JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Evaluation of Nonstarch Polysaccharides and Oligosaccharide Content of Different Soybean Varieties (*Glycine max*) by Near-Infrared Spectroscopy and Proteomics

KRISTIN HOLLUNG,*,† MARGARETH ØVERLAND,‡ MILICA HRUSTIĆ,§ PETAR SEKULIĆ,§ JEGOR MILADINOVIĆ,§ HARALD MARTENS,† BJØRG NARUM,† STEFAN SAHLSTRØM,† METTE SØRENSEN,‡ TROND STOREBAKKEN,‡ AND ANDERS SKREDE‡

Matforsk, The Norwegian Food Research Institute, Osloveien 1, N-1430 Ås, Norway, Aquaculture Protein Centre, Centre of Excellence, Post Office Box 5003, N-1430 Ås, Norway, and Institute of Field and Vegetable Crops, Maksima Gorkog 30, 21000 Novi Sad, Serbia and Montenegro

A total of 832 samples of soybeans were screened by near-infrared (NIR) reflectance spectroscopy, to identify soybean samples with a lower content of oligosaccharides and nonstarch polysaccharides (NSP). Of these, 38 samples were identified on the basis of variation in protein content and agronomic value and submitted to high-resolution NIR spectroscopy. On the basis of the NIR data, 12 samples were further selected for chromatographic characterization of carbohydrate composition (mono-, di-, and oligosaccharides and NSP). Their soluble proteins were separated by two-dimensional gel electrophoresis (2DE). Using partial least-squares regression (PLSR), it was possible to predict the content of total NSP from the high-resolution NIR spectra, suggesting that NIR is a suitable and rapid nondestructive method to determine carbohydrate composition in soybeans. The 2DE analyses showed varying intensities of several proteins, including the glycinin G1 precursor. PLSR analysis showed a negative correlation between this protein and insoluble NSP and total uronic acid (UA).

KEYWORDS: Near-infrared spectroscopy; soybean; *Glycine max*; carbohydrate composition; proteomics; nonstarch polysaccharides

INTRODUCTION

Soybean meal is a major source of protein in diets for monogastrics throughout the world, but its use for certain species is limited by the presence of a wide variety of antinutritional factors (*1*, *2*). Reduced growth performance, increased viscosity of digesta, reduced nutrient digestibility (*3*, *4*, *5*), increased incidence of diarrhea (*6*), and undesirable morphological changes of the intestinal epithelium (*7*, *8*, *9*) have been observed when soybean meal has been used in substantial amounts in feed for monogastrics. While the growth-depressing effect of heat-labile antinutritional factors in soybeans can be mostly overcome by processing, heat-stable antinutritional factors, such as nonstarch polysaccharides (NSP) and oligosaccharides, are not eliminated (*1*, *10*). Raw soybean meal contains about 30% carbohydrates: approximately 20% NSP and 10% oligosaccharides (*11*). Soybean NSP consist mainly of arabinans, arabinogalactans, and acidic polysaccharides; approximately one-third of the NSP are soluble (*12*). The primary oligosaccharides in soybeans are the galacto-oligosaccharides, raffinose, stachyose, and verbascose (*13*). Both oligosaccharides and NSPs are known to cause digestive disorders and reduced performance in monogastrics (*3*, *7*, *14*). Identifying soybean varieties with low levels of NSP and oligosaccharides would increase the potential for soybeans as an alternative protein source for monogastrics.

Near-infrared (NIR) spectroscopy is a tool to predict chemical composition of various feed ingredients. Rapid assessment of feed ingredients enables accurate feed formulation and can be used by plant breeders to select new cultivars. While NIR spectroscopy has traditionally been used to predict gross chemical content of feed ingredients (*15*), limited information exists on the use of NIR spectroscopy to predict carbohydrate composition. To our knowledge, no published papers address the efficacy of NIR to predict the content of complex carbohydrates in soybeans.

Proteomics is a powerful tool allowing identification of proteins involved in the determination of functionality of an organism (*16*, *17*). Proteomics of mature seeds provide more general information of the proteome than proteomics of metabolizing organisms such as growing plants because protein turnover is lower in mature seeds. Recently, a proteome reference map for soybean was published, which will be useful

* To whom correspondence should be addressed: Matforsk, Osloveien 1, N-1430 Ås, Norway. Telephone: +47-64-97-01-42. Fax: +47-64-97-03-33. E-mail: kristin.hollung@matforsk.no.
† The Norwegian Food Research Institute.
‡ Aquaculture Protein Centre.
§ Institute of Field and Vegetable Crops.

10.1021/jf051438r CCC: $30.25
Published on Web 10/14/2005

Table 1. Predicted Content of Moisture, Crude Protein, Fat, Carbohydrates, and Ash in the 38 Soybean Samples (% Wet Weight)

sample	genotype	variety	moisture	crude protein	crude fat	crude carbohydrates	ash
1	NS-L-115	JELICA	7.7	40.7	23.7	23.2	4.8
2	NS-L-66	KRAJINA	8.3	42.4	21.5	22.9	5.0
3	NS-L-200143	FORTUNA	8.7	44.0	19.7	22.5	5.1
4	NS-L-400007	MELI	8.7	42.2	21.2	23.2	4.7
5	NS-L-2023	AFRODITA	8.4	40.6	21.9	24.5	4.7
6	NS-L-101095	PROTEINKA	8.2	41.2	22.5	23.2	5.0
7	NS-L-201149	VALJEVKA	8.3	39.3	22.3	25.1	5.0
8	NS-L-101136	LARA	8.9	39.8	21.7	24.0	5.6
9	NS-L-201167	ALISA	8.8	41.5	21.9	23.0	4.8
10	NS-L-201187	BEČEJKA	8.4	41.5	21.4	23.9	4.9
11	NS-L-401004	TARA	8.4	41.1	21.9	24.0	4.7
12	NS-L-401009		8.5	40.2	22.2	24.2	4.8
13	NS-L-110	BALKAN	8.8	40.3	21.9	24.1	5.0
14	NS-L-2022	RAVINICA	6.8	39.5	21.8	27.0	4.9
15	NS-L-2107	NOVOSAĐANKA	7.6	42.0	21.5	24.1	4.8
16	NS-L-110146	ANA	7.5	39.9	22.2	25.8	4.6
17	NS-L-110168		8.4	38.5	23.8	24.6	4.6
18	NS-L-110175		8.0	40.5	21.8	25.1	4.7
19	NS-L-110181	TEA	8.2	40.3	23.1	23.8	4.6
20	NS-L-110190		8.1	40.0	22.8	24.7	4.5
21	NS-L-210174	ZVEZDA	7.4	39.3	20.9	27.7	4.8
22	NS-L-210200	SAVA	8.0	40.0	21.9	25.2	4.9
23	NS-L-210201	ŠAPČANKA	7.8	39.7	21.4	26.0	5.2
24	NS-L-2024	VOJVOĐANKA	6.7	39.1	21.9	27.5	4.9
25	NS-L-220203		10.3	39.2	21.1	24.6	4.9
26	NS-L-220207		10.8	40.3	20.4	23.5	5.1
27	NS-L-120169	MIMA	7.2	37.2	24.0	26.9	4.8
28	NS-L-330219	MORAVA	7.8	38.9	23.0	25.4	4.9
29	NS-L-110120	VENERA	7.5	38.6	23.0	26.2	4.8
30	NS-L-430005		11.5	39.4	21.1	23.0	5.1
31	NS-L-115	JELICA	10.1	50.1	18.2	18.4	5.1
32	NS-L-66	KRAJINA	10.8	47.3	19.0	19.5	5.1
33	NS-L-200143	FORTUNA	11.2	49.9	17.2	19.9	5.2
34	NS-L-400007	MELI	11.6	48.2	16.9	21.1	5.3
35	NS-L-2023	AFRODITA	9.7	50.4	19.0	19.3	4.9
36	NS-L-101095	PROTEINKA	7.4	36.6	26.4	25.2	4.3
37	NS-L-201149	VALJEVKA	8.4	37.2	24.5	28.1	4.4
38	NS-L-101136	LARA	7.8	38.9	24.3	28.7	4.3

for identification of proteins from 2DE (*18*). In this study, expression of 679 proteins from developing soybean seeds were profiled, and of these, 422 proteins were identified. Two other studies of mature soybean seeds have identified 44 and 100 proteins, respectively (*19*, *20*). Comparative proteomics may be used as a tool for the selection of soybean varieties with specific nutritional properties.

