JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Use of Near-Infrared Spectroscopy for Screening the Individual and Total Glucosinolate Contents in Indian Mustard Seed (*Brassica juncea* L. Czern. & Coss.)

RAFAEL FONT, MERCEDES DEL RÍO, JOSÉ M. FERNÁNDEZ-MARTÍNEZ, AND ANTONIO DE HARO-BAILÓN*

Department of Agronomy and Plant Breeding, Institute of Sustainable Agriculture (CSIC), Alameda del Obispo s/n, 14080 Córdoba, Spain

The potential of near-infrared spectroscopy (NIRS) for screening the sinigrin, gluconapin, 4-hydroxyglucobrassicin, and total glucosinolate contents of Indian mustard (*Brassica juncea* L. Czern. & Coss.) seed was assessed. Intact seed samples of this species were analyzed by NIRS and their reference values regressed against different spectral transformations by modified partial least-squares (MPLS) regression. The coefficients of determination (r^2) for sinigrin, gluconapin, 4-hydroxyglucobrassicin, and total glucosinolate contents were, respectively, 0.86, 0.95, 0.33, and 0.82. The standard deviation to standard error of prediction (SEP) ratio, and SEP to standard error of laboratory ratio were for these constituents as follows: sinigrin, 2.59 and 2.70; gluconapin, 4.16 and 2.08; 4-hydroxyglucobrassicin, 1.18 and 1.40; and total glucosinolates, 2.18 and 1.60. By comparison of commercial sinigrin spectrum with the first MPLS loadings of the sinigrin equation, it can be concluded that the molecule of sinigrin has a specific signal in the seed spectrum of *Brassica*.

KEYWORDS: Near-infrared spectroscopy; screening; *Brassica juncea*; glucosinolates; agriculture; biofumigants

INTRODUCTION

Glucosinolates (β-thioglucoside-*N*-hydroxysulfates) are naturally occurring thioglucosides that are characteristic of the Cruciferae (including the genus *Brassica*) and related families in the order Capparales. In general, glucosinolates conform to the basic structure shown in **Figure 1a**. The structural diversity of this large group of compounds is due almost entirely to the different substituents possible at the side-chain position R. The R substituent may be an alkyl or alkenyl side chain, which itself may contain substituents of hydroxyl groups or sulfur. Alternatively, the R substituent may be an aromatic or a heteroaromatic group. Glucosinolates, and specifically their hydrolysis products, are the compounds responsible for many of the beneficial and harmful properties of glucosinolate-containing plants. Among the beneficial characteristics of glucosinolates are their antibacterial and antifungal properties (*1*) and their cancer-chemoprevention activity (*2, 3*). In addition to the above-mentioned attributes, glucosinolates are also the molecules responsible for the pungent and hot flavors characteristics of the seed of some *Brassica* crops, which are highly valued in mustard spices. However, the toxic and antinutritive effects of glucosinolates have limited the use of seed meals from *Brassica* oilseeds for human food and animal feed (*4*). These negative effects of glucosinolates have been the basis for research targeting low glucosinolate contents in some *Brassica* crops (*5*).

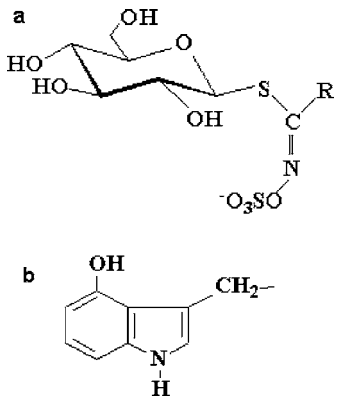

Figure 1. (**a**) General structure of glucosinolates. (**b**) Structure of the side chain R of 4-hydroxyglucobrassicin.

Chemical analysis of *Brassica* crops for determining the glucosinolate content is expensive and time-consuming. The high cost and labor input required to obtain the total glucosinolate content in the seed samples by the standard methods of analysis, such as high-performance liquid chromatography (HPLC) and gas−liquid chromatography, are serious handicaps to the analysis of large sets of samples, which is usually necessary to identify the target genotypes. The use of fast analytical techniques such as near-infrared spectroscopy (NIRS)

* Corresponding author (telephone +34 957 499247; fax +34 957 499252; e-mail deharo@cica.es).

10.1021/jf0307649 CCC: $27.50 © 2004 American Chemical Society
Published on Web 05/05/2004

results in many advantages because analysis can be made with a considerable saving of time, at a low cost, and without using hazardous chemicals. In addition, samples can be analyzed in their natural form without destruction, which is very useful in the case of scarce and valuable seed material.

