

# Predicting Protein Composition, Biochemical Properties, and Dough-Handling Properties of Hard Red Winter Wheat Flour by Near-Infrared Reflectance

Stephen R. Delwiche,<sup>1,2</sup> Robert A. Graybosch,<sup>3</sup> and C. James Peterson<sup>3</sup>

## ABSTRACT

Cereal Chem. 75(4):412-416

Breadmaking quality in wheat is one of several considerations that plant breeders face when developing new cultivars. In routine breeding programs, quality is assessed by small-scale dough-handling and bake tests, and to some extent, by biochemical analysis of gluten proteins. An alternative, not yet fully examined, method for wheat flour quality assessment is near-infrared reflectance (NIR) spectrophotometry. The present study was performed on 30 genotypes of hard red winter wheat grown during two crop years at eight to nine locations in the Great Plains area of the United States. Biochemical testing consisted of measuring protein fractions from size-exclusion HPLC ( $M_r > 100k$ ,  $M_r 25-100k$ , and  $M_r < 25k$  designated as glutenin, gliadins, and albumin and globulins, respectively), pentosan content, and SDS sedimentation volume. Dough-handling properties were

measured on a mixograph and recorded as the time to peak dough development, the peak resistance, the width of the mixing curve, and the width of the curve at 2 min past peak. Partial least squares analyses on diffuse NIR spectra (1,100–2,498 nm) were developed for each constituent or property. When applied to a separate validation set, NIR models for glutenin content, gliadin content, SDS sedimentation volume, and mixograph peak resistance demonstrated reference vs. predicted correlations ranging from  $r = 0.87$  to  $r = 0.94$ . Such models were considered sufficiently accurate for screening purposes in breeding programs. NIR spectra were responsive to each constituent or property at a level higher than expected from a correlation between the constituent or property and protein content (recognizing that protein content is modeled by NIR with high accuracy).

End-use quality in bread wheats can be defined by several characteristics, including milling yield, water absorption, dough mixing time and tolerance, loaf volume, and crumb grain appearance. Such quality is strongly determined by the composition of endosperm storage proteins, glutenin and gliadins, an axiom that is well documented in the literature (Orth and Bushuk 1972, Bietz 1987, MacRitchie et al 1990). In the present study, we were particularly concerned with the prebake quality procedures that are, or have potential to be, useful in wheat breeding programs. These procedures include size-exclusion HPLC (SE-HPLC) for protein composition, SDS sedimentation test for protein quality, and mixograph test for dough-handling properties.

SE-HPLC has been advocated as a useful tool for wheat quality improvement in breeding programs (Singh et al 1990a, Gupta et al 1993). Singh et al (1990b) claimed that sonication of the wheat flour in 2% SDS solution, in conjunction with SE-HPLC, resulted in a fractionation procedure that was more consistent in extraction efficiency across wheat genotypes than earlier methods involving differential solubility or size-exclusion chromatography alone. Furthermore, they were able to measure a strong relationship between the relative quantity of glutenin and some indicators of flour strength (loaf volume, extensigraph dough resistance and extensibility, and mixograph peak development time). By similar methodology, Graybosch et al (1996) concluded that the sonication and SE-HPLC procedure could be used to identify samples with acceptable loaf exterior and grain, primarily through a positive correlation between glutenin, expressed as a fraction of extractable protein, and these baked product attributes.

Among various biochemical tests for protein quality, the SDS sedimentation test is probably the most widely used due to its relative simplicity. The flour particle reacts when immersed in SDS solution; gluten quality is gauged by the swelling of glutenin and dissolution of gliadins (Eckert et al 1993). The test is routinely used in

breeding programs (Payne et al 1979). Along with protein content, SDS sedimentation volume in hard red winter (HRW) wheat was positively correlated with mixograph tolerance, loaf volume, and loaf exterior appearance (Graybosch et al 1996).

Despite the effectiveness of SE-HPLC, SDS sedimentation, and mixograph procedures in assessing wheat quality, analysis times are typically greater than 10 min per sample. In the first two procedures, the tests are costly in terms of chemicals and their disposal. Consequently, wheat breeders and, perhaps processors of wheat flour, desire a faster and less complex means to measure flour quality. Near-infrared reflectance (NIR) spectrophotometry for determination of protein and moisture in cereals has been commercially available since the 1970s (Norris 1978). Previously, a study on commercial bread quality by NIR spectrophotometry demonstrated feasibility for modeling water absorption and loaf height, and difficulty for modeling dough development time, tolerance, and baked product appearance (Delwiche and Weaver 1994). Higher homogeneity in quality of the commercial stock, compared to that expected in wheat breeding programs, may have contributed to the poor performance of dough and breadmaking NIR models. The purpose of the research described here was to determine the feasibility of using NIR spectrophotometry to measure the biochemical components or physical properties in wheat flour that are the determinants of breadmaking quality. The specific objectives were to assess the ability of NIR to measure concentrations of glutenin, gliadins, albumins and globulins, SDS sedimentation volume, and mixograph dough-handling properties, based upon a genetically and environmentally diverse population of HRW wheat samples.

