. Replicate spectra were later averaged before chemometric analysis.

Chemometric Analyses

Of the 197 samples, 147 by random selection were used in NIR chemometric model development; the remaining 50 were used to test the performance of the models developed. Separate models were developed for each measured constituent and calculated variables involving 24-bean average transmittance spectra [log(1/*T*)] and ground bean reflectance [log(1/*R*)] spectra. The spectral pretreatment for each set of spectra was a Savitzky-Golay second derivative (second order polynomial), with a 15-point (28 nm) and 11-point (20 nm) convolution window for the transmittance and reflectance spectra, respectively. For the transmittance samples, the wavelength range of interest was truncated to 800–1,798 nm for single bean spectra. This truncation eliminated spectral regions where no differences between samples are evident as well as regions where there was significant noise. Partial least squares (PLS) regression (theory described by Lindberg et al. [24] and Martens and Naes [25]) on mean-centered spectra was used in developing all chemometric models. PLS regression was accomplished in MATLAB (Version 7, The Mathworks Inc., Natick, MA) with a third-party chemometrics toolbox (PLSToolbox, Version 3.5, Eigenvector Research Inc., Wenatchee, WA). A unique set of PLS factors (latent variables) was developed for each constituent.

Full cross-validation was performed during calibration development, whereby each sample was temporarily removed, predicted, and then restored to the calibration set. Performance statistics were accumulated for each group of removed samples. The optimal number of factors (latent variables) for each constituent was that which either produced a minimum overall error between modeled and reference values for the samples removed during cross-validation or produced comparable results as those for the minimum factor, but with fewer factors. Model performance was reported as the coefficient of determination (R^2), the square root of the mean of the squared residuals of the calibration set (RMSEC), the square root of the mean of the squared residuals of cross-validation (RMSECV), the square root of the mean of the squared residuals of the validation set (RMSEP), and the bias of the validation set samples (Bias), which is the difference between the average value of the reference determinations and the average value of the NIR predictions.

Results and Discussion

Analysis of Storage Protein Polypeptides and Subunits

Univariate statistics for the two storage proteins studied and their subunits are presented in Table 1. These statistics were calculated separately for the calibration and validation subsets, as well as for the combined set. The ranges for the constituents of the validation samples were comparable to the corresponding ranges of the calibration samples, which indicates that the random selection of samples for each set components was appropriate. Also included in Table 1 is level of correlation between the concentration of a protein fraction and total protein content itself. The highest correlation occurred with the acidic fraction of glycinin ($r = 0.53$), followed by glycinin itself ($r = 0.51$) and β-conglycinin ($r = 0.42$), with all correlations except glycinin-to-β-conglycinin significant at $P = 0.001$ or smaller. The purpose of establishing the degree of correlation with protein content was to anticipate that the success of an NIR model of any of these protein constituents could be attributed to a sensitivity of the NIR response to general protein structure rather than that specific to the component.

Spectral Analysis

The average second derivative spectra for whole bean transmittance (197 samples × 24 beans/sample) and ground bean reflectance (197 samples × 2 packs/sample) are shown in Fig. 1a, b, respectively. A ±2σ envelope [with σ being the standard deviation of the 197-member set of 24-bean average spectra (1A) and the 197-member set of the 2-pack average spectra (1B)], plotted about each mean spectrum, is also included in both graphs. Because of the

Table 1 Descriptive statistics of soybean protein components from a population of N87-984-16 × TN03-99