Most authors using NIRS for determining the glucosinolate content of *Brassica* have focused their efforts on the species *B. napus* and its "double zero" lines (*6*−*11*), known as "canola", because of the commercial interest of this species. In contrast, those works using NIRS for predicting the glucosinolate content of Indian mustard (*Brassica juncea* L. Czern. & Coss.) are scarce (*12*, *13*).

B. juncea shows a large variability in its individual glucosinolate pattern depending on the geographical origin of the plant. Whereas *B. juncea* genotypes of European or North American origin contain from 150 to 200 μmol/g of sinigrin (R is CH_2═CH−CH_2−) (oil-extracted, air-dried meal), genotypes from the Indian subcontinent contain variable amounts of sinigrin and gluconapin (R is CH_2═CH−CH_2−CH_2−) (*14*). In addition, other minor glucosinolates are also present in the seed, such as 4-hydroxyglucobrassicin (**Figure 1b**). The glucosinolate pattern exhibited by *B. juncea* makes it one of the most promising species as a potential source of variability for glucosinolates, to be used in the fields of agriculture and medicine. In addition, its tolerance to drought permits this species to be used as an oilseed crop in Mediterranean conditions (*15*).

This study focuses on testing the potential of NIRS for determining the concentration of sinigrin, gluconapin, 4-hydroxyglucobrassicin, and total glucosinolates in the intact seed of multiple accessions of *B. juncea* from the Indian and European areas and providing some knowledge about the mechanism used by NIRS for predicting the concentration of these compounds in the seed of *B. juncea*.

MATERIALS AND METHODS

Samples. This work was conducted with 2700 seed samples (individual plants) of *B. juncea* from different geographical origins (Europe and the Indian subcontinent). The seed material, held in the Department of Agronomy and Plant Breeding (IAS, CSIC, Córdoba, Spain), was multiplied in the years 1997, 1998, and 1999 in Córdoba, Spain. Plants were grown in a typical xerofluvent soil (pH 8) (*16*), in conditions of self-pollination. Mature plants were harvested individually, and their seeds were collected in paper bags until analysis.

NIRS Analysis. Seed samples were analyzed by NIRS in an NIRSystems model 6500 spectrophotometer (Foss-NIRSystems, Inc., Silver Spring, MD) equipped with a transport module, in the reflectance mode. Intact seed samples were placed in the NIRS sample holder (3 cm diameter round cell) until it was three-fourths full, and their spectra were registered as an individual file, in the range from 400 to 2500 nm, at 2 nm intervals.

The spectrum file was then checked for spectral outliers by using principal component analysis (PCA). Spectral outliers were identified by calculating the Mahalanobis distance (*H*) of each sample spectrum to the mean spectrum of the sample population. A total of six samples showed a standardized distance from the mean (*H*) >3. After visual inspection, it was decided to leave them out of the set for NIRS analysis because of their abnormal appearance (non-well-rounded seed).

Also, on the basis of the *H* statistic, 208 samples were selected for analysis as being most relevant and most representative of those in the entire population (ISI SELECT algorithm) by their spectral features. The selected samples were tested for bulk density by applying an NIR equation previously developed (*17*), to ensure the viability of all the seed samples used in this work. The selected seed samples were then analyzed by HPLC, and their sinigrin, gluconapin, 4-hydroxyglucobrassicin, and total glucosinolate contents were determined.

Spectra in the new file (n = 208) were ordered by the sinigrin reference values of the samples, from the lowest to the highest content, and the spectra were assigned to the calibration and validation sets in a ratio of 2:1, respectively.

Using the program Global v. 1.50 (WINISI II, Infrasoft International, LLC, Port Matilda, PA), different calibration equations for sinigrin, gluconapin, 4-hydroxyglucobrassicin, and total glucosinolates were developed on the calibration set (n = 139). Calibration equations were computed using the raw optical data (log 1/*R*, where *R* is reflectance), or first or second derivatives of the log 1/*R* data, with several combinations of segment (smoothing) and derivative (gap) sizes [i.e., (0, 0, 1, 1; derivative order, segment of the derivative, first smooth, second smooth); (1, 4, 4, 1); (2, 5, 5, 2)] (*18*, *19*). To correlate the spectroscopic information (raw optical data or derivative spectra) of the samples and the content of the parameters being studied as they were determined by the reference method, modified partial least-squares (MPLS) was used as regression method (*20*), by using wavelengths from 400 to 2500 nm every 8 nm. In addition, standard normal variate and detrending transformations (SNV-DT) (*21*) were used to correct baseline offset due to scattering effects from differences in particle size among samples.