## MATERIALS AND METHODS

### Samples

The HRW wheat samples were grown in eight locations throughout Nebraska in 1990 and 1991 (Alliance, Clay Center, Grant, Lincoln, McCook, North Platte, Scottsbluff, and Sidney). The 1991 crop from Alliance was not harvested because of winter-kill. Dryland farming operations were used at all locations; however, at Sidney an additional planting was made in a field subjected to irrigation. Thirty genotypes representing former and then-current common commercial varieties of the Great Plains and promising experimental lines were grown at each location. Twenty-two genotypes did not contain a wheat-rye chromosomal translocation (Abilene, Arapahoe, Arkan, Centura, Centurk78, Cheyenne, Chisholm, Cimarron, Eagle, Karl, Lamar, Lancota, Newton,

<sup>1</sup> USDA-ARS, Instrumentation and Sensing Laboratory, Building 303, BARC-East, Beltsville, MD 20705-2350. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable.

<sup>2</sup> Corresponding author. E-mail: sdelwiche@asrr.arsusda.gov

<sup>3</sup> USDA-ARS, University of Nebraska, Department of Agronomy, Lincoln, NE.

N86L022, N86L177, NE83407, Plainsman V, RedChief, Redland, Scout 66, TAM 101, and Vona), five contained the 1AL.1RS translocation originally derived from Amigo (TAM 107, TAM 200, TAM 202, Century, and TXGH12588), and three contained the 1BL.1RS translocation (Siouxland, KS8010-72, and Rawhide) derived from the Russian cultivar Kavkaz. At each location (with two locations each year for Sidney), samples were planted in a randomized complete block design consisting of five blocks. Samples in this study originated from blocks 1 and 3 only. Additional details on growth of samples are provided in Graybosch et al (1996).

### Procedure

Grain from each sample (50 g) was tempered to 15% moisture (wb), then milled on a Brabender Quadraplex experimental mill (South Hackensack, NJ). Flour was sifted through a U.S. Standard Sieve No. 70 on a shaker (Strand Shaker Co., Minneapolis, MN) at 225 rpm for 90 sec. Samples were stored at 0–5°C in resealable plastic bags 0.1 mm thick.

Two portions (11 mg) from each sample were used in the determination of protein composition. Protein separation of the first portion was performed by SE-HPLC as described in Graybosch et al (1993), based on a procedure developed by Singh et al (1990a,b). Three fractions were obtained by SE-HPLC of the supernatant from centrifuged flour that was first sonicated in a 2% (w/v) SDS solution with 0.05M sodium phosphate buffer (pH 6.9) for 30 sec at a power level of 5W using a sonic disrupter (Tekmar, Cincinnati, OH). These fractions were: 1)  $M_r > 100k$  (polymers, primarily glutenin), designated as PT1; 2)  $M_r$  25k–100k (monomers, mostly gliadins but also secalins for genotypes possessing the wheat-rye translocations) designated as PT2; and 3)  $M_r < 25k$  (low  $M_r$  monomers, such as albumin and globulins), designated as PT3. Composition analysis of the second portion was similar to that of the first, with exception of a preliminary sonication step in 0.04M NaCl. Centrifugation was performed beforehand and the pellet was resonicated in SDS solution. Thus, the second and third SE-HPLC fractions (designated PR2 and PR3, respectively) differed from the corresponding fractions of the first portion by an absence of secalins (originally present in second fraction) and saline-extractable low  $M_r$  monomers (originally present in third fraction).

Protein fractions were eluted using an HPLC system (BioRad, Richmond, CA) equipped with a 10- $\mu$ m silica column (30-nm pore) (Waters Protein-Pak 300 SW). Absorption at 210 nm was integrated to determine the fractional area of each peak. Initially recorded as percent by weight of extractable protein (with PT1 + PT2 + PT3 = 100), the five protein fractions were subsequently expressed on a weight concentration basis by multiplication of each by crude

protein content, assuming that the proportion of unextractable proteins within all fractions was constant. Subsequent SE-HPLC work (data not shown), involving six wheat cultivars (17 samples per cultivar) and standard solutions of bovine serum albumin (BSA) used as concentration references for elution peaks, showed that  $\approx 94\%$  of all endosperm proteins were extractable. A similar value for extraction efficiency ( $\bar{x} = 0.947$ ,  $SD = 0.043$ ), based on calibration to Kjeldahl-determined protein, was reported in Lee et al (1995).