Constituent (units)	Set[a]	Mean	SD	Minimum	Maximum	Correlation to protein content
Protein content (100 × g/g)	A	38.89	1.04	35.0	41.5	–
	C	38.99	1.06	35.0	41.5	
	V	38.60	0.91	35.7	40.4	
Glycinin (g/kg)	A	135.24	12.25	108.0	177.8	0.513
	C	136.09	12.53	110.3	177.8	
	V	132.73	11.12	108.0	163.3	
Acidic fraction (g/kg)	A	75.31	7.08	60.8	104.1	0.533
	C	75.89	7.34	61.7	104.1	
	V	73.61	5.98	60.8	90.5	
Basic fraction (g/kg)	A	59.93	5.94	45.6	75.3	0.421
	C	60.20	5.99	46.6	75.3	
	V	59.12	5.79	45.6	72.8	
β-Conglycinin (g/kg)	A	87.25	12.35	64.8	128.8	0.460
	C	88.34	12.86	65.3	128.8	
	V	84.03	10.15	64.8	109.6	
α'-Fraction (g/kg)	A	19.92	3.62	10.7	28.8	0.428
	C	20.23	3.79	10.9	28.8	
	V	19.01	2.91	10.7	27.2	
α-Fraction (g/kg)	A	35.13	4.49	22.7	48.0	0.226
	C	35.31	4.64	22.7	48.0	
	V	34.59	4.02	27.0	44.6	
β-Fraction (g/kg)	A	32.20	6.84	17.4	57.1	0.456
	C	32.80	7.00	17.4	57.1	
	V	30.43	6.04	19.0	55.1	
Glycinin-to-β-conglycinin ratio	A	1.57	0.16	1.1	2.0	–0.168
	C	1.56	0.16	1.1	2.0	
	V	1.59	0.15	1.2	2.0	

[a] A = all samples (n = 197), C = calibration-cross-validation samples (n = 147), V = validation samples (n = 50)

low spectral variation in the derivative transformation of the reflectance spectra, all three curves of the lower graph appear as one overlapping trace. Compared with the ground meal spectra, the greater degree of spectral variation of the whole bean transmittance spectra is attributed to the variation in bean mass, as confirmed by a positive correlation (r = 0.62) between soybean mass and the spectral absorption at 1,650 nm, which is a region of non-specific absorption. Although the single bean derivative spectra possessed a greater level of variance, it is interesting to note that the regions of peak absorption, designated as local minima, are essentially common between the upper and lower graphs. Thus, absorption bands at 1,164, 1,210, 1,434, 1,496, 1,708, 1,726, and 1,760 nm were sensed in both reflectance of ground meal and transmittance of the whole bean. These bands are generally attributed to first and second overtone stretches of C–H and O–H, with the additional attribution of a first overtone stretch of protein N–H at 1,496 nm [26].

Results of the PLS modeling are summarized for ground meal spectra in Table 2 and for averaged single-bean spectra in Table 3. For each constituent, the RMSEP was determined when the best model from cross-validation was applied to the validation samples. Considering ground meal spectra first, it is seen that the calibrations for protein content had R^2 values of 0.73 and 0.90, depending on whether all 147 calibration and 50 validation samples were included or, in the second case, whether 3 residual outliers (absolute value of residual > 3× RMSD) were removed from the calibration set and 2 from the validation set. This resulted in corresponding RMSEP values of 0.60 and 0.32%. Such levels of model performance are on par with other reported NIR protein calibrations for soybean meal [27].

Of the protein fractions, the best performance in PLS modeling was obtained with glycinin concentration, though this was substantially lower (R^2 = 0.30) than the calibrations for protein content. All other fractions yielded even lower levels of performance, with R^2 ranging from 0.29 for the concentration of the acidic fraction of glycinin, to 0.20 for β-conglycinin concentration, and down to 0.06 for the concentration of the α fraction of β-conglycinin. Examination of the wavelength-by-wavelength products of