Cross-validation was performed on the calibration set for determining the optimum number of terms to be used in the calibration equation. The prediction error shown by the different equations in the cross-validation was computed as the standard error of cross-validation (SECV). This statistic is also used for computing the optimum number of MPLS terms to be supported in the calibration model (*22*). Additional discussion on the SECV statistic has been reported by Shenk and Westerhaus (*22*).

Cross-validation was also used to identify those samples being chemical (*t*) or spectral (*H*) outliers. *t* outliers are samples that have a relationship between their reference values and spectra that is different from the relationship of the other samples in the set and with large residuals (*t* values >2.5). Samples with a large *t* statistic often put doubt on the reference chemistry value. An *H* outlier identifies a sample that is spectroscopically different from other samples in the population and has a standardized *H* value >3.0.

Calibration equations were validated with five cross-validation groups, and those samples identified as outliers were removed from the calibration file in two elimination passes (*23*). The different calibration equations obtained in the calibration process were then validated on an external validation set (n = 69), formed with samples not included in the calibration set. The prediction ability of each of the calibration equations obtained was determined on the basis of its coefficient of determination in the external validation (r^2), ratio of the standard deviation of the external validation set to standard error of prediction (RPD), ratio of the range to standard error of prediction (RER) (*24*), and ratio of the standard error of prediction (SEP) to standard error of laboratory (SEL). On the other hand, the bias (mean of the reference values minus the mean of predicted values by NIRS) and slope were considered as an additional criteria for determining the magnitude and direction in which the NIR predictions deviate from reference data for the diverse equations.

For the calculation of the SEL, which is a measure of the reproducibility of the reference method, a subset (n = 50) of the samples used in the study were selected by their total glucosinolate content through the entire range and analyzed in duplicate over different times, analysts, and laboratories (IAS, Córdoba, Spain; University of Gembloux, Gembloux, Belgium; and CETIOM, Grignon, France). Once the SEL was determined, its value was compared with the SEP of each equation to test the precision of the different predictive models in relation to the error of the reference method.

HPLC Analysis. About 100 mg of seeds from the selected samples previously scanned by NIRS was ground in a Janke and Kunkel model A10 mill (IKA-Labortechnik) for ∼20 s, and a two-step glucosinolate extraction was carried out in a water bath at 75 °C to inactivate myrosinase. In the first step the flour was heated for 15 min in 2.5 mL of 70% aqueous methanol and 200 μL of 10 mM glucotropaeolin as an internal standard (*25*). A second extraction was applied after centrifugation (5 min, $5 \times 10^3 g$) by using 2 mL of 70% aqueous methanol. One milliliter of the combined glucosinolate extracts was pipetted onto the top of an ion-exchange column containing 1 mL of Sephadex DEAE-A25 in the formate form. Desulfation was carried out