Crude protein content ( $N \times 5.7$ ) was measured on duplicate portions (150 mg/portion) of each flour sample by combustion (model FP-428, Leco Mfg. Co., St. Joseph, MI). Protein content was reported on an as-is moisture basis. While moisture determination was not routinely measured on each sample, nine randomly selected samples exhibited a range of 13–14% using Approved Method 44-15A (AACC 1995). Water-soluble pentosan content was measured according to Hashimoto et al (1987). SDS sedimentation volume was determined on a 2-g portion of each sample using Approved Method 56-61A (AACC 1995).

Mixograph analyses were performed on all samples using a National Manufacturing mixograph (Lincoln, NE) as described in Graybosch et al (1996). Four indices were recorded for each sample: the time to peak dough development, (MIXTIME), the height of peak at maximum resistance (MIXHT), the width of the curve at maximum resistance (MIXWD), and the width of the curve at 2 min past the point of maximum resistance (MIXTOL), which is a good indicator of the dough's tolerance to overmixing.

Laboratory error for each constituent or property (constituent is inclusive of SDS sedimentation volume and mixograph properties), excluding protein content, was estimated by analysis of variance. The mean squared error of field replication within location and genotype was the best estimate of laboratory error of each constituent. For protein content, laboratory error was estimated as the standard deviation of 75 determinations of a control wheat sample that was run periodically throughout the protein by combustion analyses.

Because spectral analysis was performed as a last stage in the handling of these samples, some samples contained an insufficient amount of material for scanning. Altogether, 843 of the original 1,020 samples contained sufficient material for spectral analysis. A spectrophotometer (model 6500, NIRSystems, Silver Spring, MD) equipped with a rotating drawer attachment collected and stored the spectral readings. For each usable sample, NIR  $\log(1/R)$  spectra (1,100–2,498 nm) were collected at 2-nm intervals on two successive fillings ( $\approx 4$  g/filling) of a standard ring cell. Diffuse reflectance readings were referenced to those from a ceramic disk. Spectra within each pair of fillings were averaged and recorded as one  $\log(1/R)$  spectrum per sample.

TABLE I  
Descriptive Statistics of Samples<sup>a</sup>

Constituents <sup>b</sup>	Calibration Set		Validation Set	
	Range	Mean $\pm$ SD	Range	Mean $\pm$ SD
Crude protein	6.8–20.1	12.7 $\pm$ 2.1	6.6–19.9	12.5 $\pm$ 2.2
PT1	2.6–8.6	5.2 $\pm$ 1.0	2.4–8.5	5.2 $\pm$ 1.1
PT2	2.6–12.5	6.0 $\pm$ 1.4	2.3–11.0	5.8 $\pm$ 1.4
PT3	0.6–2.3	1.5 $\pm$ 0.2	0.3–4.4	1.6 $\pm$ 0.3
PR2	0.4–12.1	4.6 $\pm$ 1.4	1.3–9.4	4.5 $\pm$ 1.4
PR3	0.2–2.5	0.9 $\pm$ 0.2	0.5–2.0	1.0 $\pm$ 0.3
Water-soluble pentosans, $\mu$ g/100 g of flour	321–1131	597 $\pm$ 125	319–1321	616 $\pm$ 142
SDS sedimentation volume, $cm^3$	12–46	33 $\pm$ 6	13–47	32 $\pm$ 7
Mixograph indices				
MIXTIME	1.6–7.5	3.7 $\pm$ 1.1	1.7–6.9	3.7 $\pm$ 1.0
MIXHT	39–102	69 $\pm$ 12	36–102	69 $\pm$ 12
MIXWD	10–60	29 $\pm$ 9	12–56	29 $\pm$ 8
MIXTOL	7–40	20 $\pm$ 5	7–40	20 $\pm$ 5

<sup>a</sup> Calibration set:  $n = 423$  for all constituents except mixograph properties, for which  $n = 396$  is a subset. Validation set:  $n = 420$  and 385, respectively.

<sup>b</sup> Protein fractions (wt%): crude protein =  $N \times 5.7$ , PT1 = glutenin, PT2 = gliadins with secalins, PT3 = albumin and globulins, PR2 = saline unextractable albumin and globulins. MIXTIME = time to peak (min), MIXHT = height at peak (mm), MIXWD = width at peak (mm), MIXTOL = width at 2 min past peak (mm).

## Modeling

Samples were divided into two sets: calibration for the purpose of developing quantitative models, and validation for assessing the performance of the models. Calibration samples originated from block 1 exclusively, while validation samples originated from block 3 exclusively. For protein-related constituents, pentosan content, and SDS sedimentation volume, sample numbers were 423 and 420 for calibration and validation sets, respectively. Slightly smaller subsets ( $n = 396$  and  $385$ , respectively) of these parent sets were used for models of mixograph properties, owing to the lack of complete mixograph data on 62 of 843 samples. The range, mean, and standard deviation of each constituent within a set are contained in Table I. Generally, the ranges, means, and standard deviations in the validation set were equivalent to the corresponding values in the calibration set.

