Prediction of Total Dietary Fiber by Near-Infrared Reflectance Spectroscopy in Cereal Products Containing High Sugar and Crystalline Sugar

Sandra E. Kays,* Franklin E. Barton, II, William R. Windham, and David S. Himmelsbach

Richard B. Russell Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, P.O. Box 5677, Athens, Georgia 30604

Crystalline sugar has unique spectral characteristics that influence the near-infrared (NIR) region of the spectrum often used to assess product composition. The current study investigated the potential of expanding an existing NIR spectroscopic calibration for the prediction of total dietary fiber in cereal products to include products with high sugar and crystalline sugar content. Using AOAC Procedure 991.43 as the reference method, an NIRSystems monochromator and ISI software for scanning and data analysis, and a selection algorithm to select representative high-sugar samples, a sugar-expanded partial least-squares model ($n = 100$) was developed for prediction of dietary fiber. The standard error of cross-validation, R^2, standard error of performance ($n = 45$ independent validation samples), and r^2 were 1.88%, 0.98, 1.40%, and 0.99, respectively. The NIR model for prediction of total dietary fiber in cereal products was, thus, expanded to include samples with high sugar and crystalline sugar content.

Keywords: *Dietary fiber; near-infrared; chemometrics; cereal*

INTRODUCTION

A previous paper (Kays et al., 1996) has described the development of a near-infrared, partial least-squares (PLS) model for the prediction of total dietary fiber in cereal and grain products. Development of this model was prompted by changes in U.S. food labeling legislation, namely the Nutrition Labeling and Education Act (*Code of Federal Regulations*, 1995), which has led to an increase in food composition analysis, expanding the need for rapid and accurate methods for the measurement of nutrients. Dietary fiber, due to its beneficial effects on human health, is one of the nutrients included in the Nutrition Labeling and Education Act. The current method of analyzing dietary fiber in the United States is the AOAC approved method (AOAC, 1990, 1992; Lee et al., 1992). Dietary fiber is defined as the polysaccharides and lignin that are not hydrolyzed by human alimentary enzymes (Trowell et al., 1976; Lee and Prosky, 1995), and the AOAC method meets this definition in measuring the nonstarch polysaccharides, lignin, and residual starch present in foods. Although the AOAC method provides precise, repeatable results with a wide range of foods, it is very time-consuming, taking a minimum of 2 days to complete a determination. Near-infrared spectroscopy (NIRS) has provided a valuable analytical tool in agriculture and in the food industry for rapid and accurate analysis of certain product constituents (Williams and Norris, 1987; Osborne et al., 1993; Kays et al., 1997). In addition to being rapid, NIRS requires very little sample preparation, requires no chemicals, produces no chemical waste, and has, therefore, been investigated as a means of analyzing fiber in several foods (Baker, 1983, 1985; Horvath et al., 1984; Williams et al., 1991; Kays et al., 1996).

The NIR reflectance model developed for total dietary fiber prediction in cereal products (Kays et al., 1996) utilized a variety of grains including wheat, oats, corn, rice, rye, barley, millet, buckwheat, multiple-grain products, and mixtures of cereal products with commercial oat and wheat fibers with a range in total dietary fiber from 0 to 52%. The model was sufficiently accurate for food labeling and monitoring purposes with a standard error of cross-validation (SECV) and multiple coefficient of determination (R^2) of 1.6% total dietary fiber and 0.99, respectively, and a standard error of performance (SEP) of 1.5% dietary fiber (Kays et al., 1996). The model appeared to be primarily influenced by effects due to C−H and O−H groups in the carbohydrate and water absorbing regions of the spectrum, although minor influences from constituents such as lipid and protein were also apparent. Further investigation demonstrated that prediction of total dietary fiber is based on information related to both dietary fiber and water concentration, and a calibration was developed to predict fiber in cereal products having a wide range in moisture content (Windham et al., 1997).

The original model did not include cereal products with >20% sugar and did not predict total dietary fiber as well when tested with these products—unless sugar was extracted prior to NIR scanning (Kays et al., 1996). The sugar-coated cereals, some granolas, muesli, many other breakfast cereals, and the muffin and cake mixes contain >20% sugar, and a large number of these products contain sugar in the crystalline form. The high-sugar and crystalline sugar products represent a significant portion of the cereal product market. High-sugar-containing cereal products were not included in the original study partly to avoid too diverse a data set initially and partly because of difficulties encountered with the reference method. Extraction of excess amounts of sugar is required before using the AOAC procedure (Lee et al., 1992). The current work demonstrates the spectral characteristics of the monosaccharides and disaccharide commonly present in cereal products and high-sugar cereal products and investigates expansion of the original NIR calibration for total dietary fiber

* Author to whom correspondence should be addressed [telephone (706) 546-3338; fax (706) 546-3607].

S0021-8561(97)00326-9

prediction to include cereal products with high sugar content and crystalline sugar. In addition, observations are made on associations between heavily loaded values in the PLS loadings of the calibration and known vibrations of interest in the study of NIR spectra.

MATERIALS AND METHODS

Instrumentation. The NIRSystems 6500 monochromator (NIRSystems, Silver Spring, MD) used for this study is a visible/near-infrared scanning monochromator having a tungsten source and a holographically ruled grating. Diffusely reflected radiation is detected from 400 to 2500 nm in 2 nm intervals. The lower wavelength region (400−1098 nm) is observed by a pair of silicon detectors located 20 cm from the surface of the sample cell and at an angle of 45° to the incident beam. The upper wavelength region (1100-2500 nm, used for the present study) is observed by two pairs of lead sulfide detectors in the same orientation as the silicon detectors. Reference reflectance values are obtained using a ceramic block. Samples are presented in a sample cell that is placed on an oscillating shaft with axis of rotation parallel to the incident radiation.