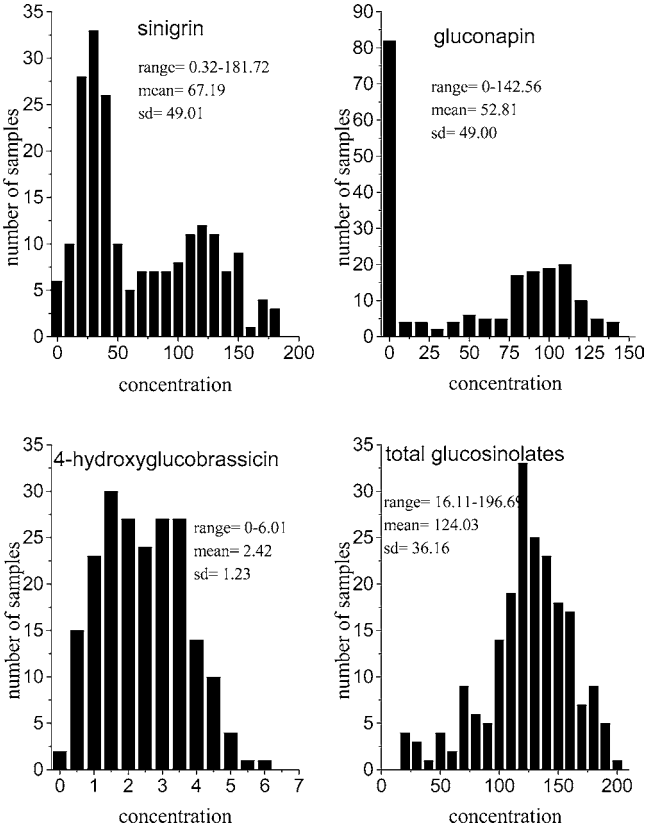

Figure 2. Distribution plots for sinigrin, gluconapin, 4-hydroxyglucobrassicin, and total glucosinolates in the whole set of samples ($n = 208$). Values are expressed as micromoles per gram of dry weight.

by the addition of 75 μL of purified sulfatase (EC 3.1.6.1, type H-1 from *Helix pomatia*) (Sigma) solution. Desulfated glucosinolates were eluted with 2.5 mL (0.5 mL × 5) of Milli-Q (Millipore) ultrapure water and analyzed with a model 600 HPLC instrument (Waters) equipped with a model 486 UV tunable absorbance detector (Waters) at a wavelength of 229 nm. Separation was carried out by using a Lichrospher 100 RP-18 in Lichrocart 125-4 column, 5 μm particle size (Merck). The amount of each individual glucosinolate present in the sample was calculated by means of the internal standard and expressed as micromoles per gram of seed dry weight (dw). The total glucosinolate content was computed as the sum of all the individual glucosinolates present in the sample.

MPLS Loading Plots. The MPLS loading plots of the first three factors generated from the MPLS regression performed on the second-derivative transformation of the raw optical data (2, 5, 5, 2; SNV + DT) were calculated. MPLS regression constructs its factors by capturing as much of the variation in the spectroscopic data as possible by using the reference values actively during the decomposition of the spectroscopic data. By balancing the spectroscopic and chemical information, the method reduces the impact of large, but irrelevant, spectroscopic variations in the calibration modeling (*26*).

The loading plots show the regression coefficients of each wavelength to the parameter being calibrated for each factor of the equation. Wavelengths represented in the loading plots as more highly participating in the development of each factor are those of greater variation and are better correlated to the parameter in the calibration set.

RESULTS

Individual and Total Glucosinolate Frequency Distributions for the Selected Samples. Figure 2 shows the distribution plots of the individual and total glucosinolate contents of the samples ($n = 208$) used in this work. The glucosinolate composition of *B. juncea* showed two patterns, characterized by high sinigrin and high gluconapin concentrations. Many samples showed gluconapin concentrations close to 0, and therefore, the total glucosinolate content was mainly determined by its sinigrin concentration. These results are in agreement with previous studies on this species (*27*, *28*).

Three constituents (sinigrin, gluconapin, and total glucosinolates) showed wide ranges, whereas 4-hydroxyglucobrassicin exhibited a narrow range. Glucosinolate concentrations shown by the samples used in this work are those found in naturally occurring and cultivated varieties of this species (*29*, *30*).

NIRS Analysis. *Sinigrin.* Sinigrin was better modeled by the second-derivative transformation (2,5,5,2; SNV + DT) equation (SDTE) of the raw 1/log R optical data than by any other equation. SDTE showed a higher coefficient of determination in the calibration ($R^2 = 0.93$) and a lower standard error of calibration (SEC = 12.90 μmol/g of dw) than any other mathematical treatment used in this work. R^2 and SEC values indicate the best theoretical accuracy obtainable for a set of variables used to develop a calibration (*31*).

Two samples were identified as being chemical outliers in cross-validation, and seven terms were selected as the optimum number to fit the final model. After visual inspection of these two seed samples identified as chemical outliers, it was decided definitely to eliminate them from the calibration set because of their high chlorophyll content (green appearance). The ratio of the SD to SECV exhibited by the SDTE was the highest (2.74) of the three equations. The coefficient of determination in the cross-validation (1-VR) was high (0.86) for the SDTE and slightly greater than that of the first-derivative transformation equation (FDTE) (0.83). Although the FDTE showed a prediction ability in cross-validation similar to that of the SDTE, it was modeled by using a higher number of terms to fit the final model.

Of the three equations, SDTE showed the highest prediction ability (**Table 1**). The r^2 shown by this equation (0.86) was high, meaning that 86% of the sinigrin variability contained in the seed samples of the validation set was explained by the model (**Figure 3A**). The RPD and RER values shown by the SDTE were the highest of the three equations and were close to those values recommended for screening purposes (*24*). The bias and slope corresponding to this equation were the closest to 0 and 1, respectively, of the three equations.