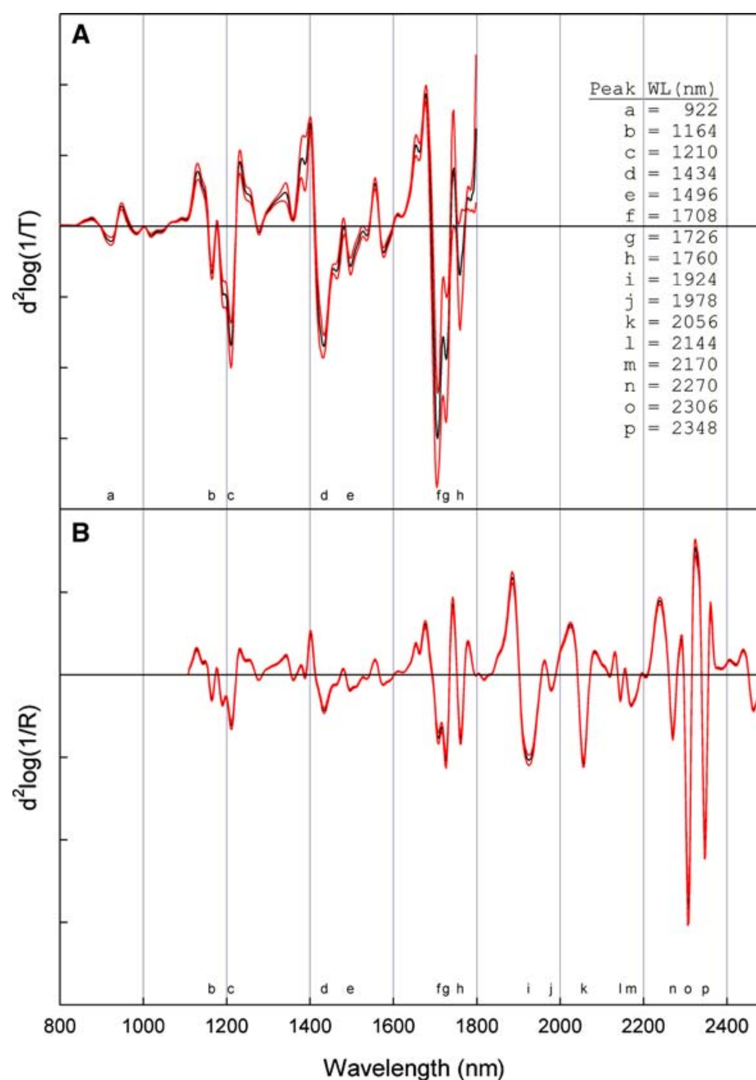

Fig. 1 Average second derivative spectrum of 24-bean average transmittance spectra (*upper*) and ground meal reflectance spectra (*lower*) of all samples used in study (n = 197). A ±2 standard deviation envelope about the mean is included in each graph

the average second derivative spectrum and the PLS regression vector of the calibrations for protein, glycinin, and β-conglycinin (Fig. 2) yields the finding of an overall similarity in the calibration equations. Thus, the modeling success of the protein fractions, albeit very modest, appears to arise from the success of the model for protein content itself. The poorest performance was noted for the non-concentration component, the glycinin-to-β-conglycinin

Table 2 Summary of partial least squares (PLS) models for protein components, based on ground meal reflectance spectra

Constituent (units)	Latent variables	Calibration-cross-validation set[a]			Validation set[a]	
		R^2	RMSEC	RMSECV	RMSEP	Bias
Protein content (PC) (100 × g/g)	3	0.732	0.55	0.58	0.60	0.12
PC, less outliers[b] (100 × g/g)	4	0.901	0.31	0.34	0.32	0.02
Glycinin (g/kg)	3	0.305	10.4	10.9	11.9	4.6
Acidic (g/kg)	3	0.288	6.2	6.5	6.3	2.5
Basic (g/kg)	3	0.270	5.1	5.3	6.2	2.1
β-Conglycinin (g/kg)	2	0.198	11.5	12.1	11.0	3.6
α' (g/kg)	3	0.225	3.3	3.5	3.0	0.8
α (g/kg)	2	0.060	4.5	4.7	4.4	1.3
β (g/kg)	3	0.266	6.0	6.3	6.3	1.5
11S/7S	1	0.022	0.16	0.16	0.15	–0.01

[a] n = 147 and 50 for the calibration-cross-validation and validation sets, respectively

[b] Three samples removed from the calibration set, two samples removed from the validation set