The Nicolet 850 (Nicolet Instrument Corp., Madison, WI) was employed as a mid-infrared Fourier transform (FT) spectrometer by configuring it with a globar source, a KBr beamsplitter, and a MCT/B liquid nitrogen cooled detector. Mid-infrared diffuse reflectance spectra were collected using the Collector accessory with macro cups. The sample compartment was continuously purged with air supplied by a Balston Model 75-62 (Balston Inc., Haverhill, MA) purge gas generator, which removed most of the CO_2 and water vapor to a dew point of −73 °C. Data collection and processing were performed using Omnic software (Nicolet Instrument Corp., Madison, WI).

Reagents. Heat stable α-amylase, A 3306; protease, P 3910; amyloglucosidase, A 9913; acid-washed Celite, C 8656; total dietary fiber control kit, TDF-100A; 2-(*N*-morpholino)-ethanesulfonic acid (MES), M 8250; tris(hydroxymethyl)-aminomethane (Tris), T 1503; D-(+)-glucose, G 7528; D-(−)-fructose, F 0127; and sucrose, S 7903; were purchased from Sigma Chemical Co., St. Louis, MO. Buffer (MES/Tris, 0.05 M) was prepared and adjusted to pH 8.2 at 24 °C. Adjustment of the buffer to pH 8.3 is required if the temperature is 20 °C and to pH 8.1 if the temperature is 28 °C, with interpolation for intermediate temperatures.

Sample Preparation. The cereal and grain products used for the original calibration ($n = 77$) and validation ($n = 30$) data sets (Kays et al., 1996 and present work) contained <20% sugar and <10% fat and included breakfast cereals, crackers, pastas, brans, and flours. Products were dry milled to <500 μm in a cyclone mill (Cyclotec 1093 sample mill, Perstorp Analytical, Silver Spring, MD). Due to limited numbers of cereal products with high total dietary fiber content, commercial oat and wheat fibers (range in total dietary fiber 90−98%) were mixed with several processed cereals to provide 23 samples with high, medium, or low dietary fiber content. Fifteen samples in the original calibration set and eight samples in the original validation set were prepared in this way (Tables 1 and 2). Commercial oat and wheat fibers were provided by Canadian Harvest U.S.A. L.P. (Cambridge, MN).

High-sugar cereal products used for expansion of the calibration and validation data sets contained >20% sugar and included breakfast cereals, sugar-coated breakfast cereals, crackers, cookies, and muffin and cake mixes. Thirty-nine high-sugar samples were available for expansion of the calibration data set and 15 for validation data set expansion. High-sugar cereal samples, sucrose, glucose, and fructose were mixed with liquid nitrogen to facilitate grinding in the cyclone mill. On the basis of product label values, the range, mean, and standard deviation of sugar content of the high-sugar cereal samples used in the sugar-expanded calibration data set were 21.8−53.3, 35.4, and 9.3%, respectively, and those in the sugar-expanded validation data set were 22.2−55.6, 34.4, and 9.2%, respectively.

Table 1. Cereal and Grain Products in the Calibration Data Set [Range, Mean, and Standard Deviation (SD) of Total Dietary Fiber (TDF) Percent][a]

sample	product type	no. of products	range in TDF%	mean TDF%	SD TDF%
original	wheat	25	2.9−41.1	13.3	9.7
	oats	6	8.1−19.0	12.6	4.32
	corn	4	1.2−13.6	6.5	5.4
	rice	8	0.7−4.3	2.2	1.5
	rye	1	38.4		
	barley	2	12.4−19.4	15.86	
	millet	3	2.7−3.6	3.1	0.5
	multiple grain	13	6.5−41.0	16.9	13.4
	commercial fiber and cereal product mixes	15	6.4−52.1	31.2	14.7
high sugar	wheat	5	3.7−14.9	8.8	4.2
	oats	3	6.7−10.6	8.5	2.0
	corn	3	0.6−2.1	1.4	0.8
	rice	1	1.5		
	multiple grain	11	2.2−11.9	7.3	2.9

[a] Grains represented in multiple-grain products in the original sample set are as follows: wheat (12), oats (10), corn (5), rice (5), rye (7), barley (10), millet (3), buckwheat (4), and amaranth (2). Grains represented in the multiple-grain products in the high-sugar sample set are as follows: wheat (4), oats (3), corn (2), rice (2) rye (1), and barley (1). The number of products with each grain type is in parentheses.

Table 2. Cereal and Grain Products in the Validation Data Set [Range, Mean, and Standard Deviation (SD) of Total Dietary Fiber (TDF) Percent][a]

sample	product type	no. of products	range in TDF%	mean TDF%	SD TDF%
original	wheat	7	3.6−43.7	13.0	14.1
	oats	1	15.4		
	corn	2	1.8−4.9	3.4	
	rice	3	1.2−2.2	1.5	0.6
	rye	2	13.3−18.1	15.7	
	multiple grain	10	2.0−39.3	16.8	13.1
	commercial fiber and cereal product mixes	5	12.5−42.8	29.4	12.7
high sugar	wheat	1	5.8		
	oats	1	6.9		
	corn	1	2.4		
	rice	1	1.4		
	multiple grain	9	1.0−10.6	5.3	3.1

[a] Grains represented in multiple-grain products in the original sample set are as follows: wheat (6), oats (3), corn (6), rice (3), rye (3), barley (2), millet (1), amaranth (2), and buckwheat (1). Grains represented in the multiple-grain products in the high-sugar sample set are as follows: wheat (8), oats (6), corn (5), and rice (4). The number of products with each grain type is in parentheses.