To evaluate the prediction ability of the equations in relation to the overall error of the reference method, the SEL was calculated and related to SEP. Sinigrin determination by the reference method showed an SEL of 6.94 μmol/g of dw, and the corresponding SEP/SEL ratio for the SDTE was 2.70.

Gluconapin. The various calibration equations developed to predict the gluconapin content showed performances in calibration and cross-validation that were similar to those exhibited by the different sinigrin equations. Two samples were identified as being chemical outliers in the cross-validation (these samples matched those outliers for the sinigrin equation) and were eliminated from the calibration set. As previously occurred for sinigrin, the SDTE for gluconapin showed higher prediction ability in cross-validation than any of the other mathematical treatments. High R^2 (0.95) and 1-VR (0.89) values were shown by the SDTE and, also, the SD/SECV (2.99), which was the highest of all the equations for any of the parameters studied. The three equations were modeled by using 10 terms as the optimum number from cross-validation.

The r^2 shown by the SDTE in the external validation was high (0.95) (**Table 1**; **Figure 3B**). In addition, the SDTE showed RPD and RER values that were over the minimum value

Table 1. External Validation Statistics for Sinigrin, Gluconapin, 4-Hydroxyglucobrassicin, and Total Glucosinolates (Micromoles per Gram of Dry Weight) NIR Predictive Equations in the Intact Seed of *B. juncea* ($n = 69$)

component	range[a]	mean[b]	SD[c]	SEP[d]	r^{2e}	RPD[f]	RER[g]
sinigrin	4.70–178.25	68.51	48.58	18.74	0.86	2.59	9.26
gluconapin	0.51–142.56	54.92	48.89	11.74	0.95	4.16	12.09
4-hydroxyglucobrassicin	0.30–5.55	2.56	1.16	0.98	0.33	1.18	5.35
total glucosinolates	21.89–187.29	127.62	34.15	15.65	0.82	2.18	10.56

[a] Range of the reference data in the external validation set. [b] Mean value of the reference data in the external validation set. [c] Standard deviation of the reference data in the external validation set. [d] Standard error of the prediction. [e] Coefficient of determination in the external validation. [f] Ratio of the standard deviation to SEP. [g] Ratio of the range to SEP.