Table 3 Summary of PLS models for protein components, based on 24-bean average spectra

Constituent (units)	Latent variables	Calibration-cross-validation set[a]			Validation set[a]	
		R^2	RMSEC	RMSECV	RMSEP	Bias
Protein content (PC) (100 × g/g)	6	0.688	0.59	0.64	0.78	0.24
PC, less outliers[b] (100 × g/g)	6	0.822	0.41	0.46	0.51	0.06
Glycinin (g/kg)	7	0.296	10.5	11.8	11.8	–0.18
Acidic (g/kg)	8	0.320	6.0	6.9	6.1	–0.3
Basic (g/kg)	7	0.254	5.1	5.7	6.4	–0.6
β-Conglycinin (g/kg)	1	0.218	11.3	11.6	10.5	2.4
α' (g/kg)	2	0.337	3.1	3.2	3.0	–0.5
α (g/kg)	1	0.040	4.5	4.6	4.0	0.4
β (g/kg)	1	0.214	6.2	6.3	6.3	1.4
11S/7S	1	0.124	0.15	0.15	0.16	–0.01

[a] n = 147 and 50 for the calibration-cross-validation and validation sets, respectively

[b] Three samples removed from the calibration set, two samples removed from the validation set

ratio, with $R^2 = 0.02$. The small number of PLS latent variables, ranging from 1 to 4, is a telltale indication of the difficulty with NIR being able to discern individual protein fractions.

When the PLS models derived from the averaged single bean spectra are considered (Table 3), it is not surprising to discover lower modeling performance. Again, the best performance was associated with the calibrations for protein content [$R^2 = 0.69$ (all calibration and validation samples included) and 0.82 (3 and 2 residual outliers removed from calibration and validation sets, respectively, which were the same outlying samples in the ground meal analysis)]. The corresponding RMSEP values (0.78 and 0.51%) were slightly higher than those for ground meal. For the protein fractions, model performance was again low, with R^2 ranging from 0.04 (α fraction of β-conglycinin) to 0.34 (α' fraction of β-conglycinin). For glycinin and β-conglycinin, the R^2 values were 0.30 and 0.22, respectively. These values are quite similar to the corresponding values from the ground meal calibrations, despite the observation that the (second derivative × regression coefficient) curves for these two minor constituents (Fig. 3) have smaller resemblance to the corresponding curve for protein content than the similarities among the curves for the ground meal calibrations (Fig. 2). Given the low modeling performance of models based on either ground meal reflectance or averaged whole bean transmittance, single bean analysis was not attempted. Though single bean analysis has been applied for macro constituent determination, such as moisture, with some degree of success [13], the results of the current study on protein constituents allude to the difficulty, if not impossibility, of their measurement by single bean NIR analysis. Results by Abe et al. [28] also put this into perspective. Obtaining an SEP of 0.77 and 0.67% in predicting protein in wheat and soybean, respectively, using single kernel spectra, Abe et al. [28] concluded that the results were insufficient to be of practical use.

Pazdernick et al. [12] found better apparent success than the current study in developing NIR calibrations for glycinin content. However, the reported values of R^2, ranging from 0.71 to 0.84, were associated with calibrations in which 12–13 samples were removed as outliers from a set of 98 samples, which beforehand had undergone spectral outlier removal and "selection" procedures starting with a set of 319 genotypes. When an equivalent fraction of samples (20 of 147, based on their high residual values) were manually removed from the current study's calibration set, the resulting R^2 value of 0.83 was in agreement with Pazdernick et al. [12].

Another contributing reason to the difficulty in chemometric modeling of protein fractions is in the use of total protein (crude protein) by combustion when converting the storage protein subunits, initially expressed as the percentage of protein in an electrophoresis gel lane, into mass units. The protein represented by a gel lane is actually just the soluble protein, and soluble protein is a major, though not complete, fraction of total protein. This approach loses validity when the fraction of soluble-to-total protein is not constant across samples.