A study was carried out to (1) evaluate different soybean varieties by the use of NIR and proteomics with respect to the content of protein and different carbohydrates to identify suitable alternative protein sources for monogastrics, (2) evaluate the use of NIR as a rapid screening method to predict content of complex carbohydrates of soybeans, and (3) determine whether specific seed proteins are correlated to the amount of specific NSP and oligosaccharides in soybeans.

MATERIALS AND METHODS

Selection of the Soybean Genotypes. The soybean genotypes were obtained by the Institute of Field and Vegetable Crops, Novi Sad, Serbia and Montenegro. A total of 832 genotypes of soybeans were initially screened by NIR spectroscopy in the 950−1700 nm region (DA 7000 FLEXI-MODE NIR/vis spectrophotometer, Perten Instruments, Huddinge, Sweden). The screening was performed on whole soybeans by transmittance measurements. On the basis of predictions of chemical data on protein and carbohydrate composition, 30 genotypes were chosen for analysis by high-resolution NIR spectroscopy. Samples 1−30 were grown in the experimental fields in the Novi Sad area of Vojvodina, Serbia and Montenegro (45° 16′ N and 19° 51′ E) in 2004. The dominant soil type at the field where the genotypes were grown is the calcareous chernozem with a pH of 7.65 and organic matter content of 3.3%. Samples 1−8 and 31−38 are the same genotypes, but these samples were obtained from various test plots in the Vojvodina region in 2004 (*21*). There were thus 38 soy samples in total. The predictions of chemical composition are shown in Table 1. Ground soybean samples were made by milling 2 × 100 g samples in an IKA Universalmühle (IKA Works, Wilmington, NC) for 15 s in liquid nitrogen. The predicted crude protein values of the 38 soybean samples were confirmed by chemical analysis (Kjeldahl N × 6.25) (EU Dir. 93/28).

NIR Spectroscopy of Soybeans. For the 38 samples, high-resolution reflectance spectra from 400 to 2498 nm was measured in a Foss NIRSystems model 6500 scanning monochromator (Silver Spring, MD) equipped with a transport module (2 nm steps, 32 scans per sample). Absorbance values were recorded as log 1/*R*, where *R* is the sample reflectance. The analysis of unground and ground samples was carried out using a coarse granular transport cell. This cell is rectangular with internal dimensions of 4.1-cm wide, 17.2-cm long, and 1.4-cm deep. When the ground samples were analyzed, the instrument was turned 90° (placed on a stand with the coarse cell in the horizontal direction), not to permit any of the sample to sprinkle out of the cell. Ceramic was used as a reference. Each sample was placed into the cell and scanned twice. These data formed the basis for selecting 12 particularly informative samples for detailed chemical analysis.

Carbohydrate Analysis. For the 12 selected samples, total and insoluble neutral NSP contents and their sugar compositions were determined according to the method described by Englyst et al. (*22*), by measuring the neutral sugars in acidic hydrolysates by gas chromatography (GC) as alditol acetates. The soluble fraction was calculated by the difference. Uronic acid (UA) was analyzed spectrophotometrically (*22*).

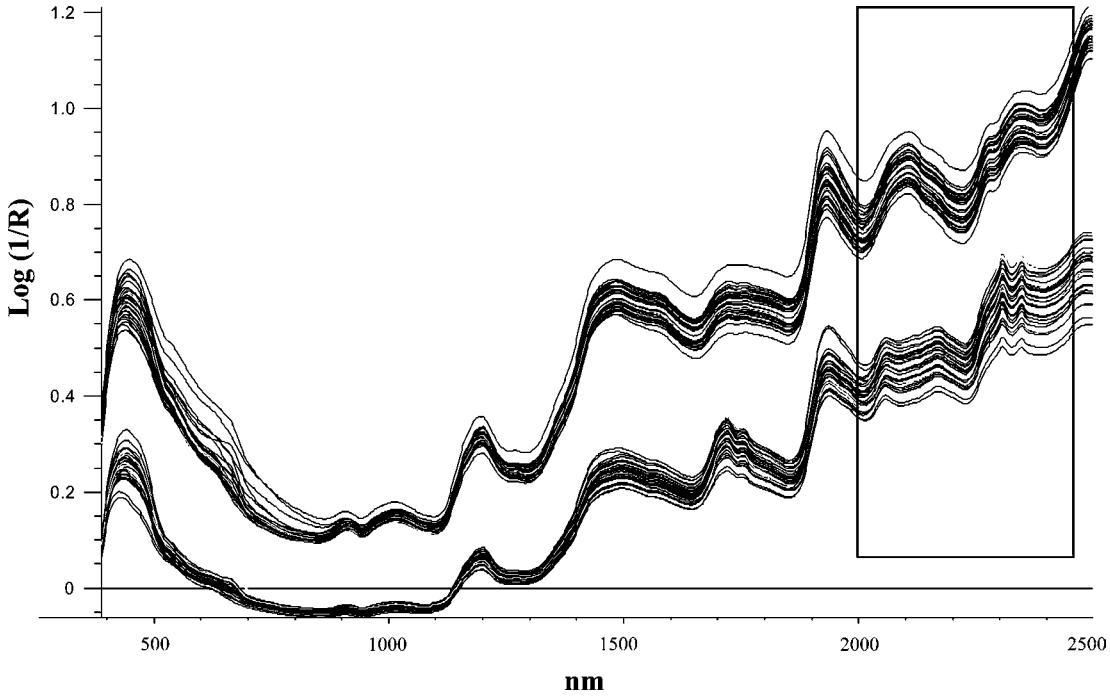

Figure 1. High-resolution NIR spectroscopy of the 38 soybean samples. Spectra were collected from 400 to 2498 nm and expressed as absorbance. The upper 38 spectra are log(1/*T*) from whole soybeans, and the lower 38 spectra are log(1/*R*) from ground soybeans. The rectangle marks the region from 2000 to 2450 nm known to contain information on the carbohydrate composition.

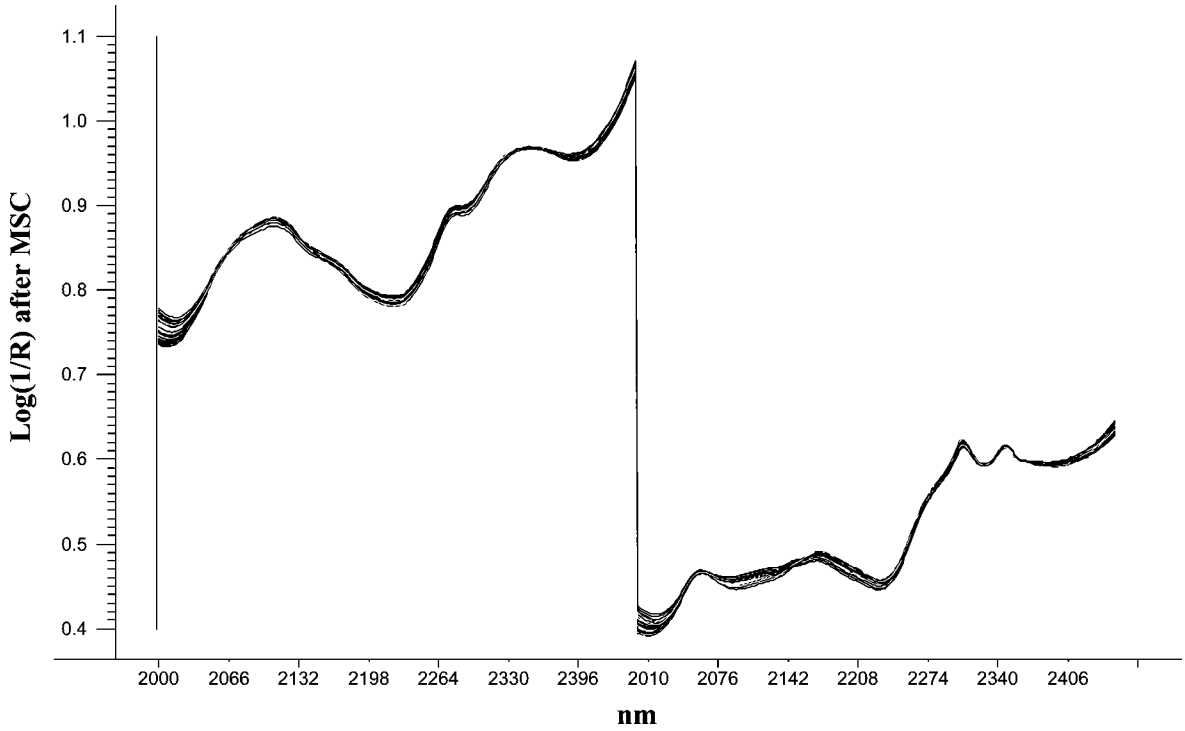

Figure 2. NIR spectra of the 38 soybean samples from the carbohydrate-relevant region 2000–2450 nm after MSC. Spectra for whole soybeans are shown to the left, and spectra from ground soybeans are shown to the right.