Chemometric modeling consisted of an individual partial least squares (PLS) regression on mean-centered spectra for each constituent. Data were pretreated with a multiplicative scatter correction to the mean spectrum of the calibration set (Geladi et al 1985) to minimize sample-to-sample differences in filling density. All 700 readings per spectrum were used in a one-sample-out, 20-factor cross-validation procedure during model development. At each PLS factor ranging from 1 to 20, squares of residual values (model minus reference) of the rotated-out samples were accumulated. For each constituent, the optimal model was the one that possessed the fewest factors with a sum of squares of residuals, defined as the PRESS (ASTM 1995), fell within 5% of that of the model that produced the minimum sum.

Model performance was evaluated on the validation set and reported as: 1) correlation coefficient ( $r$ ) determined from the linear regression of the validation reference values onto model-predicted values; 2) standard deviation of validation errors (SDV), also commonly referred to as standard error of performance (SEP); 3) bias, defined as the mean difference between modeled and reference values; 4) ratio of the standard deviation of the reference values (SD) to the SDV (RPD, defined by Williams [1987]) where  $RPD = SD/SDV$ ; and 5) ratio of the variance of the reference values with vari-

ance of the NIR model removed to the variance of the reference values with variance caused by the reference technique removed where  $RAP = (SD^2 - SDV^2)/(SD^2 - \text{laboratory error}^2)$  (Martens and Naes 1989).

The RPD is often used as a dimensionless gauge of the ability of an NIR model to predict a constituent beyond that which is attributable to chance. A value of 1.0 indicates that the variability attributable to the NIR model (assuming negligible laboratory error) has the same level as the variance of the naturally occurring constituent, thus indicating a lack of modeling power. Values of  $\geq 2.5$  indicate that the NIR model may be suitable for screening and breeding programs, while those  $>5.0$  are potentially useful in quality control (Williams and Sobering 1993). RAP is also a dimensionless gauge, but differs from RPD in that the variance of the reference method (or its best estimate) is removed. A value of 1.0 represents the best attainable condition for the NIR model and indicates that the modeling error is at the same level as the laboratory error. Although both gauges are generally similar in definition, RAP is designed to examine the performance of the NIR model with respect to the inherent error in the reference technique.

## RESULTS AND DISCUSSION

The NIR model validation statistics for each constituent are summarized in Table II. The optimal number of factors, as determined by cross-validation of the calibration set, ranged from 7 (PR2) to 17 (MIXTIME). The excellent performance for crude protein (SDV = 0.126%, bias = -0.005%) is consistent with the literature (Osborne and Fearn 1983, Williams et al 1983). Based on the RPD and RAP, the best protein component models, aside from crude protein itself, were those for intermediate molecular weight ( $M_r$ , 25–100k) proteins, with (PT2) and without secalins (PR2). The value of RPD = 3.0 for PT2 was slightly higher than RPD = 2.7 for PR2, though

TABLE II  
Summary of Near-Infrared Reflectance (NIR)  
Partial Least Squares Regression Analyses<sup>a</sup>

Constituents <sup>b</sup>	Factors <sup>c</sup>	SDV <sup>d</sup>	Bias <sup>e</sup>	RPD <sup>f</sup>	RAP <sup>g</sup>
Crude protein	11	0.126	-0.005	17.2	1.00
PT1	13	0.442	0.078	2.4	0.89
PT2	12	0.493	0.026	3.0	0.92
PT3	11	0.271	-0.110	1.0	0.08
PR2	7	0.530	0.080	2.7	0.95
PR3	9	0.263	-0.140	1.0	0.09
Water-soluble pentosans, μg/100 g of flour	15	99.2	-15.08	1.4	0.65
SDS sedimentation volume, cm <sup>3</sup>	13	3.33	0.301	2.0	0.84
Mixograph indices					
MIXTIME	17	0.659	-0.007	1.6	0.81
MIXHT	12	4.186	0.319	2.8	0.98
MIXWD	12	3.693	0.381	2.2	0.93
MIXTOL	10	2.902	0.025	1.8	0.79

<sup>a</sup>  $n = 420$  validation samples for all constituents, except mixograph properties for which  $n = 385$  is a subset.

<sup>b</sup> Protein fractions (wt%): crude protein =  $N \times 5.7$ , PT1 = glutenin, PT2 = gliadins with secalins, PT3 = albumin and globulins, PR2 = gliadins, PR3 = saline unextractable albumin and globulins. MIXTIME = time to peak (min), MIXHT = height at peak (mm), MIXWD = width at peak (mm), MIXTOL = width at 2 min past peak (mm).

<sup>c</sup> Optimal number determined by cross-validation of calibration samples.