Conclusion

Models based on single bean transmission spectra respectively weighted by individual bean mass failed to accurately predict soybean storage protein fraction measurements from bulk material. NIR spectroscopy involving single bean transmission spectra therefore does not show promise as a tool for fast, non-destructive screening of lines for high glycinin concentration at early stages of plant breeding. This limitation stems from an apparent lack of specificity to the spectral sensing of the protein fractions, be it measured by whole bean transmittance or ground meal reflectance. Liu [2] noted that because of the different

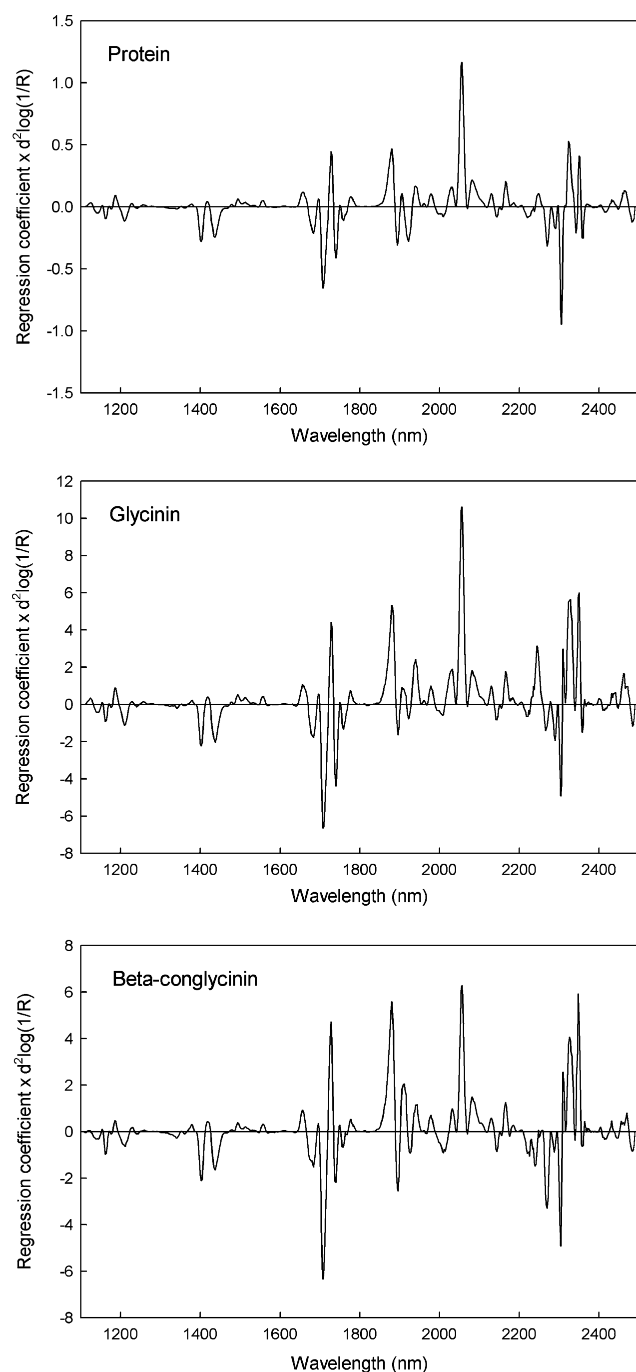

Fig. 2 (Average second derivative spectrum × PLS regression coefficient) vectors for the protein content (*upper*), glycinin (*middle*), and β-conglycinin (*lower*) calibrations, based on ground meal reflectance spectra