The concentration of low-molecular-weight sugars was quantified using an HPLC technique based on that of Copp et al. (*23*) with a few modifications. A total of 5 g of dried and milled sample was shaken with 25 mL of 50% methanol with 1.5 mg mL^{-1} mannitol as an internal standard for 30 min. Activated carbon (2.5 g) was added, and the suspension was shaken for 60 min at room temperature and left at 4 °C for 60 min. The samples were filtrated by paper filters. Filtrate was collected; 5 mL of the filtrates was incubated at 35 °C for 16 h; and precipitates were removed by centrifugation. A total of 1 mL of filtrate was evaporated by vacuum centrifugation (ISS110 SpeedVac, Termo Savant) and redissolved in 1 mL of distilled water. After filtration through Millex-HV (0.45 μm, 13 mm), the samples were analyzed by HPLC on a Shimadzu pump (LC-10AD) controlled by Class VP software. The sample (20 μL) was injected with a SIL-10 autoinjector into a Varian Carbohydrate PB-column (300 × 7.8 mm) eluted with water (0.4 mL/min) at 80 °C and equipped by a refractive detector

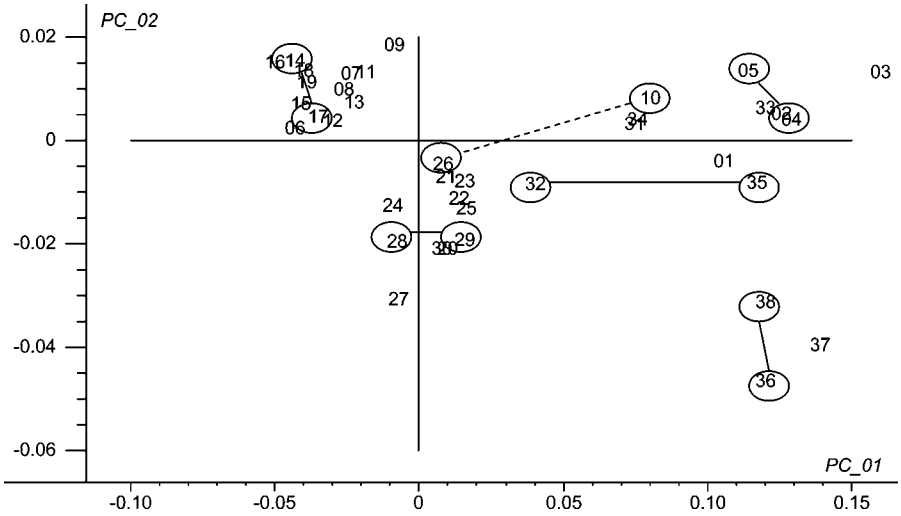

Figure 3. PLSR of NIR spectra from the 38 soybean samples. NIR spectra from both whole and ground soybeans are used as *x* variables, and samples are used as *y* variables. Predictions of protein, ash, moisture, fat, and carbohydrate contents are used as guidance in the analysis.

Table 2. Total Neutral NSP Content, Uronic Acid Content, and Content of the Individual Monosaccharides of Soybean Samples as (g kg^{-1}) Dry Matter Basis[a]

sample	rhamnose	fucose	arabinose	xylose	mannose	galaktose	glucose	total NSP	total uronic acid
4	11.3 a[b]	10.3 a	36.5 c	31.1 cd	13.2 b	60.1 bc	107.1 bcd	269.6 de	24.7 abc
5	9.1 a	4.5 a	42.8 b	43.1 a	21.7 a	72.6 abc	125.5 a	319.3 a	26.5 a
14	8.9 a	5.3 a	46.0 ab	38.3 ab	17.8 ab	74.1 ab	119.7 ab	310.1 ab	25.2 ab
17	8.9 a	5.9 a	46.9 ab	31.2 cd	17.3 ab	73.4 abc	111.5 bc	295.1 bc	26.2 ab
28	8.2 a	5.3 a	43.2 b	28.0 de	16.0 ab	78.0 a	91.4 ef	270.1 de	21.9 bc
29	8.6 a	5.6 a	43.1 b	31.2 cd	16.7 ab	76.5 a	89.5 ef	271.2 de	23.1 abc
10	9.0 a	5.7 a	45.6 ab	31.4 cd	17.4 ab	77.7 a	104.7 cd	291.5 bc	25.5 ab
26	9.3 a	6.5 a	47.4 ab	34.9 bc	16.1 ab	81.4 a	95.5 de	291.1 bc	23.5 abc
32	8.8 a	5.0 a	43.6 b	28.2 de	13.7 b	66.7 abc	90.7 ef	256.7 e	23.4 abc
35	6.9 a	4.3 a	35.3 c	23.5 e	14.4 b	57.4 c	78.7 f	220.5 f	20.4 c
36	8.2 a	5.4 a	49.8 a	34.6 bc	18.3 ab	80.2 a	88.4 ef	284.9 cd	25.0 ab
38	10.2 a	5.8 a	49.8 a	34.0 bc	16.2 ab	78.1 a	99.2 cde	293.3 bc	24.5 abc
p value	0.128	0.485	<0.001	<0.001	0.01	0.001	<0.001	<0.001	0.005

[a] Data are means of two replicates. [b] Means followed by a different letter within the same column are significantly different ($p < 0.05$).

Table 3. Insoluble Neutral NSP Content, Uronic Acid Content, and Content of the Individual Monosaccharides of Soybean Samples as (g kg^{-1}) Dry Matter Basis[a]

sample	rhamnose	fucose	arabinose	xylose	mannose	galactose	glucose	insoluble NSP	insoluble uronic acid
4	6.8 a[b]	4.0 bc	31.5 cd	32.4 abc	11.1 ab	51.3 bc	78.1 de	215.2 de	23.4 ab
5	6.9 a	4.1 abc	36.2 b	34.8 ab	13.3 a	56.6 ab	108.4 a	260.3 a	25.9 a
14	6.9 a	4.4 ab	38.1 ab	35.8 a	12.2 ab	58.7 ab	99.3 ab	255.4 ab	24.3 ab
17	6.6 a	4.8 ab	40.6 a	28.1 bcde	11.3 ab	57.4 ab	89.9 bc	238.7 bc	26.1 a
28	5.7 a	4.1 abc	34.7 bc	24.0 de	11.0 ab	59.6 ab	70.0 ef	209.1 de	24.4 ab
29	5.7 a	4.5 ab	34.9 bc	26.4 cde	11.5 ab	60.0 ab	69.3 ef	212.3 de	24.7 ab
10	6.7 a	4.7 ab	36.5 b	22.9 e	11.6 ab	60.9 ab	82.9 cd	226.2 cd	23.2 ab
26	5.4 a	4.9 a	36.4 b	25.4 cde	9.8 ab	60.3 ab	68.0 ef	210.2 ab	24.9 ab
32	6.5 a	4.0 bc	34.9 bc	21.6 e	10.7 ab	52.7 ab	71.1 ef	201.5 e	26.6 a
35	5.3 a	3.5 c	28.8 d	23.0 e	9.8 ab	42.1 c	63.4 f	175.9 f	20.6 b
36	5.9 a	4.7 ab	40.8 a	25.5 cde	9.7 ab	62.1 a	76.6 de	225.3 cd	25.3 a
38	5.4 a	4.2 abc	38.3 ab	30.2 abcd	10.1 b	55.2 ab	70.7 ef	214.1 de	25.5 a
p value	0.05	<0.001	<0.001	<0.001	0.04	<0.001	<0.001	<0.001	0.009

[a] Data are means of two replicates. [b] Means followed by a different letter within the same column are significantly different ($p < 0.05$).

(RID-6A). The sugars were quantified on the basis of their areas relative to the internal standard and corrected for their individual response factors. All samples were analyzed in triplicates.