<sup>d</sup> Standard deviation of validation errors.

<sup>e</sup> Mean NIR value minus mean biochemical value.

<sup>f</sup> Standard deviation of biochemical or physical value/SDV ( $= SD / SDV$ ), dimensionless.

<sup>g</sup> Relative ability of prediction =  $[SD^2 - SDV^2]/[SD^2 - (\text{error of biochemical or physical measurement technique})^2]$ , dimensionless.

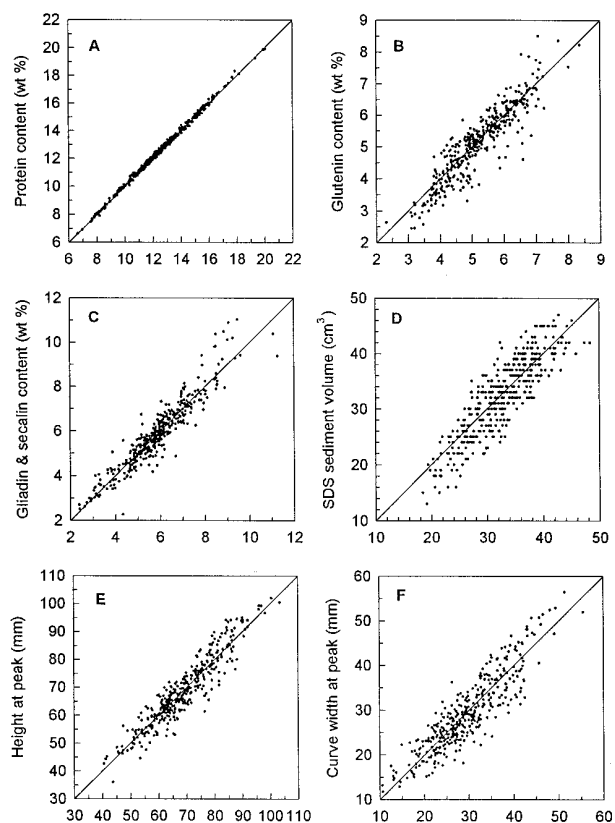


Fig. 1. Reference (y-axis) vs. near-infrared reflectance (x-axis) modeled values for six of 12 constituents studied, evaluated on validation samples, 45° line included. A, Crude protein =  $N \times 5.7$ ; B, PT1 = glutenin; C, PT2 = gliadins with secalins; D, SDS sedimentation volume; E, MIXHT = height at peak; F, MIXWD = width at peak.  $n = 420$  for A–D, 385 for E and F.

the opposite ranking occurred for the RAP values (0.92 and 0.95, respectively). Among the models for other protein components, only the glutenin (PT1) model, with RPD = 2.4 and RAP = 0.89, demonstrated a predictive ability beyond that of random event. At RPD values of 1.0, both albumin and globulin models, PT3 and PR3, could not be modeled by NIR, most likely because of their relatively low concentrations.

Among the nonprotein models, SDS sedimentation volume (RPD = 2.0, RAP = 0.84) was more accurately modeled than pentosan content (RPD = 1.4, RAP = 0.65). Again, the actual values for pentosan content were probably too low for quantification by NIR. Of the four mixograph properties analyzed, the best NIR model was that for MIXHT (RPD = 2.8 and RAP = 0.93), followed by that for MIXWD (RPD = 2.2 and RAP = 0.93).

Plots of reference vs. NIR predicted (validation set) values for the six constituents (crude protein, PT1, PT2, SDS, MIXHT, and MIXWD) of highest RPD are shown in Fig. 1. Visual examination of these plots indicates that NIR predictions of each constituent resulted in a uniform envelope about the line of zero error (45° line) throughout the range in data. With the exception of PR2, the performance of the models for the remaining constituents (PT3, PR3, pentosan, MIXTIME, and MIXTOL) was probably insufficient for future application.

Because of the inherently strong performance of the crude protein model, it was reasoned that successful NIR modeling of the other biochemical constituents (especially the protein fractions) and physical properties could arise as a result of correlations between these constituents (or properties) and protein content. Therefore, performance of the NIR models was additionally evaluated by partial correlation analysis in which the protein content as a covariate (CORR procedure) (SAS Institute, Cary, NC) was first retained and then removed in the determination of the correlation coefficient (Table III). Linear regression of crude protein by combustion onto crude protein by NIR resulted in  $r = 0.999$ . Similarly, a strong correlation ( $r = 0.91$ ) was found between the reference and NIR values

for PT1. Because PT1 and crude protein were positively correlated ( $r = 0.80$ ), the removal of crude protein as a covariate during partial correlation analysis resulted in a drop in correlation ( $r = 0.73$ ) between the reference and NIR values for PT. Similar to the results for PT1, NIR models for PT2 and PR2 (for reference vs. NIR,  $r = 0.94$  and  $0.93$ , respectively) exceeded that which could be ascribed to the correlation between these constituents and crude protein ( $r = 0.90$  and  $0.92$ ). However, with removal of crude protein as a covariate, the correlations between reference and NIR values for these constituents dropped to  $r = 0.64$  and  $0.36$ , respectively. This suggests that the NIR models for these protein components are indeed aided by, but not wholly dependent on, crude protein. Williams et al (1984) reached a similar conclusion on the NIR prediction of four limiting amino acids (lysine, methionine, threonine, and tryptophan) in wheat.