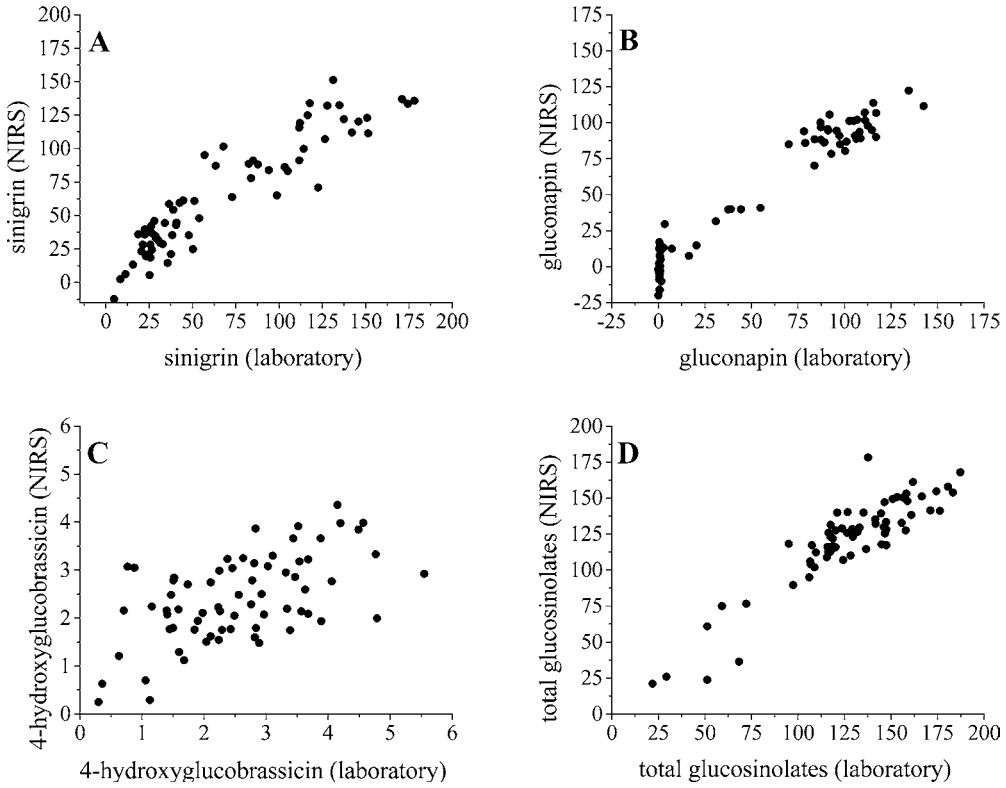

Figure 3. External validation scatter plots for sinigrin (**A**), gluconapin (**B**), 4-hydroxyglucobrassicin (**C**), and total glucosinolates (**D**). Values are expressed as micromoles per gram of dry weight.

recommended for screening of seed material (*24*). The SDTE for gluconapin showed bias and slope similar to those exhibited by sinigrin, although the bias was slightly higher than those shown by the other equations. The SEL and SEP/SEL ratio shown by the SDTE for gluconapin were 5.63 μmol/g of dw and 2.08, respectively.

4-Hydroxyglucobrassicin. High SECs (from 0.92 to 1.00 μmol/g of dw) and low R^2 values (from 0.36 to 0.45) were shown by the equations for 4-hydroxyglucobrassicin in calibration. Similar prediction abilities were shown by the three equations in cross-validation, which were modeled with three and four terms, and showed the lowest prediction abilities of all the glucosinolate equations, on the basis of the SD/SECV ratio. A low coefficient of determination in cross-validation (0.33) was also exhibited by these equations.

The external validation was conducted with the equations developed in the calibration process. The r^2, RPDs, and RERs exhibited by all of the equations were low (**Table 1**; **Figure 3C**), the diverse equations showing similar performances. Slopes for the three equations were far from 1, predicted values for 4-hydroxyglucobrassicin being underestimated (positive biases).

The SEL and the SEP/SEL ratio for the SDTE for 4-hydroxyglucobrassicin were 0.70 μmol/g of dw, and 1.4, respectively, the latter being the lowest value of the different equations obtained for any parameter.

Total Glucosinolates. FDTE and SDTE showed almost identical SECs (17.00 and 17.36 μmol/g of dw, respectively) and R^2 values (0.79 and 0.78, respectively) in calibration and also 1-VR (0.70) and SD/SECV (1.84 and 1.82, respectively) in cross-validation. FDTE was modeled with seven terms, whereas five terms were selected in cross-validation as the optimum number to fit the second-derivative equation. Three samples were identified in cross-validation as being chemical outliers, two of them matching those outliers of the sinigrin and gluconapin equations.

In the external validation, the FDTE showed the lowest SEP (15.65 μmol/g of dw) and, therefore, the higher RPD (2.18) and RER (10.56) of the three equations (**Table 1**). The r^2 showed by the FDTE was high (0.82) (**Figure 3D**). The various predictive models for total glucosinolates showed similar slopes.

The SEL and its respective SEP/SEL ratio for the FDTE were 9.76 μmol/g of dw, and 1.60, respectively.

MPLS Loading Plots. Figure 4 shows the MPLS loading spectra for factors 1, 2, and 3, respectively, for sinigrin (2, 5, 5, 2; SNV + DT), and **Figure 5** displays the NIR reflectance