Protein Extraction and 2DE. For the 12 samples, a total protein extract was prepared according to the protocol in Weiss et al. (*24*). All chemicals were analytical-grade from either Merck (Whitehouse Station, NJ), Sigma−Aldrich (St. Louis, MO), or BioRad Laboratories (Hercules, CA). Briefly, 50 mg of ground soybeans was extracted in 500 μL of buffer [6 M urea, 1% Triton-X-100, 0.5% DTT, and 0.5% (w/v) IPG buffer pH 3−10 (Amersham Biosciences, Piscataway, NJ)]. Protein content was measured using the RC-DC protein assay (BioRad). The total protein extract (50−500 μg) was mixed with rehydration buffer (8 M urea, 2 M thiourea, 2% Nonidet P-40, 2% IPG buffer pH 3−10, and 60 mM DTT) at a total of 350 μL and used for in-gel rehydration of homemade 18-cm immobilized pH-gradient (IPG) strips pH 4−9. Isoelectric focusing was performed using a stepwise protocol for 35 kVh using the IPGphor apparatus (Amersham Biosciences). IPG strips were equilibrated in buffer containing 50 mM Tris-HCl at pH 8.8, 6 M urea, 30% glycerol, 2% SDS, and 10 mg/mL DTT for 15 min and then for 15 min in equilibration buffer containing 25 mg/mL

Table 4. Content of Low-Molecular-Weight Carbohydrates in 12 Selected Soybean Varieties as (mg g^{-1}) Dry Matter Basis[a]

sample	stachyose	raffinose	sucrose	maltose	glucose	xylose	fructose
4	20.20 c[b]	7.68 bcde	50.31 abc	1.95 cd	1.77 cde	9.18 bc	0.29 a
5	28.60 a	6.74 cde	47.74 abcd	3.29 ab	2.91 a	9.55 b	0.43 a
14	24.09 abc	9.15 ab	48.02 abcd	1.30 d	1.59 e	8.67 bcd	0.48 a
17	21.09 bc	8.60 bc	52.80 ab	2.30 bcd	2.40 abcde	12.87 a	0.63 a
28	25.30 ab	5.92 e	40.96 def	3.32 ab	2.29 abcde	8.85 bc	0.66 a
29	27.11 a	6.68 cde	44.00 cde	3.72 a	2.78 ab	9.25 bc	0.66 a
10	24.75 abc	7.95 bcd	55.01 a	1.75 cd	1.69 de	6.48 e	0.50 a
26	28.46 a	6.86 cde	45.77 bcd	2.63 abc	2.64 abc	7.00 de	0.37 a
32	27.30 a	9.51 ab	37.57 ef	2.02 cd	1.92 bcde	7.68 cde	0.70 a
35	24.61 abc	11.05 a	33.44 f	1.82 cd	1.81 cde	8.63 bcd	0.52 a
36	26.44 a	6.41 de	48.29 abcd	3.55 a	2.27 abcde	12.16 a	0.57 a
38	24.51 abc	7.07 cde	49.79 abc	3.60 a	2.58 abcd	13.28 a	0.30 a
p value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.4824

[a] Data are means of three replicates. [b] Means followed by a different letter within the same column are significantly different ($p < 0.05$).

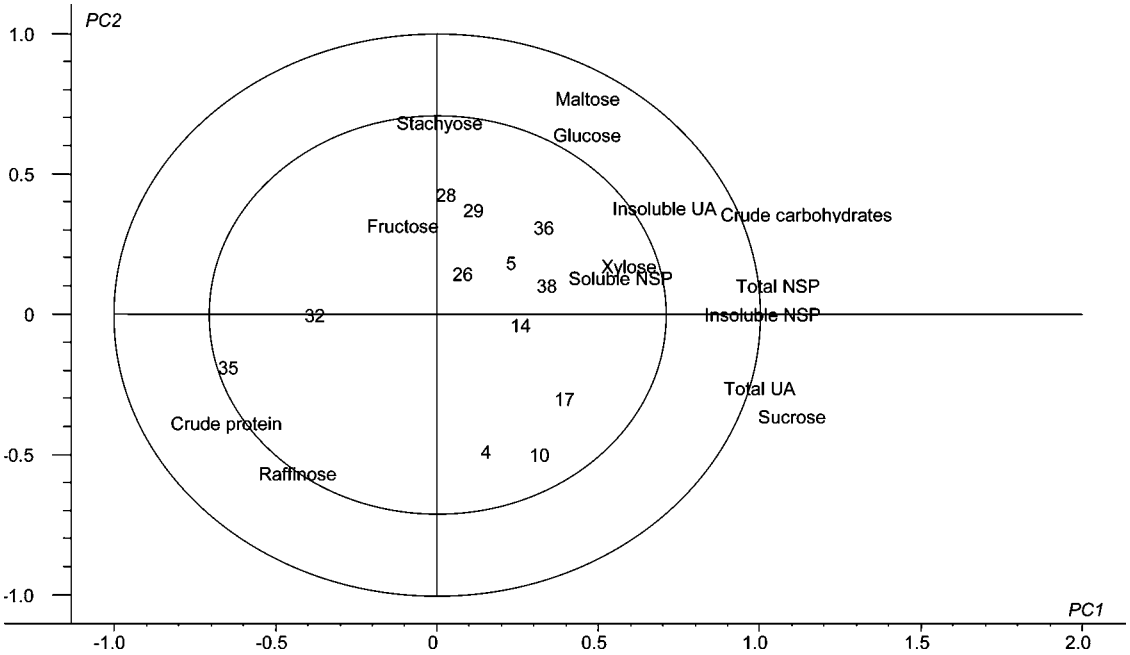

Figure 4. Loading plot of the two first dimensions of the PLSR of carbohydrate composition of the 12 selected soybean samples. Chromatographically determined carbohydrates were used as *x* variables, and sample identifiers were used as *y* variables. Predictions of crude protein and carbohydrates are included as passive variables in the figure.

iodoacetamide instead of DTT. After equilibration, proteins were separated on vertical 12.5% polyacrylamide gels on the Ettan DALT *twelve* apparatus (Amersham Biosciences). Analytical gels were silver-stained according to the protocol of Blum et al. (*25*). For preparative gels, 500 μg of protein was used in the IEF and the gels were silver-stained, but glycine was replaced by 5% acetic acid after development. Gel alignments were performed with the Z3 3.0 program package from Compugen, Inc. (San Jose, CA).

Protein Identification by Mass Spectrometry. Gel plugs containing the protein spot of interest were washed twice for 15 min in 50 mM ammonium carbonate at pH 8.0−8.5/acetonitril (1:1) and dried in a speed-vac centrifuge for 20 min. A total of 30 μL of 50 mM ammonium carbonate at pH 8.0−8.5 containing 0.15 μg of trypsin (Sequence grade, Promega, Madison, WI) was added to each gel plug and incubated for 45 min on ice before transfer to 37 °C and further incubation overnight. The supernatant was removed, added 3 μL of formic acid, and diluted to a final concentration of 0.25% TFA. The sample was enriched and purified using OMIX microcolumns (Varian, Palo Alto, CA) according to instructions from the manufacturer. The peptide samples were eluted in α-cyano-4-hydroxy-cinnamic acid, 0.1% TFA/50% ACN (1:1), and spotted directly onto the target plate. Peptide samples were analyzed by mass spectrometry (MS) using matrix-assisted laser desorption ionization time-of-flight (MALDI−TOF) in an Ultraflex MALDITOF/TOF (Bruker Daltronics, Billerica, MA). A peptide standard (Bruker Daltronics) was used for external calibration of the instrument, and internal calibration was performed using the trypsin autolysis peaks. Proteins were identified by peptide mass fingerprinting, searching the NCBI database using the MASCOT search engine at http://www.matrixscience.com, and 100 ppm tolerance. All identified proteins were confirmed by MS/MS using the LIFT module on 1−2 selected parent ions and repeated database searches.

Statistics. The data were analyzed by ANOVA using the GLM procedure of SAS (SAS version 6.08, SAS Institute, Inc., Cary, NC). Tukey test ($p < 0.05$) was performed to determine minimum significant differences (MSD). Principal component analysis (PCA) and the multivariate analysis partial least-squares regression (PLSR) were carried out using The Unscrambler version 9.05 (CAMO ASA, Trondheim, Norway). Because the outlined region from 2000 to 2450 nm of the NIR spectra is known to be particularly informative about carbohydrate composition (*26*), this region was used for the multiplicative signal correction (MSC).