Reference Laboratory Method for Total Dietary Fiber. Total dietary fiber in all samples was measured in the laboratory by AOAC Approved Method 991.43 (AOAC, 1992; Lee et al., 1992), with modifications described previously (Kays et al., 1996). Purity and activity of enzymes used in the AOAC procedure were monitored as described previously (Kays et al., 1996). Before the AOAC procedure was performed, samples containing >20% sugar were desugared by extracting four times with 85% ethyl alcohol (10 mL/g of sample) for 15 min with stirring and evaporated in a vacuum oven overnight at 30 °C. The sugar extracted was calculated for each sample on the basis of sample weight before and after extraction. Total dietary fiber values for desugared samples were adjusted for the percent sugar extracted, and total dietary fiber values for all samples were calculated on a dry weight basis (Kays et al., 1996).

Spectroscopic Analysis. All dry-milled cereal samples and mono- and disaccharides were scanned with the NIRSystems 6500 monochromator. High-sugar samples were scanned

before desugaring. Duplicates of each sample were presented in cylindrical sample cells (internal diameter = 38 mm, depth = 9 mm) with optical quartz surface and cardboard backing. Each sample was scanned 16 times, and the data were averaged and transformed to $\log_{10}(1/R)$. The duplicate scans of each sample were examined visually for consistency and averaged.

Spectra of sucrose and selected cereal samples were collected in the mid-infrared range (4000−400 cm^{-1}) using the Nicolet 850 FT spectrometer (Nicolet Instruments Corp., Madison, WI). Samples were diluted in powdered KBr (8% sample), to limit absorbance to approximately 0.8 absorbance units, and presented in open surface cylindrical metal cups (13 mm internal diameter, 2 mm depth). Data were averaged over 256 scans at 4 cm^{-1} resolution after the samples were allowed to stand for 10 min in a purged-air environment. Sample spectra were Fourier transformed, ratioed against a KBr background, and presented in the absorbance mode.

NIR Calibration on the Original Data Set. The original calibration was developed using a method similar to that previously described (Kays et al., 1996). Seventy-seven cereal samples were scanned with the NIRSystems 6500 monochromator and analyzed for dietary fiber using the reference method. The wavelength region used for NIR analysis was 1100−2500 nm with 2 nm intervals. A commercial spectral analysis program (NIRS3, Version 4.01, Infrasoft International Inc., Port Matilda, PA) was used to process the data and develop chemometric models. First, $\log_{10}(1/R)$ spectra were transformed with standard normal variate and detrending procedures (Barnes et al., 1989), to remove multiplicative interferences of scatter, and then transformed with second-derivative processing (gap = 20 nm, smoothing interval = 10 nm). Data were subsequently centered using the CENTER program, available via NIRS3, which allows centering of samples based on constituent values as well as spectral characteristics, i.e. partial least squares-1 (PLS1; Lindberg et al., 1983). Prior to calibration, $\log_{10}(1/R)$ spectra were mean centered, transformed with standard normal variate and detrending procedures (Barnes et al., 1989) to remove multiplicative interferences of scatter, and then transformed with second-derivative processing (gap = 8 nm, smoothing interval = 8 nm). Calibration was performed using modified PLS regression available through NIRS3. The modification to PLS scaled the reference method data and reflectance data at each wavelength to have a standard deviation of 1.0 before each PLS regression term (Shenk and Westerhaus, 1991a). The optimum number of PLS factors used for total dietary fiber prediction was determined by cross-validation (Martens and Naes, 1989). During cross-validation one-sixth of the calibration samples at a time was temporarily removed from the calibration set and used for prediction. Performance statistics were accumulated for each group of removed samples. The optimal number of factors for total dietary fiber was that which produced a minimum in overall error between modeled and reference values (SECV). The preprocessing transformations used were the optimum required to improve the SECV compared to PLS analysis with untransformed data.

Upon completion of the calibration, the model was validated using an independent set of 30 cereal samples. Model performance was reported as the coefficient of determination (r^2), the SEP, the slope, and the average difference between measured and modeled values (bias) (Hruschka, 1987).

NIR Calibration on the Sugar-Expanded Data Set. To expand the original calibration, which contained products with low and medium amounts of sugar, to include products with high sugar and crystalline sugar content, 39 cereal products containing >20% sugar were purchased from grocery retailers. Using an algorithm called SELECT (NIRS3; Shenk and Westerhaus, 1991a,b) high-sugar samples from the group of 39 were selected for calibration expansion. The SELECT algorithm identifies samples within and outside the neighborhoods previously defined by the original calibration. Eleven PLS1 components were used by SELECT and, with the scores in 11-dimensional space, the neighborhood H distance was calculated between all spectral pairs in the original model and each of the high-sugar samples. A neighborhood H value of 0.6 was used to define the neighborhoods. Any high-sugar sample whose neighborhood H value was <0.6 H from any sample in the original model was eliminated, as neighbors of the central sample are considered to be spectrally similar and therefore not needed for calibration expansion. This process was performed with all of the high-sugar samples in the pool, until every sample was in either the calibration update or the eliminated set. Twenty-five high-sugar samples were chosen by the SELECT algorithm, out of the 39 available, for calibration expansion. Two of the 25 selected samples (Fudge Brownie Mix and Blueberry Muffin Mix) were discarded as spectral outliers (Mahalanobis distance >3), based on PLS analysis. Rescanned samples were still recognized as outliers. Twenty-three high-sugar samples were, thus, combined with the original 77 calibration samples to generate a 100-sample, sugar-expanded calibration data set. Log_{10} I/R spectra were transformed and centered as described for the original data set, and a sugar-expanded calibration model was developed using modified PLS (Shenk and Westerhaus, 1991a) with the same preprocessing spectra transformations used for the original calibration.

The sugar-expanded model was tested using the original 30 independent validation samples alone and the original 30 independent validation samples plus 15 independent high-sugar samples. Model performance was reported as the SEP, r^2, slope, and bias.