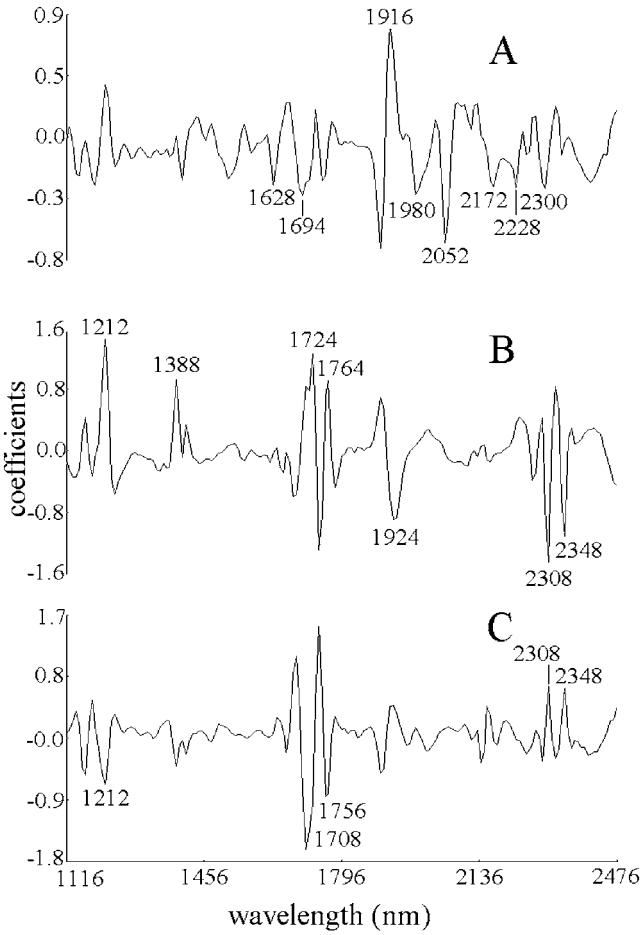

Figure 4. MPLS loading spectra for sinigrin in *B. juncea* seed in the second derivative (2, 5, 5, 2; SNV + DT) transformation. Panels **A**, **B**, and **C** represent loadings for factors 1, 2, and 3, respectively.

second-derivative spectrum of commercial sinigrin monohydrate ($C_{10}H_{16}NO_9S_2K{\cdot}H_2O$) (Sigma).

The first MPLS factor of the SDTE was mainly influenced by bands at 2052 nm related to N−H symmetric stretch plus amide II or N−H asymmetric stretch plus amide III and at 1916 nm related to O−H stretch/OH deformation hydroxyl (*32*). Other wavelengths displayed in the first MPLS loading plot that matched those of the sinigrin spectrum were 1628 nm, related to C−H stretch first overtone of =CH groups; 1694 nm, related to S−H stretch first overtone or C−H stretch first overtone of the CH_3 groups; 1980 and 2172 nm, related to N−H stretch of amides; and 2300 nm, assigned to C−H combination tones (*32*, *33*).

The second MPLS loading was mainly influenced by C−H combination tones at 2348 and 2308 nm, C−H stretch first overtone at 1764 and 1724 nm, C−H combinations at 1388 nm, and C−H stretch second overtone at 1212 nm and also by water. The third MPLS factor was the most correlated to sinigrin, and its loading plot resembled that of the second factor.

DISCUSSION

The accuracy of an NIR equation to predict unknowns is usually established on the basis of the r^2 (*34*) and RPD and RER (*24*, *34*) statistics, as generated in the external validation. However, other authors consider that the SECV and the R^2 are the two basic statistics to establish the acceptable level of accuracy for a calibration (*22*), as SEP can be quite variable depending on laboratory errors and sample representation in the validation set.

The values for R^2 shown by the equations in this work indicated excellent quantitative information (FDTE and SDTE of sinigrin and gluconapin), good quantitative information (FDTE and SDTE of total glucosinolates), or correct separation of the samples into high and low groups (4-hydroxyglucobrassicin) (*22*). On the basis of the r^2 shown in the external validation, excellent (SDTE of gluconapin) and good quantitative information (SDTE of sinigrin and FDTE of total glucosinolates) was obtained, except for the 4-hydroxyglucobrassicin. The low R^2 and r^2 exhibited by the 4-hydroxyglucobrassicin equations are explained here by the low mean content of this glucosinolate in the samples of the *B. juncea* seed and, also, by the larger inaccuracies in the determination of this indole glucosinolate in comparison to the others, factors that seriously affect the existing correlation (*34*). According to Williams (*34*), accuracy in the reference analysis is essential to setting up efficient NIR calibrations. Usually, the lower accuracy exhibited by NIRS is due to inaccuracies in their respective reference analyses and the SEP has to be put in perspective by relating it to SEL. The increment of the error of the laboratory expressed by the 4-hydroxyglucobrassicin could be explained in this work, by its extremely low concentration in the seed (mean = 2.42 μmol/g of dw), which would made it more prone to variations in the analytical process. This assertion is supported by the comparative study of the different SEP/SEL values for the equations obtained in this work. Thus, on the basis of this ratio, the FDTE of the 4-hydroxyglucobrassicin and total glucosinolates, which showed the lowest r^2 in the external validation of all the parameters (**Table 1**), were classified as equations showing excellent precision. In contrast, the SDTE of sinigrin and gluconapin, which, on the other hand, showed high r^2, were classified at a lower level of precision (good precision) on the basis of the SEP/SEL ratio.