RESULTS

High-Resolution NIR Spectroscopy of Soybeans. To investigate how to get maximum information about soybean samples at minimum laboratory work, two different ways of using NIR was tested. High-resolution NIR absorbance spectra

from 400 to 2500 nm of both whole and ground soybeans are shown in **Figure 1**. Clear absorbances are evident in the visible wavelength range. Characteristic water absorbance peaks are visible around 1200, 1450, and 1940 nm, and around 2100 nm the characteristic carbohydrate peak is visible, especially in the upper, whole soybean curves. In contrast, the ground soybean spectra show small but relatively sharp lipid peaks around 1720 and 2350 nm. (The more complex peak patterns of the soybean proteins, known to contribute to the observed spectra, are more difficult to identify by visual inspection alone.) Thus, the two ways of measuring the soybeans gave somewhat different NIR information, judging already from the visual appearance of the curves. However, because of the extreme precision of the NIR measurements, detailed multivariate mathematical feature extraction is normally more informative than just visual inspection. The obvious cross baseline and scaling variations in the spectra represent light-scattering variations, which may be caused by differences in the physical structure in the samples and thus to differences in the behavior upon grinding. To separate the physical information from the chemical information of the samples, the spectra were corrected using MSC. **Figure 2** shows the effect of spectral preprocessing by MSC in the 2000−2450 nm region of the NIR spectra, intended to separate the two main types of variation in the spectra in **Figure 1**, physical light scattering and chemical light absorption. The MSC was performed on NIR spectra in the 2000−2400 nm region for both the whole and ground soybean samples. **Figure 2** shows differences in the spectra from whole and ground soybean samples. Although most of the light-scattering variations are removed, differences can still be observed both in general baseline and scaling levels as well as in specific chemical absorption peaks. To select a subset of the soybean samples for further analyses, a PLSR model $y \approx f(x)$ was generated from the information in the MSC-treated spectra (x variables). Predictions of protein, ash, and carbohydrate contents based on initial NIR spectroscopic measures were used as y variables for modeling guidance in the selection model. This multivariate calibration model demonstrated that the soybean samples had different chemical composition. PLSR is a bilinear regression modeling method, where a few latent variables ("PLS components, PCs") $T = [\, t_1, t_2, t_3, ...]$ are generated as y-relevant linear combinations of the x variables, $T \approx f_1(x)$, and y is then modeled by regression on these latent variables, $y \approx f_2(T)$. **Figure 3** represents the bilinear "score plots" of t_1 versus t_2, showing the position of each of the 38 soybean samples with respect to the two most important NIR variation patterns. This plot and the corresponding score plot of t_1 versus t_3 was used for selecting 12 soybean samples for further investigation. These 12 samples were chosen to form 6 different pairs (connected by line segments in **Figure 3**). Samples 5 and 35 are the same genotype but grown at different locations, even though they appear in separate groups based on the NIR spectroscopic analysis and measured chemical content. The pair containing samples 10 and 26 was chosen based on the originally predicted contents of carbohydrates and protein. The predicted protein content was highly correlated to the analyzed content ($r = 0.96$).

Composition of Carbohydrates. The content of different nonstarch carbohydrates was determined chromatographically in the selected 12 samples. **Tables 2** and **3** show the content of total and insoluble NSPs and UA, respectively. There were significant differences among the soybean samples for the content of all measured total monosaccharides, except for rhamnose and fucose, and for the content of all measured insoluble monosaccharides, except for rhamnose. The UA

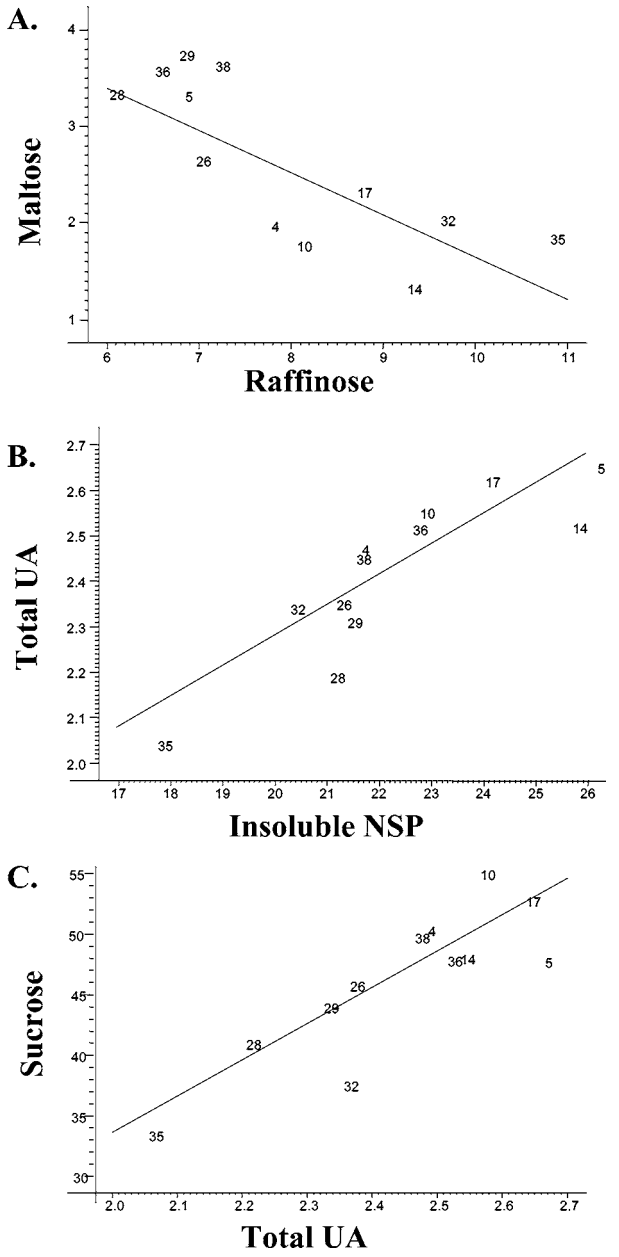

Figure 5. Pairwise raw-data plots of selected carbohydrates for verification of the PLSR results. (A) Correlation of maltose and raffinose content. (B) Correlation of total UA and insoluble NSP. (C) Correlation of sucrose and total UA.

content varied significantly among the samples, and most of the UA was found in the insoluble NSP fraction. The oligosaccharide composition of the 12 different soybean samples is presented in **Table 4**. Among the raffinose group of oligosaccharides, stachyose was predominant in these soybean samples. With the exception of fructose, there were significant differences among the 12 soybean samples for the content of low-molecular-weight carbohydrates. A bilinear PLSR model was developed, using carbohydrates as x variables and sample indicator variables (0/1) as noninformative y variables. Additional relevant variables are included for interpretation: crude protein and carbohydrate contents predicted from the initial NIR analyses are included as passified (down-weighted and thus noninfluential) y variables. The variation in the carbohydrate content was mainly explained by the first two components. This model is summarized as the scale-free correlation loading plot for the first two components in **Figure 4**, where the axes represent the correlation coefficient

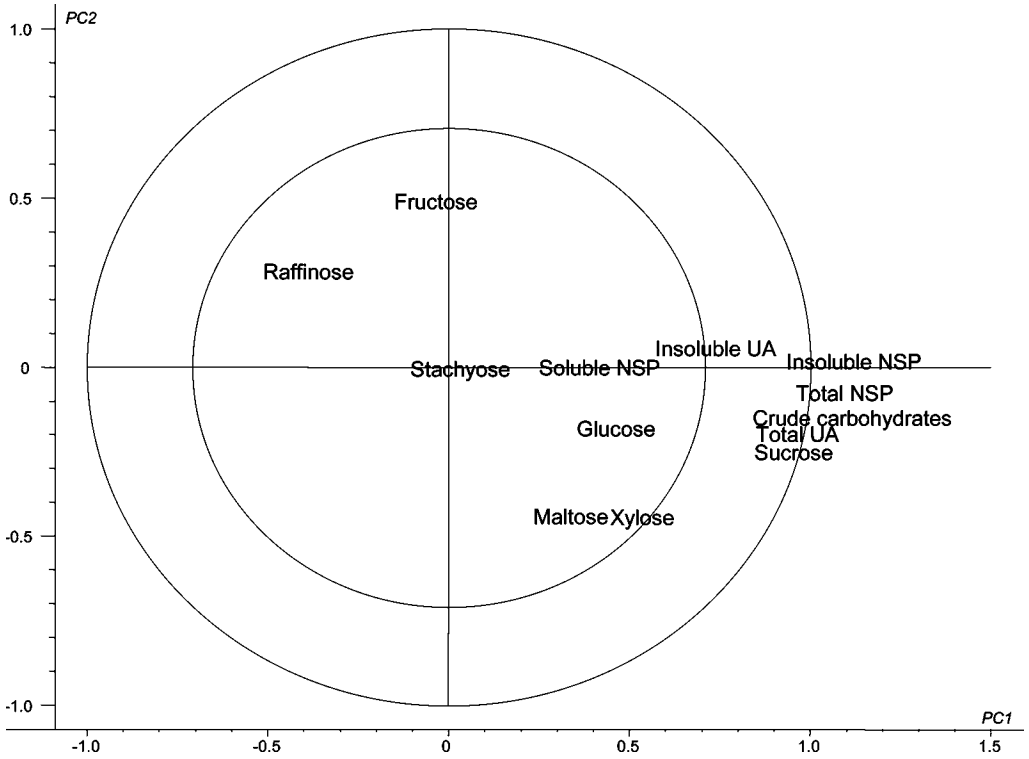

Figure 6. Correlation loading plot of carbohydrates modeled by NIR in the selected 12 soybean samples. NIR spectra of ground soybeans are used as *x* variables, and compositions of carbohydrates are used as *y* variables.