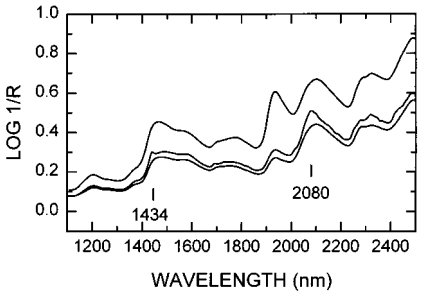

Figure 1. NIR spectra of Corn Flakes (7% sugar, upper plot), Frosted Flakes (43% sugar, middle plot), and Golden Crisp cereal (56% sugar, lower plot).

RESULTS

Total Dietary Fiber Measured by the Reference Method. The range for total dietary fiber in all cereal samples, determined by the AOAC enzymatic−gravimetric procedure, was from 0.6 to 52.1% (N = 161). For the high-sugar samples the ranges in total dietary fiber for the calibration and validation data sets were from 0.6 to 14.9% and from 1.0 to 10.6%, respectively. The distributions of samples for each grain type in the calibration and validation data sets are given in Tables 1 and 2. The range, mean, and standard deviation of total dietary fiber percent for each grain type are also presented. The standard error of the AOAC laboratory determinations (Windham et al., 1989) was 0.73% for the samples in the original calibration and validation data sets and 0.88% for the high-sugar samples.

Spectral Characteristics of Samples. Examples of NIR spectra of cereal samples with various amounts of sugar are given in Figure 1. The upper plot is the spectrum of Corn Flakes cereal, which contains approximately 7% sugar, present as sucrose and high fructose corn syrup (predominantly fructose). The middle plot is the spectrum of Frosted Flakes cereal with 43% sugar, present as sucrose and corn syrup (predominantly glucose). The lower plot is the spectrum of Golden Crisp cereal with 56% sugar present as sucrose, corn syrup, and honey (a mixture of sucrose, glucose, and fructose). The spectrum of Frosted Flakes is typical of cereal products containing large amounts

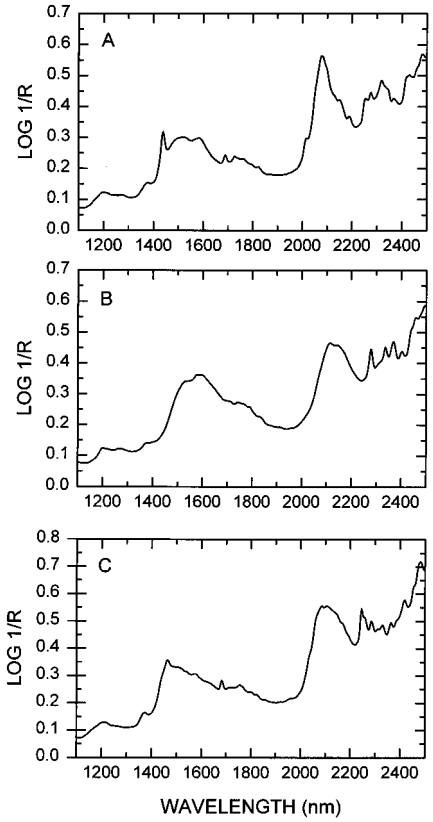

Figure 2. NIR spectra of sucrose (A), glucose (B), and fructose (C).

of crystalline sugar (Murray and Williams, 1987) with a sharp peak at 1434 nm and another sharp peak at 2080 nm. Seventeen of the 23 high-sugar samples selected for calibration and 8 of the 15 high-sugar validation samples showed evidence of crystalline sugar in the near-infrared spectra. Although Golden Crisp cereal contains a large amount of sugar, it does not appear to contain crystalline sugar (Figure 1). Sucrose, D-(+)-glucose, and D-(−)-fructose are the most common sugars found in cereal products and have unique spectra as demonstrated in Figure 2, panels A, B, and C, respectively. Sucrose has particularly sharp absorbances at 1434 and 2070 nm. The monomeric sugars, glucose and fructose, do not exhibit the same characteristics as sucrose at 1434 and 2070 nm. However, they mimic sucrose in other areas of the spectrum. The fundamental O−H and C−H bands that are responsible for absorbances at 1434 and 2070 nm in the near-infrared spectrum for sucrose occur in the 4000−2000 cm^{-1} portion of the mid-infrared region and are shown in Figure 3 (solid line). Figure 3 also shows the difference spectrum for Frosted Flakes minus Corn Flakes (Figure 3, dashed line) and Honey Crunch Corn Flakes minus Corn Flakes (Figure 3, broken line) in the same portion of the mid-infrared spectrum. The difference in the spectrum of the former reflects a difference in the sucrose amount, which is greater in Frosted Flakes. In the latter, a difference can be detected in the amount of sucrose and in a long chain-hydrocarbon (2930 and 2854 cm^{-1}), which may be lipid (Barton and Himmelsbach, 1992). The additivity of the spectra in Figure 3 would indicate that the sugar frosting is not part of the matrix but merely mixed with it. The requirement for a single new factor to account for added sugar (see Figures 4 and 5 and discussion) also suggests that sugar is present in the matrix of many samples but not part of it.

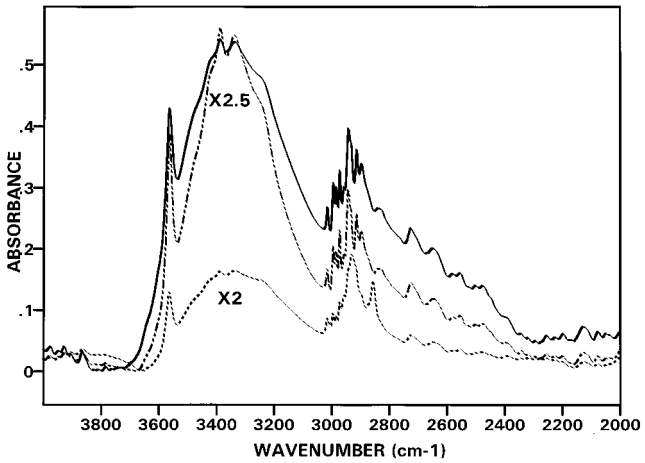

Figure 3. Mid-IR spectrum of sucrose (solid line), and mid-IR difference spectra of Frosted Flakes minus Corn Flakes (ratio 1:1, dashed line) and Honey Crunch Corn Flakes minus Corn Flakes (ratio 1:1, broken line). The difference spectra are multiplied by 2.5 and 2, respectively.