For SDS sedimentation volume, the dependency on crude protein for success in NIR modeling was less than that for PT1, PT2, or PR2. Possessing a lower correlation with crude protein ( $r = 0.70$ ), the correlation between reference and NIR values of SDS sedimentation volume, when compared to the protein components, was less affected by the removal of crude protein as a covariate ( $r = 0.87$  before removal to  $r = 0.74$  after removal). Lastly, the success of the NIR models for MIXHT and MIXWD appears to be strongly dependent on the correlation to protein content ( $r = 0.87$  and  $0.83$  for MIXHT and MIXWD, respectively) as seen by the decline in correlation between the reference and NIR model when protein content was removed as a covariate (from  $r = 0.94$  to  $r = 0.60$  for MIXHT, and from  $r = 0.89$  to  $r = 0.60$  for MIXWD).

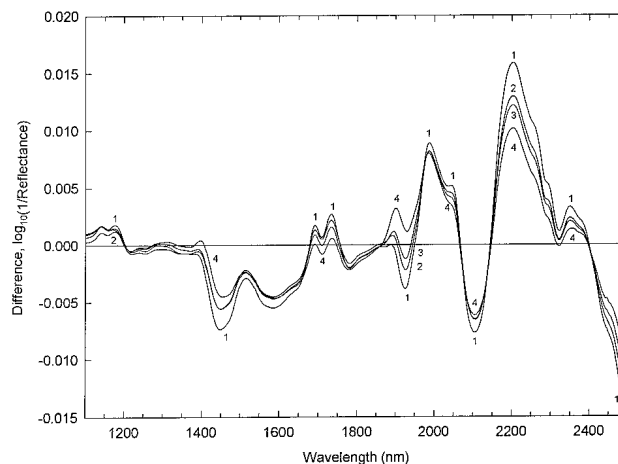
Spectral interpretation of NIR PLS models through examination of the regression vector or individual loadings vectors is typically difficult and therefore was not attempted. Alternatively, to understand the spectral sensitivity to a constituent, a difference spectrum was formed by taking the average of the 25 multiplicatively scatter-corrected spectra that corresponded to the highest values of the constituent within the calibration set and subtracting the similarly formed average spectrum that corresponded to the lowest 25 values of the set. Figure 2 contains the difference spectra, labeled as curves 1–4, respectively, for protein content ( $\bar{X}_{\text{high-25}} - \bar{X}_{\text{low-25}} = 17.1\% - 8.6\%$ ), PT1 (7.2% - 3.2%), MIXHT (94.6 mm - 47.0 mm), and SDS sedimentation volume (43.5 cm<sup>3</sup> - 19.4 cm<sup>3</sup>). All four difference spectra were generally similar, with positive peaks occurring at 1,146, 1,180, 1,392, 1,518, 1,694, 1,736, 1,894, 1,990, 2,204, and 2,352 nm, and negative peaks at 1,450, 1,588, 1,782, 1,928, 2,108, and 2,324 nm. The greatest dif-

**TABLE III**  
Correlation ( $r$ ) Between Constituent and Crude Protein and Between Constituent as Measured by Reference Method and Near-Infrared Reflectance (NIR) Modeled Value, Before and After Removal of Crude Protein as a Covariate<sup>a</sup>

Constituent <sup>b</sup>	Reference Method vs. NIR Model		
	Constituent vs. Crude Protein	Before Covariate Removal	After Covariate Removal
Crude protein	—	0.999	—
PT1	0.805	0.911	0.729
PT2	0.899	0.941	0.641
PT3	-0.026	0.243	0.313
PR2	0.916	0.927	0.361
PR3	0.087	0.264	0.253
Water-soluble pentosans, $\mu\text{g}/100\text{ g}$ of flour	-0.322	0.714	0.675
SDS sedimentation volume, cm <sup>3</sup>	0.702	0.873	0.737
Mixograph indices			
MIXTIME	-0.359	0.773	0.737
MIXHT	0.872	0.935	0.688
MIXWD	0.827	0.893	0.598
MIXTOL	0.648	0.826	0.672

<sup>a</sup>  $n = 420$  validation samples except for mixograph properties, for which  $n = 385$  is a subset. All correlations are significant ( $P = 0.001$ ), except PT3 vs. crude protein and PR3 vs. crude protein, for which both correlations are not significant.