between each of the input *x* and *y* variables and PLSR score t_1 (abscissa) and t_1 (ordinate), respectively. Positions are given one character to the left of the variable names. The figure shows that within the obtained model, total UA, total NSP, insoluble NSP, sucrose, and crude carbohydrates are strongly positively intercorrelated and negatively intercorrelated to the crude protein measurement, which are particularly high in soybean sample 35. Maltose and raffinose are likewise negatively intercorrelated. Stacchyose shows a strong, unique variation pattern and is particularly high in sample 29 and low in sample 4. The most salient features in these raw data are plotted in more detail in **Figure 5**, which shows the best correlation among individual carbohydrates**.** As seen in **Figure 4**, **Figure 5** shows that there is a negative correlation between raffinose and maltose ($r = -0.77$) and a positive correlation between total UA and both insoluble NSP ($r = 0.87$) and sucrose ($r = 0.85$).

NIR as a Predictor of Carbohydrate Composition. Using the MSC-treated NIR spectra as *x* variables and the carbohydrate measurements as *y* variables, a PLSR model was made. Cross-validation between the 12 samples indicated 2 PLS PCs to give the best predictive ability for all of the carbohydrate variables, explaining on the average 42% of the 13 *y* variables. **Figure 6** shows the correlation loading plot for the *y* variables for the first two PCs. Because different variation patterns might affect the different carbohydrate variables, individual PLSR calibration models were made for each individual carbohydrate variable (*y*) versus the MSC-treated spectra (*x*). Some of the best prediction abilities of individual carbohydrates from NIR spectra are shown in **Figure 7**. Full cross-validation was used for finding the optimal number of PCs in each case. There was a positive correlation between the NIR spectra and total NSP ($r = 0.91$), insoluble NSP ($r = 0.80$), total UA ($r = 0.86$), and raffinose ($r = 0.75$).

Changes in Protein Expression Related to Carbohydrate Composition. The selected 12 samples were analyzed by 2DE (**Figure 8**). Approximately 1000 proteins were resolved on the gels, and 590 protein spots were matched in the data set. Several proteins appeared to vary substantially in intensities among the different samples, although the cross-validation did not reveal any significant PLS PCs, probably because the number of samples was too low compared to the heterogeneity of the sample set. The protein marked by an arrow in the figure was identified as the glycinin G1 precursor protein from soybean (Swissprot ID P04776) with 30% sequence coverage. The observed and theoretical pI and MW are in agreement (pI 5.9; MW, 6.3 kD), and the gel position also matches the position in the reference maps described in Hajduch et al. (*18*). A PLSR model (**Figure 9**) of the spot intensities of the glycinin G1 precursor demonstrates that the intensity of the protein correlates to some of the carbohydrates. There was a negative correlation between the glycinine G1 precursor and both insoluble NSP ($r = -0.70$) and total UA ($r = -0.80$) (**Figure 10**).

DISCUSSION

NIR spectroscopy was used in combination with traditional chemical analyses to predict protein and carbohydrate composition in soybeans. The results of the present study suggest that NIR spectroscopy is a promising tool to predict the content of complex carbohydrates such as total and insoluble NSP, total and insoluble UA, and some low-molecular-weight carbohydrates. NIR spectroscopy might thus be used as a basis to select soybeans with desirable carbohydrate composition for defined nutritional purposes. Changes in protein expression were found to be associated with certain carbohydrate components in soybeans, suggesting that protein composition is related to the composition of carbohydrates.

A variation in the content of NSP and total UA among the 38 soybean samples was observed. Glucose comprised the largest proportion of the total and insoluble NSP, followed by galactose, arabinose, xylose, mannose, rhamnose, and xylose. These results are similar to results presented earlier (*27*). Sucrose

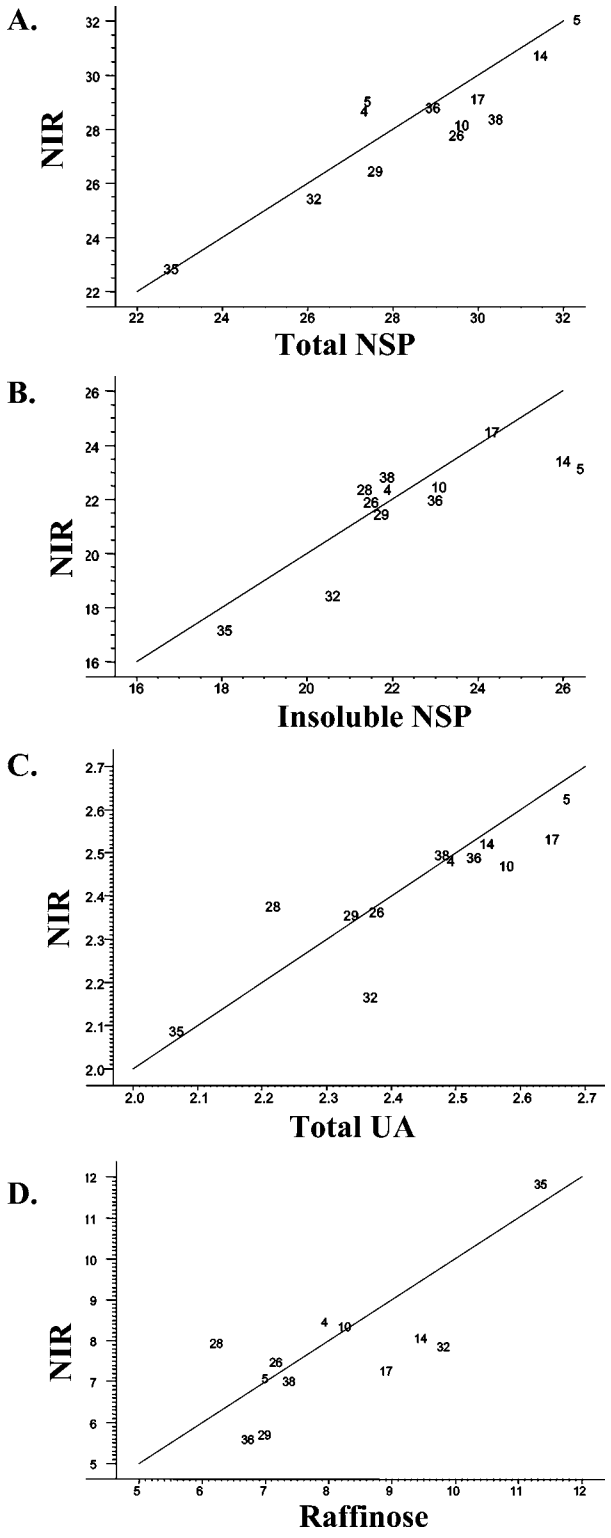

Figure 7. Correlation plots between some NIR-predicted and chromatographically measured carbohydrates in the selected 12 soybean samples. (A) Prediction of total NSP. (B) Prediction of insoluble NSP. (C) Prediction of total uronic acid. (D) Prediction of raffinose.

comprised the largest proportion of the oligosaccharides, followed by raffinose, stachyose, xylose, maltose, and glucose. This is in agreement with earlier studies reporting that the main oligosaccharides in soybean meal are the α-galactosides sucrose (6−7%), raffinose (1−2%), and stachyose (5−6%), accounting for a total soluble carbohydrate content of 12−15% (*1*, *28*).