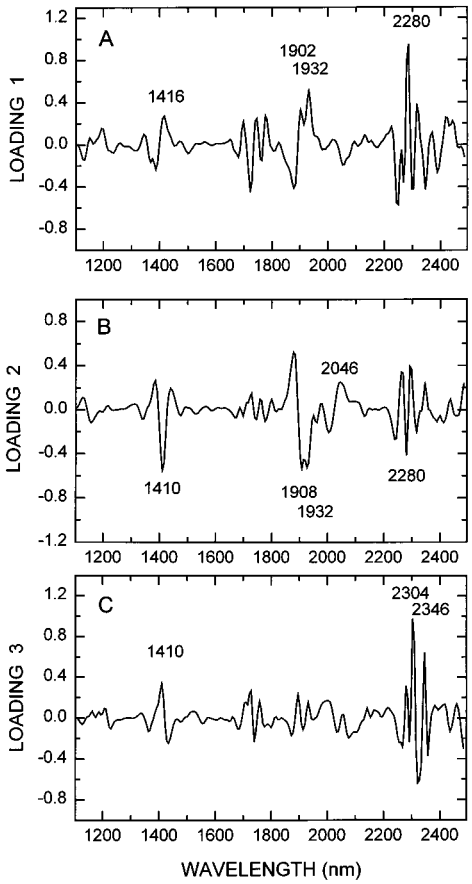

Figure 4. PLS loading spectra for total dietary fiber in cereal products in the original model. Panels A, B, and C represent loadings for factors 1, 2, and 3, respectively.

Original NIR Calibration Model for Total Dietary Fiber. As reported previously (Kays et al., 1996) a NIR calibration was obtained, using PLS, for determination of the concentration of total dietary fiber in cereal products. This model was designated the "original" calibration in the present study and contained products with <20% sugar and <10% fat. For the original calibration the overall error between modeled and reference values (SECV), using six cross-validation groups, was 1.64%, with R^2 of 0.99 (Table 3). Independent validation samples were predicted using the original calibration model. When NIR-predicted values for

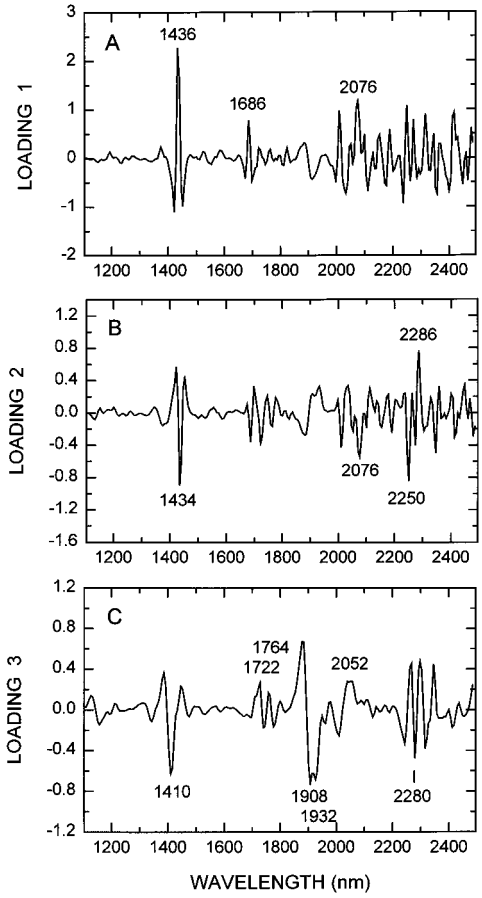

Figure 5. PLS loading spectra for total dietary fiber in cereal products in the sugar-expanded model. Panels A, B, and C represent loadings for factors 1, 2, and 3, respectively.

dietary fiber were compared statistically with AOAC-determined values, the SEP was 1.39% and r^2 0.99 (Table 3). The SEP was lower than the SECV, possibly because the range in total dietary fiber values for the validation data set falls within the range of values for the calibration data set. Linear regression of AOAC determined dietary fiber against NIR-predicted dietary fiber for the calibration data ($Y = 0.07 + 1.00X$) and the validation data ($Y = -0.18 + 1.02X$) gave intercepts and slopes not significantly different from 0.0 and 1.0, respectively ($p > 0.05$).

Sugar-Expanded NIR Calibration Model for Total Dietary Fiber. The sugar-expanded calibration data set contained products that spanned the range of high (>20%), medium, and low sugar content. Thus, a NIR calibration was obtained, using modified PLS, for determination of total dietary fiber in cereal products containing a wide range of sugar and crystalline sugar. Using six cross-validation groups, the SECV for the sugar-expanded calibration was 1.88% and R^2 was 0.98 (Table 3). The intercept and slope of the linear regression line (AOAC-determined dietary fiber versus NIR-predicted dietary fiber, $Y = -0.23 + 1.00X$), were not significantly different from 0.0 and 1.0, respectively ($p > 0.05$). Initially, 15 independent high-sugar validation samples were purchased. Fifteen high-sugar validation samples were combined with the original 30 validation samples and predicted with the sugar-expanded model. One validation sample (Low Fat Apple Cinnamon Muffin Mix) was discarded as a spectral outlier (Mahalanobis distance >3.0). The SEP was 1.58%, r^2 0.98, and bias −0.30. One sample (Cocoa Pebbles) was identified as a residual (t statistic) outlier (the sample was rescanned and reanalyzed for dietary fiber using the reference method and still recognized as an outlier). When this sample was removed, the SEP was 1.40%, r^2 0.99, and bias −0.40 (Table 3). As with the original calibration the SEP was lower than the SECV. The intercept and slope of the linear regression line (AOAC-determined versus NIR-predicted dietary fiber for the sugar-expanded validation data set, $Y = -0.64 + 1.02X$) were not significantly different from 0.0 and 1.0, respectively ($p > 0.05$). When the original model was used for prediction of total dietary fiber in the original 30 validation samples plus the 13 high-sugar validation samples, the SEP, r^2, bias, and slope were 4.22%, 0.91, 1.21, and 0.87, respectively. The large SEP and bias are primarily due to markedly underpredicted values for high-sugar products that exhibit spectral evidence of crystalline sugar. A comparison of AOAC-determined values versus NIR-predicted values, using both models, for the individual high-sugar validation samples is presented in Table 4. The sugar-expanded model was used to predict the original validation samples ($N = 30$) with a resulting SEP, r^2, bias, and slope of 1.42%, 0.99, −0.66, and 1.04, respectively.