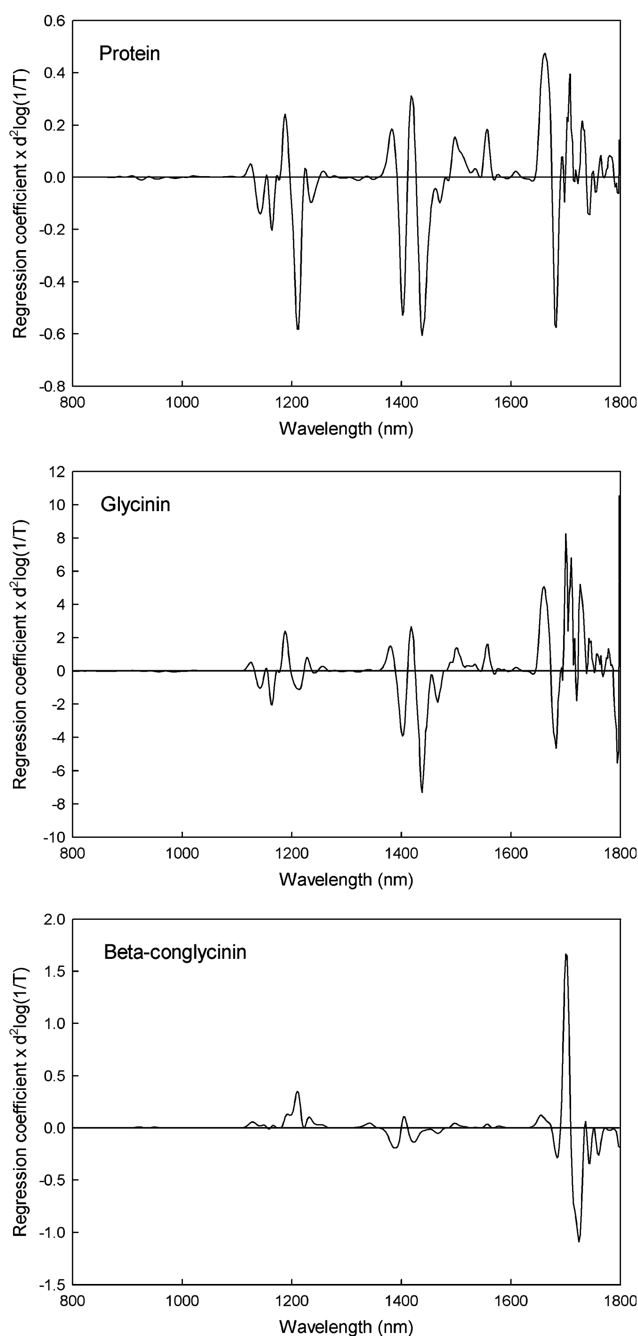

Fig. 3 (Average second derivative spectrum × PLS regression coefficient) vectors for the protein content (*upper*), glycinin (*middle*), and β-conglycinin (*lower*) calibrations, based on whole bean transmittance spectra

methods used in the isolation of the different protein fractions, reports by different researchers on the characterization of soy protein fractions are sometimes controversial. This appears to be an issue in this study as well. Further studies could be conducted that utilize a direct measurement of soluble protein concentration, rather than a reliance on total protein content (by combustion, or Kjeldahl); however, it is unclear whether an NIR method for anything beyond screening of soybean lines for desirable protein characteristics will be practical.

Acknowledgments The authors give thanks to B.E. Stetzler (ARS, Beltsville) for NIR reflectance and combustion protein measurements.