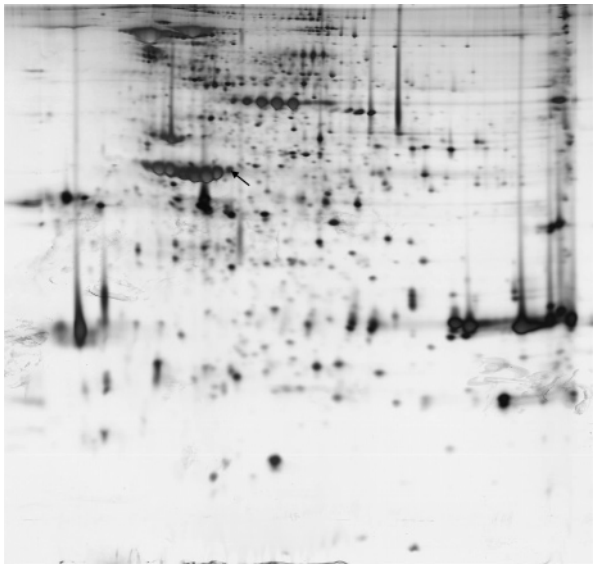

Figure 8. Representative 2DE image of soybean. A total protein extract of soybean was separated on IPG at pH 4–9 in the first dimension and 10% SDS–PAGE in the second dimension. Proteins are silver-stained, and the spot identified as the glycinin G1 precursor is marked by an arrow.

Although soybean meal is a major protein supplement for monogastric animals throughout the world, the oligosaccharides and NSP from soybean have been shown to cause digestive disorders and reduced performance in several species (*3*, *7*, *14*). A negative effect of soybean NSP on the digestibility of fat and protein has been reported in broiler chickens and salmon (*7*). Oligosaccharides are not digested by enzymatic hydrolyses in monogastrics because of the lack of the enzyme α-galactosidase, which is necessary to hydrolyze the α-1,6 linkages present in oligosaccharides (*29*), but are fermented by the bacterial population in the gastrointestinal tract. Fermentation of oligosaccharides in the gastrointestinal tract has been associated with adverse effects on nutrient digestibility and energy availability of soybean meal in poultry (*30*, *31*). In pigs, the effect of oligosaccharides on nutrient digestibility is not as clear (*6*, *32*), but reduced nutrient digestibility (*4*) and increased diarrhea (*6*) have been reported. In fish, indigestible soybean oligosaccharides have been reported to cause osmotic diarrhea and alterations in the intestinal flora (*33*).

Despite the limited number of soybean samples analyzed for complex carbohydrate composition, it was possible to obtain a reasonably good model to predict total NSP, unsoluble NSP, raffinose, and total UA content from the high-resolution NIR spectra of the soybean samples. For most low-molecular-weight carbohydrates, it was not possible to make a prediction model based on the NIR spectra because of the large variation in the content of these carbohydrates in the samples. NIR calibration normally requires many more calibration samples to optimize and assess the statistical calibration model. In particular, the cross-validation was negatively impacted by sample 35, which was rather unique, both in the NIR spectrum and in the carbohydrate composition, compared to sample 32, which was within the same group as sample 35 in the original selection of the 12 samples. Hence, improved results for the NIR-based prediction of the patterns of covariation among the carbohydrate fractions may be attainable, but this needs further validation and refinement based on more data.

Composition of soybean carbohydrates may be influenced by genotype and environmental conditions, including soil type,

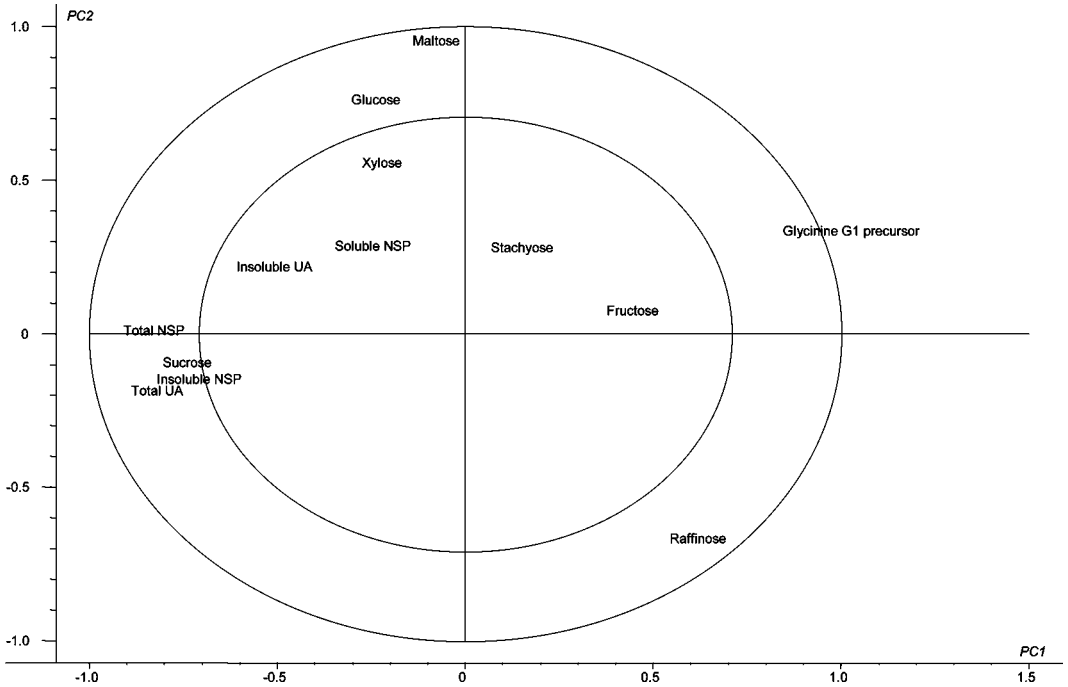

Figure 9. Loading plot of the two first dimensions of the PLSR of carbohydrate composition and glycinin G1 precursor content in the selected 12 soybean samples. Carbohydrate compositions are used as *x* variables, and contents of the glycinin G1 precursor are used as *y* variables.

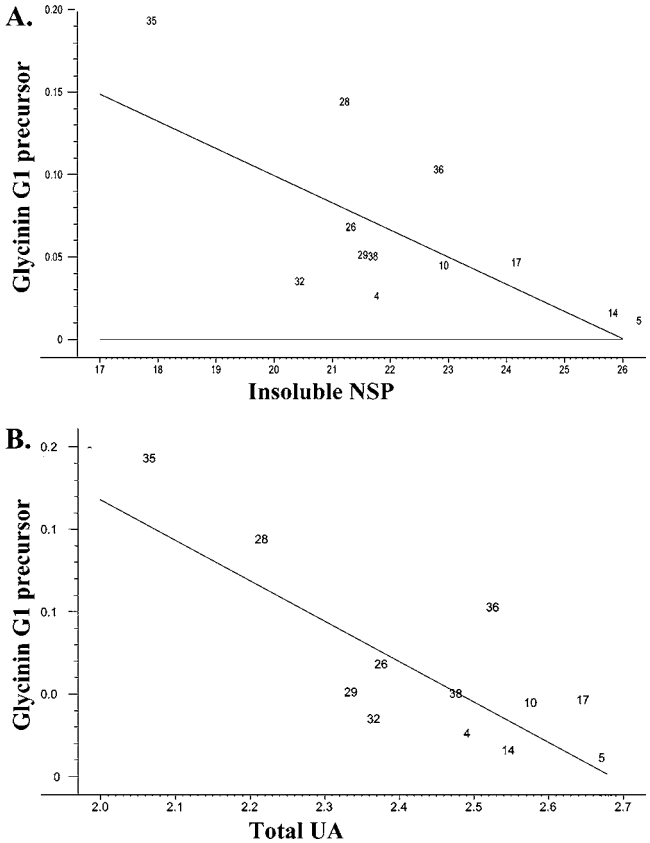

Figure 10. Pairwise raw-data plots of the glycinin G1 precursor content and selected carbohydrates for verification of the PLSR results. (A) Correlation of glycinin G1 content and insoluble NSP. (B) Correlation of glycinin G1 content and total UA.

fertilizer application, and climate (*13*, *21*). In the present study, large variation in the chemical content among the soybean samples was observed among and within genotypes. For instance, soybean samples 5 and 35 were from the same genotype but grown under different environmental conditions. The protein content in these two samples differed by 10%, with the total protein and relative amount of glycinin G1 precursor being higher in sample 35. In addition, the total NSP differed by 10%, insoluble NSP differed by 8.4%, and the content of raffinose was twice as high in sample 35 as compared with sample 5. This indicates that the environmental conditions were more important for the variation in the content of protein and NSP than the effect of the genotype. However, because the soybeans are grown in the field, it cannot be excluded that cross-fertilization influenced the chemical composition.