PLS Loadings. In the original calibration nine factors were employed in the model and explained 99.1% of the spectral variation. Sample scores having the highest correlation (calculated by Pearson correlation coefficient) with dietary fiber were for factors 1, 2, and 3 with correlation coefficients of 0.78, 0.56, and 0.19, respectively.

PLS loadings, which are the regression coefficients of each wavelength to dietary fiber for each factor, can indicate wavelengths of high variation in a calibration set which may show association with wavelengths of chemical origin known *a priori*. The two calibration sets in the present study appear to have loadings of high relative positive or negative value at wavelengths that are known for carbohydrates, water, or lipids. Association of these known vibrations with the appearance of key values in loadings, along with the correlation of individual loadings independently to fiber or sugar content, lends an estimation of what chemical component is the primary contributor for each loading. Loading plots for factors 1−3 of the original model are presented in Figure 4. Factor 1 had the highest correlation with dietary fiber, and its loading has peaks correlated to O−H absorption in water bands at 1416 and 1902/1932 nm and C−H absorption in the carbohydrate band at 2280 nm (Figure 4A; Murray and Williams, 1987). The loading for factor 2 has regression coefficients related to absorption by O−H in the water bands at 1410 and 1908/1932 nm, C−H groups of carbohydrate at 2280 nm, and amide in the protein band at 2046 nm (Figure 4B). The loading for the third factor exhibited effects associated with C−H stretch groups in lipid at 2304 and 2346 nm (Figure 4C). As reported previously (Williams et al., 1991; Kays et al., 1996; Windham et al., 1997), the model appears to be predominantly influenced by carbohydrate and water absorbance with smaller influences due to protein and lipid.

Eight factors were employed in the sugar-expanded model and explained 98.5% of the spectral variation. Sample scores having the highest correlation with dietary fiber were for factors 1, 2, and 3 with correlation coefficients of 0.48, 0.76, and 0.37, respectively. The loading for factor 2 showed strong correlation to absorbance at 1434 nm, typical of O−H groups in crystalline

Table 3. Calibration and Validation Statistics for Dietary Fiber Prediction by the Original and Sugar-Expanded NIR Models[a]

		calibration					validation[b]						
model	method	*n*	mean	SD	SECV	R^2	*n*	mean	SD	SEP	r^2	bias	slope
original	AOAC	77	15.83	13.70			30	15.50	13.53				
	NIRS	77	15.83	13.66	1.64	0.99	30	15.34	13.17	1.39	0.99	0.16	1.02
expanded	AOAC	100	13.74	12.73			43	12.31	12.36				
	NIRS	100	13.93	12.54	1.88	0.98	43	12.61	12.23	1.40	0.99	−0.40	1.01

[a] Mean, standard deviation (SD), standard error of cross-validation (SECV), and multiple coefficient of determination (R^2) for calibration. Mean, standard deviation, standard error of performance (SEP), and coefficient of determination (r^2) for validation. [b] Validation samples (n = 30) used to test the original model contained <20% sugar. Validation samples (n = 43) used to test the sugar-expanded model consisted of 30 validation samples containing <20% sugar plus 13 samples with >20% sugar.

Table 4. NIR Prediction of Dietary Fiber in High-Sugar Cereal Products

		NIR-predicted TDF%[a]	
product	AOAC TDF%	original model	sugar-expanded model
Fruit and Fiber	10.58	15.93	12.12
Healthy Choice	9.36	11.31	10.01
Golden Crisp	3.59	9.37[b]	6.15
Honey Grahams	4.80	4.52	4.36
Oatmeal Crunch	4.52	4.18	4.91
Blueberry Morning	3.38	3.89	3.71
Golden Grahams	3.54	3.33	4.43
Honey Nut Cheerios[c]	7.29	0.24	7.04
Nut and Honey Crunch[c]	2.45	−3.97	0.93
Apple Cinnamon Toasted Oats[c]	6.87	−4.00	5.79
Honey Crunch Corn Flakes[c]	1.84	−7.09	0.93
Oat Bran Muffin Mix[c]	4.65	−10.62	2.44
Double Dip Crunch[c]	1.01	−9.97	−0.77

[a] Total dietary fiber. [b] Global outlier (Mahalanobis distance >3). [c] Spectral evidence of crystalline sugar.