<sup>b</sup> Protein fractions (wt%): crude protein =  $N \times 5.7$ , PT1 = glutenin, PT2 = gliadins with secalins, PT3 = albumin and globulins, PR2 = gliadins, PR3 = saline unextractable albumin and globulins. MIXTIME = time to peak (min), MIXHT = height at peak (mm), MIXWD = width at peak (mm), MIXTOL = width at 2 min past peak (mm).



**Fig. 2.** Difference spectra, formed by ranking (lowest to highest) calibration samples ( $n = 423$ , except for MIXHT subset  $n = 396$ ), then subtracting average spectrum of ranks 1–25 from average spectrum of ranks ( $n - 24$ )– $n$ . Ranking procedure performed separately for each constituent. Spectra, before averaging were multiplicatively scatter-corrected to the average spectrum ( $n = 423$  or  $396$ ). Difference spectra identified by rank sets are 1 = crude protein, 2 = PT1 (glutenin), 3 = MIXHT (height at peak), 4 = SDS sedimentation volume.

ferences among all four difference spectra, including the nonamplitude differences such as band shape and peak position, occurred between 1,850 and 2,350 nm, a region predominated by combination bands of CH, OH, and NH (Osborne and Fearn 1986).

The similarity among the spectra in Fig. 2 is partly because of the commonality in samples making up the high-25 (or low-25) in each set. For instance, ≈60% of the high-25 protein samples are also high-25 SDS samples, with a similar percentage of overlap for most of the other pairs of categories. The noncommon samples are still enough to cause the clear differences (1,850 – 1,950 nm) between the SDS and protein difference spectra. The similarity between the PT1 and MIXHT difference spectra appears to agree with Bangur et al (1997), who determined highly significant correlations between extensigraph maximum resistance and the proportion of polymeric protein (measured by SE-HPLC) to total protein. Gupta et al (1993) observed that the amount of polymeric protein unextractable in 0.5% SDS solution relative to either the total amount of polymeric protein or total amount of protein has a very large influence on extensigraph maximum resistance or mixograph peak dough development time

Because SE-HPLC separates the protein fractions based on semiautomatic molecular weight, the three main separation categories (glutenin, gliadins, and albumin and globulins) do not correspond perfectly with those fractions more classically defined by solubility methodologies (Singh et al 1990a). Rather, small but significant quantities of the two minor solubility fractions appear within each major SE-HPLC fraction. Thus, if the assumption is made that differences in infrared absorption are caused by the same molecular interactions that are responsible for the unique solubility characteristics of the protein fractions, then an inherent error in the NIR model for each SE-HPLC determined fraction will exist to the extent of this difference.

## CONCLUSIONS

This study indicates that certain quality-related biochemical components and dough-handling properties of wheat flour may be determined by NIR at accuracies sufficient for screening purposes in breeding programs. Contents of glutenin ( $M_r > 100$  kDa) and gliadins, with and without secalins from wheat-rye translocations ( $M_r 25-100$  kDa), were reasonably well modeled by PLS analysis on NIR spectra, as were SDS sedimentation volume, mixograph peak time, and mixograph peak width. The success of modeling goes beyond the degree of correlation between each of these quality parameters and protein content, a constituent easily modeled by NIR. NIR models for albumin and globulin content and for pentosan content were not sufficiently accurate for breeding programs. Further research directed toward spectral interpretation of wheat gluten macromolecules and their interaction with starch may lead to enhancements of NIR quantitative models that predict breadmaking quality.

## ACKNOWLEDGMENTS

Partial support of this work was by a grant from Anheuser-Busch Companies, St. Louis, MO. We are thankful for the technical assistance of Joyce Shaffer (Beltsville) and Vern Hansen (Lincoln).