In this study, we observed a negative correlation between the seed storage protein G1 glycinin precursor and insoluble NSP and raffinose. Glycinin and β-conglycinin have shown to produce immunological reactions in young farm animals. β-conglycinin appears to be more immunoreactice than glycinin in young calves (*9*). To our knowledge, a correlation between carbohydrate composition and seed storage proteins has not been previously reported. Proteomics could be a powerful tool to reveal more information on the proteome in soybeans related to the composition of carbohydrates. Recently, an oilseed proteomics initiative has published proteome maps with expression profiles of 679 protein spots during seed filling in soybeans (*18*). These maps may be used as reference tools to identify soybean seed proteins.

We have demonstrated that high-resolution NIR has a potential to identify heat-stable antinutritional factors, such as NSP and oligosaccharides in soybeans. This can aid in the selection of suitable soybeans for use in diets for target monogastric species.

ACKNOWLEDGMENT

We thank Anne Helene Bjerke for skillful technical assistance. The equipment at the MS/Proteomics platform in the Food Science Alliance, Ås, Norway, was used for identification of proteins.

LITERATURE CITED

(1) Francis, G.; Makkar, H. P. S.; Becker, K. Antinutritional factors present in plant-derived alternate fish feed ingredients and their effects in fish. *Aquaculture* **2001**, *199*, 197−227.

(2) Anderson, R. L.; Wolf, W. J. Compositional changes in trypsin inhibitors, phytic acid, saponins, and isoflavones related to soybean processing. *J. Nutr.* **1995**, *125*, S581−S588.

(3) Smits, C. H. M.; Annison, G. Non-starch plant polysaccharides in broiler nutrition—Towards a physiologically valid approach to their determination. *Worlds Poult. Sci. J.* **1996**, *52*, 203−221.

(4) Kornegay, E. T. Feeding value and digestibility of soybean hulls for swine. *J. Anim. Sci.* **1978**, *47*, 1272−1280.

(5) Skrede, A.; Krogdahl, A. Heat affects nutritional characteristics of soybean meal and excretion of proteinases in mink and chicks. *Nutr. Rep. Int.* **1985**, *32*, 479−489.

(6) Zhang, L. Y.; Li, D. F.; Qiao, S. Y.; Johnson, E. W.; Li, B. Y.; Thacker, P. A.; Han, I. K. Effects of stachyose on performance, diarrhoea incidence, and intestinal bacteria in weanling pigs. *Arch. Anim. Nutr.-Archiv. Fur Tierernahrung* **2003**, *57*, 1−10.

(7) Refstie, S.; Svihus, B.; Shearer, K. D.; Storebakken, T. Nutrient digestibility in Atlantic salmon and broiler chickens related to viscosity and non-starch polysaccharide content in different soyabean products. *Anim. Feed Sci. Technol.* **1999**, *79*, 331−345.

(8) Krogdahl, A.; Bakke-McKellep, A. M.; Baeverfjord, G. Effects of graded levels of standard soybean meal on intestinal structure, mucosal enzyme activities, and pancreatic response in Atlantic salmon (*Salmo salar* L.). *Aquacult. Nutr.* **2003**, *9*, 361−371.

(9) Lallès *Soy Products as Protein Sources for Preruminants and Young Pigs*; Drackley, J. K., Ed.; Federation of Animal Science Societies: Savoy, IL, 2000; pp 106−126.

(10) Leske, K. L.; Jevne, C. J.; Coon, C. N. Effect of oligosaccharide additions on nitrogen-corrected true metabolizable energy of soy protein concentrate. *Poult. Sci.* **1993**, *72*, 664−668.

(11) Snyder, H. E.; Kwon, T. W. *Soybean Utilization*; Van Nostrand Reinhold: New York, 1987; p 60.

(12) Knudsen, K. E. B. Carbohydrate and lignin contents of plant materials used in animal feeding. *Anim. Feed Sci. Technol.* **1997**, *67*, 319−338.

(13) Karr-Lilienthal, L. K.; Kadzere, C. T.; Grieshop, C. M.; Fahey, G. C., Jr. Chemical and nutritional properties of soybean carbohydrates as related to nonruminants: A review. *Livestock Prod. Sci.* **2005**, in press.

(14) Coon, C. N.; Leske, K. L.; Akavanichan, O.; Cheng, T. K. Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. *Poult. Sci.* **1990**, *69*, 787−793.

(15) Perez-Marin, D. C.; Garrido-Varo, A.; Guerrero-Ginel, J. E.; Gomez-Cabrera, A. Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: Chemical composition and open-declaration. *Anim. Feed Sci. Technol.* **2004**, *116*, 333−349.

(16) Hunter, T. C.; Andon, N. L.; Koller, A.; Yates, J. R.; Haynes, P. A. The functional proteomics toolbox: Methods and applications. *J. Chromatogr., B: Biomed. Sci. Appl.* **2002**, *782*, 165−181.

(17) Tyers, M.; Mann, M. From genomics to proteomics. *Nature* **2003**, *422*, 193−197.

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

(20) Herman, E. M.; Helm, R. M.; Jung, R.; Kinney, A. J. Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol.* **2003**, *132*, 36−43.

(21) Vidic, M.; Hrustic, M.; Jockovic, D.; Miladinovic, J.; Dordevic, V. *38th Seminar of Agronomists*; Institute of Field and Vegetable Crops: Novi Sad, Serbia and Montenegro, pp 129−139.

(22) Englyst, H. N.; Quigley, M. E.; Hudson, G. J.; Cummings, J. H. Determination of dietary fibre as non-starch polysaccharides by gas−liquid chromatography. *Analyst* **1992**, *117*, 1707−1714.

(23) Copp, L. J.; Blekinsop, R. W.; Yada, R. Y.; Marangoni, A. G. The relationship between respiration and chi color during long-term storage of potato tubers. *Am. J. Potato Res.* **2000**, *77*, 279−287.

(24) Weiss, W.; Vogelmeier, C.; Gorg, A. Electrophoretic characterization of wheat grain allergens from different cultivars involved in bakers' asthma. *Electrophoresis* **1993**, *14*, 805−816.

(25) Blum, H.; Beier, H.; Gross, H. J. Improved silver staining of plant proteins, RNA, and DNA in polyacrylamide gels. *Electrophoresis* **1987**, *8*, 93−99.

(26) Munck, L.; Pram Nielsen, J.; Møller, B.; Jacobsen, S.; Søndergaard, I.; Engelsen, S. B.; Nørgaard, L.; Bro, R. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. *Anal. Chim. Acta* **2001**, *446*, 171−186.

(27) Refstie, S.; Sahlström, S.; Bråthen, E.; Baeverfjord, G.; Krogedal, P. Lactic acid fermentation eliminates indigestible carbohydrates and antinutritional factors in soybean meal for Atlantic salmon (*Salmo salar*). *Aquaculture* **2005**, *246*, 331−345.

(28) Honig, D. H.; Rackis, J. J. Determination of the total pepsin−pancreatin indigestible content (dietary fiber) of soybean products, wheat bran, and corn bran. *J. Agric. Food Chem.* **1979**, *27*, 1262−1266.

(29) Slominski, B. A. Hydrolysis of galactooligosaccharides by commercial preparations of α-galactosidase and β-fructofuranosidase—Potential for use as dietary additives. *J. Sci. Food Agric.* **1994**, *65*, 323−330.

(30) Parsons, C. M.; Zhang, Y.; Araba, M. Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poult. Sci.* **2000**, *79*, 1127−1131.

(31) Smiricky-Tjardes, M. R.; Flickinger, E. A.; Grieshop, C. M.; Bauer, L. L.; Murphy, M. R.; Fahey, G. C. *In vitro* fermentation characteristics of selected oligosaccharides by swine fecal microflora. *J. Anim. Sci.* **2003**, *81*, 2505−2514.

(32) Smiricky, M. R.; Grieshop, C. M.; Albin, D. M.; Wubben, J. E.; Gabert, V. M.; Fahey, G. C. The influence of soy oligosaccharides on apparent and true ileal amino acid digestibilities and fecal consistency in growing pigs. *J. Anim. Sci.* **2002**, *80*, 2433−2441.

(33) Wiggins, H. S. Nutritional value of sugars and related compounds undigested in the small gut. *Proc. Nutr. Soc.* **1984**, *43*, 69−75.

Received for review June 17, 2005. Revised manuscript received September 8, 2005. Accepted September 15, 2005.

JF051438R