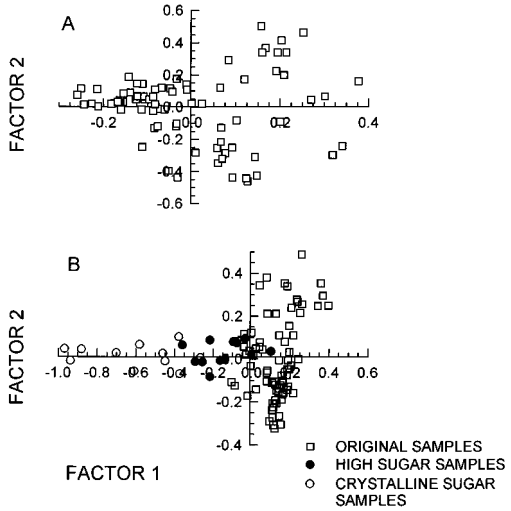

Figure 6. Plots of the PLS scores for factor 1 versus factor 2 for the original model (A) and the sugar-expanded model (B).

sugar, and to the 2200−2300 nm carbohydrate absorbing region of the NIR spectrum (Figure 5B). The loading for factor 1 also has a sharp peak at 1436 nm (Figure 5A) and peaks at 2076 and 2200−2300 nm correlated to carbohydrate (Figure 5A). The third factor of the sugar-expanded model had a loading plot with peaks at 1410 and 1908 nm, which are commonly associated with water, and at 2280 nm, associated with carbohydrate (Figure 5C). Overall, the sugar-expanded model appears to be predominantly influenced by carbohydrate with sharp peaks at 1434 nm and peaks at 2076 and 2200−2300 nm. Lesser influences due to water, lipid, and protein are indicated by the third loading.

PLS Scores. Plots of the factor 1 versus factor 2 scores for total dietary fiber are represented in Figure 6. In the original model, the contribution of dietary fiber (Pearson correlation coefficient = 0.78) is appreciable, resulting in the visualization of samples of low to high fiber primarily along the factor 1 axis (Figure 6A). For the sugar-expanded model (Figure 6B), the predominant contribution to factor 1 appears to be sugar content, based on product label values (Pearson correlation coefficient = −0.84), whereas the predominant contribution to factor 2 appears to be dietary fiber (Pearson correlation coefficient = 0.76). Crystalline sugar samples, in general, have the highest sugar content, and separation of these samples along the factor 1 axis is illustrated in Figure 6B.

DISCUSSION

The original calibration for prediction of total dietary fiber in cereal products has been expanded to include cereal products with high sugar and crystalline sugar content. The accuracy of total dietary fiber prediction, overall, for high-sugar samples was markedly improved by the sugar-expanded model compared to the original model. Products containing crystalline sucrose, for example, were greatly underpredicted by the original equation; however, using the sugar-expanded model, predicted values for total dietary fiber were acceptably close to the AOAC-determined values. The SECV is slightly higher for the sugar-expanded model compared to the original model. Thus, for samples containing <20% sugar, either model could be used. Overall, it is a distinct advantage to be able to predict a wide range of products, including high-sugar samples, with one calibration.

Sucrose, glucose, and fructose are the most commonly found sugars in cereal products. Many cereal products contain corn syrup, which is predominantly glucose, high fructose corn syrup,which is predominantly fructose with glucose making up the difference, and honey, which contains a mixture of sucrose, fructose, and glucose (Whitney and Rolfes, 1996). Each of the three sugars has unique NIR spectral characteristics, which are apparent in cereal products as sharp absorbance bands in specific areas of the product's NIR spectra. Correlations to these bands are observed in loadings 1 and 2 of the sugar-expanded model. Spectral properties of specific cereal products may depend on the concentration of free sugars or honey added to the product, in particular the ratios of crystalline sucrose to sucrose and the presence of glucose or fructose. The unique spectra of products containing large amounts of crystalline sugar demonstrate the necessity for developing an expanded equation to encompass such products. The rationale for an expanded equation for high-sugar products that do not exhibit spectral evidence of crystalline sugar is not as apparent. Some, though not all, high-sugar samples were predicted poorly by the origi-

nal model (one was a global outlier); however, the sugar-expanded model has corrected this deficiency.

In a previous study that investigated the use of near-infrared spectroscopy to predict neutral detergent fiber in ready-to-eat breakfast cereals (Baker, 1983), the author avoided the problem from the spectral uniqueness of high-sugar samples through the wavelengths chosen for the regression analysis. The wavelengths selected avoided interference by both sugar and fat.

In the present study 3 of the 59 high-sugar samples were identified as spectral outliers (2 from the calibration data set and 1 from the validation data set). All 3 samples displayed spectral evidence of crystalline sugar, and all 3 were mixes (1 brownie mix and 2 muffin mixes). Discarding the 3 samples eliminated the only muffin and brownie mixes remaining in the calibration data set. A substantially greater number of samples of this type may be required to expand the calibration to include these products.

The need for extraction of sugar before the AOAC procedure is performed increases the length of the assay from 2 days to 3 days and can also potentially decrease the accuracy of the assay. For near-infrared analysis, prior extraction of sugar is not required and sample preparation merely consists of grinding. The calibration obtained to associate AOAC total dietary fiber values with NIR spectra can be used for analysis of new prediction samples merely by obtaining NIR spectra of the products. Thus, use of the NIR calibration reduces the time required for total dietary fiber determination from 2 or 3 days to several minutes.

Examination of the loadings for the three factors most highly correlated with total dietary fiber suggests that the sugar-expanded model is predominantly influenced by O−H and C−H groups from the carbohydrate fractions with smaller influences from water, lipid, and protein. This is in contrast to the original model in which major influences appear to be from carbohydrate and water with minor influences from protein and lipid [in agreement with Williams et al. (1991), Kays et al. (1996), and Windham et al. (1997)]. It is important to note that the second loading of the original model is almost identical to loading 3 of the expanded model and that loading 1 of the original model is very similar in many respects to loading 2 of the expanded model. However, a new loading was introduced in the expanded model, loading 1, which appears to account for, or counteract, the effect of the free sugars present. The sharp peaks at 1434/1436 nm in loading 1 correlate with absorbance bands of O−H groups found in crystalline sugar. The importance of factor 1 is further demonstrated in the score plots. By interchanging the axes in Figure 6B it is observed that the pattern of the low-sugar samples is similar to the pattern in Figure 6A. The position of the high-sugar samples in Figure 6B suggests that the effect of factor 1 in the new model is to account for the presence of sugar.