## LITERATURE CITED

- American Association of Cereal Chemists. 1995. Approved Methods of the AACC, 9th ed. Method 44-15A, approved October 1975, revised October 1981 and October 1994; Method 56-61A, approved April 1964, reviewed October 1982 and October 1994. The Association: St. Paul, MN.
- ASTM. 1995. American Society for Testing and Materials Standard Practices. E1655, infrared, multivariate, quantitative analysis. The Society: West Conshohocken, PA.
- Bangur, R., Batey, I. L., McKenzie E., and MacRitchie, F. 1997. Dependence of Extensograph parameters on wheat protein composition measured by SE-HPLC. *J. Cereal Sci.* 25:237-241.
- Bietz, J. A. 1987. Genetic and biochemical studies of nonenzymatic endosperm proteins. Pages 215-241 in: *Wheat and Wheat Improvement*, 2nd ed. E. G. Heyne, ed. ASA-CSSA-SSSA: Madison, WI.
- Delwiche, S. R., and Weaver, G. 1994. Bread quality of wheat flour by near-infrared spectrophotometry: Feasibility of modeling. *J. Food Sci.* 59:410-415.
- Eckert, B., Amend, T., and Belitz, H. D. 1993. The course of the SDS and Zeleny sedimentation tests for gluten quality and related phenomena studied using the light microscope. *Z. Lebensm. Unters. Forsch.* 196:122-125.
- Geladi, P., MacDougall, D., and Martens, H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39:491-500.
- Graybosch, R. A., Peterson, C. J., Hansen, L. E., Worrall, D., Shelton, D. R., and Lukaszewski, A. 1993. Comparative flour quality and protein characteristics of 1BL/1RS and 1AL/1RS wheat-rye translocation lines. *J. Cereal Sci.* 17:95-106.
- Graybosch, R. A., Peterson, C. J., Shelton, D. R., and Baenziger, P. S. 1996. Genotypic and environmental modification of wheat flour protein composition in relation to end-use quality. *Crop Sci.* 36:296-300.
- Gupta, R. B., Khan, K., and MacRitchie, F. 1993. Biochemical effects of flour properties in bread wheats. I. Effects of variation in the quantity and size distribution of polymeric protein. *J. Cereal Sci.* 18:23-41.
- Hashimoto, S., Shogren, M. D., and Pomeranz, Y. 1987. Cereal pentosans: Their estimation and significance. I. Pentosans in wheat and milled wheat products. *Cereal Chem.* 64:30-34.
- Lee, J. H., Graybosch, R. A., and Peterson, C. J. 1995. Quality and biochemical effects of a 1BL/1RS wheat-rye translocation in wheat. *Theor. Appl. Genet.* 90:105-112.
- Martens, M., and Naes, T. 1989. Methods for calibration. Pages 73-163 in: *Multivariate calibration*. H. Martens and T. Naes, eds. John Wiley and Sons: New York.
- MacRitchie, F., du Cros, D. L., and Wrigley, C. W. 1990. Flour polypeptides related to wheat quality. Pages 79-145 in: *Advances in Cereal Science and Technology*, Vol. 10. Y. Pomeranz, ed. Am. Assoc. Cereal Chem.: St. Paul, MN.
- Norris, K. H. 1978. Near infrared spectroscopy—The present and future. Pages 245-251 in: *Cereals '78: Better Nutrition for the World's Millions*. Y. Pomeranz, ed. Am. Assoc. Cereal Chem.: St. Paul, MN.
- Orth, R. A., and Bushuk, W. 1972. A comparative study of the proteins of wheats of diverse baking qualities. *Cereal Chem.* 49:268-275.
- Osborne, B. G., and Fearn, T. 1983. Collaborative evaluation of universal calibrations for the measurement of protein and moisture in flour by near infrared reflectance. *J. Food Technol.* 18:453-460.
- Osborne, B. G., and Fearn, T. 1986. Theory of near-infrared spectrophotometry. Pages 20-42 in: *Near Infrared Spectroscopy in Food Analysis*. Longman Scientific and Technical: Essex, UK.
- Payne, P. I., Corfield, K. G., and Blackman, J. A. 1979. Identification of a high-molecular-weight subunit of glutenin whose presence correlates with breadmaking quality in wheats of related pedigree. *Theor. Appl. Genet.* 55:153-159.
- Singh, N. K., Donovan, G. R., Batey, I. L., and MacRitchie, F. 1990a. Use of sonication and size-exclusion high-performance liquid chromatography in the study of wheat flour proteins. I. Dissolution of total proteins in the absence of reducing agents. *Cereal Chem.* 67:150-161.
- Singh, N. K., Donovan, G. R., and MacRitchie, F. 1990b. Use of sonication and size-exclusion high-performance liquid chromatography in the study of wheat flour proteins. II. Relative quantity of glutenin as a measure of breadmaking quality. *Cereal Chem.* 67:161-170.
- Williams, P. C. 1987. Variable affecting near-infrared reflectance spectroscopic analysis. Pages 143-167 in: *Near-Infrared Technology in the Agricultural and Food Industries*. P. C. Williams and K. H. Norris, eds. Am. Assoc. Cereal Chem.: St. Paul, MN.
- Williams, P. C., and Sobering, D. C. 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* 1:25-32.
- Williams, P. C., Norris, K. H., Gehrke, C. W., and Bernstein, K. 1983. Comparison of near-infrared methods for measuring protein and moisture in wheat. *Cereal Foods World* 28:149-152.
- Williams, P. C., Preston, K. R., Norris, K. H., and Starkey, P. M. 1984. Determination of amino acids in wheat and barley by near-infrared reflectance spectroscopy. *J. Food Sci.* 49:17-20.

[Received September 2, 1997. Accepted March 24, 1998.]