References

1. Kwanyuen P, Pantalone VR, Burton JW, Wilson RF (1997) A new approach to genetic alteration of soybean protein composition and quality. J Am Oil Chem Soc 74:983–987
2. Liu K (1997) Soybeans: chemistry, technology and utilization. Chapman & Hall, London
3. Murphy PA, Resurrecion AP (1984) Varietal and environmental differences in soybean glycinin and β-conglycinin concentration. J Agric Food Chem 32:911–915
4. Yagazaki K, Takagi T, Sakai M, Kitamura K (1997) Biochemical characterization of soybean protein consisting of different subunits of glycinin. J Agric Food Chem 45:656–660
5. Kitamura K (1995) Genetic improvement of nutritional and food processing quality in soybean. Jap Agric Res Quart 29:1–8
6. Wilson RF (2004) Seed composition. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production and uses, 3rd edn. Agron. Monogr. 16. ASA, CSSA, SSA, Madison, pp 621–677
7. Krishnan HB (2000) Biochemistry and molecular biology of soybean seed storage proteins. J New Seeds 2:1–25
8. Panthee DR, Kwanyuen P, Sams CE, West DR, Saxton AM, Pantalone VR (2004) Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. J Am Oil Chem Soc 81:1005–1012
9. Takahashi K, Mizumo Y, Yumoto S, Kitamura K, Nakamura S (1996) Inheritance of the a-sub-unit deficiency of ß-conglycinin in soybean (*Glycine max* L) line induced by γ-ray irradiation. Breed Sci 46:251–255
10. Kitagawa S, Ishimoto M, Kikuchi F, Kitamura K (1991) A characteristic lacking or decreasing remarkably 7S globulin subunit induced with X-ray irradiation in soybean seeds. Jpn J Breed 41:460–461
11. Kwanyuen P, Wilson RF (2000) Optimization of coomassie staining for quantitative densitometry of soybean storage proteins in gradient gel electrophoresis. J Am Oil Chem Soc 77:1251–1254
12. Pazdernick DL, Plehn SJ, Halgerson JL, Orf JH (1996) Effect of temperature and genotype on the crude glycinin fraction (11S) of soybean and its analysis by near-infrared reflectance spectroscopy. J Agric Food Chem 44:2278–2281
13. Lamb DT, Hurburgh CR (1991) Moisture determination in single soybean seeds by near-infrared transmittance. Trans ASAE 34:2123–2129
14. Delwiche SR, Massie DR (1996) Classification of wheat by visible and near-infrared reflectance from single kernels. Cereal Chem 73:399–405
15. Delwiche SR (1998) Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. J Cereal Sci 27:241–254
16. Delwiche SR (1995) Single wheat kernel analysis by near-infrared transmittance: Protein content. Cereal Chem 72:11–16
17. Pearson TC, Wicklow DT, Maghirang EB, Xie F, Dowell FE (2001) Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. Trans ASAE 44:1247–1254
18. Panthee (2005) DR genetic mapping of quantitative trait loci conditioning protein concentration and quality, and other seed characteristics in soybean [*Glycine max* (L.) Merrill]. Ph.D. dissertation, The Univ of Tennessee, Knoxville, USA
19. Burton JW, Carter TE, Wilson RF (1999) Registration of Prolina soybean. Crop Sci 39:294–295
20. Pantalone VR, Allen FL, Landau-Ellis D (2003) Registration of 'TN93–99' Soybean germplasm. Crop Sci 43:1137
21. Chua NH (1980) Electrophoretic analysis of chloroplast proteins. Meth Enzymol 69:434–446
22. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254
23. Delwiche SR, Pordesimo LO, Scaboo AM, Pantalone VR (2006) Measurement of inorganic phosphorus in soybeans with near-infrared spectroscopy. J Agric Food Chem 54:6951–6956
24. Lindberg W, Persson JA, Wold S (1983) Partial least squares method for spectrofluorimetric analysis of mixtures of humic acid and lignosulfonate. Anal Chem 55:643–648
25. Martens H, Næs T (1987) Multivariate calibration by data compression. In: Williams PC, Norris KH (eds) Near-infrared technology in the agricultural and food industries. Am Assoc Cereal Chem, St Paul, pp 57–87
26. Osborne BG, Fearn T (1986) Near infrared spectroscopy in food analysis. Longman Scientific and Technical, Harlow Essex, pp 20–42
27. Panford JA, Williams PC, deMan JM (1988) Analysis of oilseeds for protein, oil, fiber and moisture by near-infrared reflectance spectroscopy. J Am Oil Chem Soc 65:1627–1634
28. Abe H, Kusama T, Kawano S, Iwamoto M (1995) Non-destructive determination of protein content in a single kernel of wheat and soybean by near infrared spectroscopy. In: Davies AMC, Williams PC (eds) Near infrared spectroscopy: the future waves. NIR, Chichester, pp. 457–461