CONCLUSIONS

A calibration equation for the prediction of total dietary fiber in cereal and grain products has been expanded to include cereal products with high sugar and crystalline sugar content. Use of the sugar-expanded NIR calibration decreases the time required to analyze high-sugar cereal samples for dietary fiber content from 3 days to several minutes. Examination of PLS loadings for the high-sugar model suggests that analytically useful absorbance is from effects related to O−H and C−H groups in the carbohydrate regions of the spectrum.

ACKNOWLEDGMENT

We thank Dr. Richard G. Leffler for valuable technical suggestions, Ms. Judy E. Davis and Mr. A. Terrell Boynton for technical assistance, and Ms. Lee Dexter of Canadian Harvest U.S.A. L.P. for providing some of the samples used to develop the calibration.

LITERATURE CITED

AOAC. Total Dietary Fiber in Foods: Enzymatic-gravimetric Method. *Official Methods of Analysis*, 15th ed.; AOAC: Arlington, VA, 1990.

AOAC. Total, Soluble, and Insoluble Dietary Fiber in Foods. *Official Methods of Analysis*, 15th ed., 3rd supplement; AOAC: Arlington, VA, 1992.

Baker, D. The determination of fiber in processed cereal foods by near-infrared reflectance spectroscopy. *Cereal Chem*. **1983**, *60*, 217−219.

Baker, D. The determination of fiber, starch, and total carbohydrate in snack foods by near-infrared reflectance spectroscopy. *Cereal Foods World* **1985**, *30*, 389−392.

Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard normal variate and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772−777.

Barton, F. E., II; Himmelsbach, D. S.; Duckworth, J. H.; Smith, M. J. Two dimensional vibration spectroscopy: correlation of mid- and near-infrared regions. *Appl. Spectrosc.* **1992**, *46*, 420−429.

Code of Federal Regulations **1995**, FDA, HHS; 21, Part 101.9,

Horvath, L.; Norris, K. H.; Horvath-Mosonyi, M.; Rigo, J.; Hegedus-Volgyesi, E. Study into determining dietary fiber of wheat bran by NIR-technique. *Acta Aliment.* **1984**, *13*, 355−382.

Hruschka, W. R. Data analysis: wavelength selection methods. In *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P. C., Norris, K. H., Eds.; American Assocication of Cereal Chemists: St. Paul, MN, 1987; pp 35−54.

Kays, S. E.; Windham, W. R.; Barton, F. E., II. Prediction of total dietary fiber in cereal products using near-infrared reflectance spectroscopy. *J. Agric. Food Chem*. **1996**, *44*, 2266−2271.

Kays, S. E.; Barton, F. E., II; Windham, W. R. NIR analysis of dietary fiber. In *Complex Carbohydrates: Definition, Analysis, and Applications*; Lee, S. C., Prosky, L., Eds.; Dekker: New York, 1997, in press.

Lee, S. C.; Prosky, L. International survey on dietary fiber: definition, analysis, and reference materials. *J. AOAC Int.* **1995**, *78*, 22−36.

Lee, S. C.; Prosky, L.; De Vries, J. W. Determination of total, soluble, and insoluble dietary fiber in foods—enzymatic-gravimetric method, MES-TRIS buffer: collaborative study. *J. AOAC Int.* **1992**, *75*, 395−416.

Lindberg, W.; Persson, J.-A.; Wold, S. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. *Anal. Chem.* **1983**, *55*, 643−648.

Martens, H.; Naes, T. Assessment, validation and choice of calibration method. In *Multivariate Calibration*; Wiley: New York, 1989; pp 237−266.

Murray, I.; Williams, P. C. Chemical principles of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P. C., Norris, K. H., Eds.; American Association of Cereal Chemists: St. Paul, MN, 1987; pp 17−34.

Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*; Longman Scientific and Technical: Harlow, England, 1993.

Shenk, J. S.; Westerhaus, M. O. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci*. **1991a**, *31*, 469−474.

Shenk, J. S.; Westerhaus, M. O. Population structuring of near infrared spectra and modified partial least squares regression. *Crop Sci*. **1991b**, *31*, 1548−1555.

Trowell, H.; Southgate, D. A. T.; Wolever, T. M. S.; Leeds, A. R.; Gassull, M. A.; Jenkins, D. J. A. Dietary fiber redefined. *Lancet* **1976**, 967.

Whitney, E. N.; Rolfes, S. R. *Understanding Nutrition*, 7th ed.; West Publishing: St. Paul, MN, 1996; p 132.

Williams, P. C.; Norris, K. H. *Near-Infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists: St. Paul, MN, 1987.

Williams, P. C.; Cordeiro, H. M.; Harnden, M. F. T. Analysis of oat bran products by near-infrared reflectance spectroscopy. *Cereal Foods World* **1991**, *36*, 571−574.

Windham, W. R.; Mertens, D. R.; Barton, F. E., II. Protocol for NIRS calibration: sample selection and equation development and validation. In *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*; Marten, G. C., Shenk, J. S., Barton, F. E., II, Eds.; U.S.D.A. Agriculture Handbook 643; U.S. GPO: Washington, DC, 1989; pp 96−103.

Windham, W. R.; Kays, S. E.; Barton, F. E., II. Effect of cereal product residual moisture content on total dietary fiber determined by near-infrared reflectance spectroscopy. *J. Agric. Food Chem*. **1997**, *45*, 140−144.

Received for review April 22, 1997. Revised manuscript received July 17, 1997. Accepted July 17, 1997.[⊗] Reference to company or trade names is for the purpose of description only and does not imply endorsement by the U.S. Department of Agriculture.

JF9703260

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1997.