The cross-validation and external validation resulted in similar SD/SECV and RPD values for all of the parameters except for gluconapin. The SDTE for gluconapin showed a higher RPD performance in the external validation (RPD = 4.16) than in cross-validation (SD/SECV = 2.99). An explanation for this is that SEP values are limited by the degree of correlation between reference data and NIR predictions (*34*). Thus, the higher r^2 shown by gluconapin in the external validation (0.95) with respect to that of the cross-validation (0.89) would lead to a lower SEP, thus increasing the RPD. The higher r^2 obtained in the external validation is probably related to a particular representation of the samples in the validation set, and this supports the idea that SECV is the best single estimate of the prediction capability of the equation (*22*).

Many authors have reported variable NIR prediction data on *Brassica* glucosinolates, but mainly in rapeseed (*6*−*10*), because of the commercial interest of this species. Coefficients of determination for total glucosinolates in the external validation reported in "canola" commodity by Williams and Sobering (*9*) and by Daun et al. (*10*), in NIR reflectance, were 0.74 and 0.82, respectively, these results being similar to those reported in the present work. In addition, the above-mentioned authors obtained RPDs for total glucosinolates that ranged from 1.94 to 1.36, values that were lower than those obtained by us in this study. An RPD value (2.29) similar to those shown by us in this study for total glucosinolates has been reported by Daun et al. (*10*). However, other authors have reported high-accuracy equations for total glucosinolates (*6*). An RPD of 10.78 is inferred from

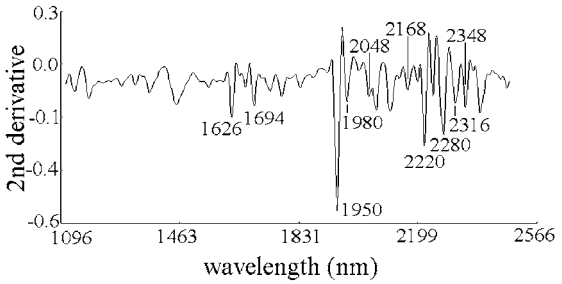

Figure 5. NIR reflectance second-derivative (2, 5, 5, 2) spectrum of commercial sinigrin monohydrate ($C_{10}H_{16}NO_9S_2K \cdot H_2O$).

the SEP and SD data reported, which indicates a predictive model with a high accuracy.

In a previous work (*13*) performed with *B. juncea* intact seed, we obtained coefficients of determination in cross-validation for sinigrin (1-VR = 0.74) and gluconapin (1-VR = 0.80) that were similar to those shown in this work, and the value obtained for total glucosinolates (1-VR = 0.88) was higher than that presented in this study (**Table 1**). In our opinion, the lower coefficient of determination for the total glucosinolate equation shown in this work could be due to the higher variability exhibited by the samples, in comparison to those of the previous work. Basically, this increment of variability was done by adding a significant number of samples with a low total glucosinolate content (**Table 1**). This could lead to an increment of the error of the HPLC analysis, thus decreasing the prediction ability of the new equation.

On the basis of the similarities between the second-derivative transformation of the sinigrin monohydrate spectrum (**Figure 5**) and the first MPLS loading for this constituent (**Figure 4A**), it seems that absorbers of the sinigrin molecule participated directly in modeling this factor. However, some of the groupings occurring in the sinigrin molecule also occur in the major cell constituents (protein, lipids, starch, cellulose, etc.) as overtones or combinations. This makes it difficult to conclude the degree of participation of the structural proteins, cellulose, or oil in the calibration of sinigrin. Because the −R group of glucosinolates is derived from amino acids, it is possible that some correlation between the protein and the total glucosinolate content of the embryo exists. In addition, protein and oil have been reported to show a high negative correlation in the seed of *Brassica* (*14*). These inherent correlations between glucosinolates and some major constituents in the seed could also explain the NIR modeling capability for sinigrin.

Results reported in this work show that NIRS is able to predict the sinigrin, gluconapin, and total glucosinolate contents in the intact seed of *Brassica juncea*, with sufficient accuracy for screening purposes. Each sample that we analyzed by using the NIRS method took us ∼1.5 min, and prediction results for the individual and total glucosinolate contents were monitored instantaneously. The equations shown in this work have been recently applied to the evaluation of the glucosinolate composition of ∼3000 individual plants of *B. juncea*. This has allowed the rapid identification and selection of the genotypes of interest, which would not have been possible by using HPLC. NIRS is thus ideal for mass screening programs in large-scale plant monitoring.

ABBREVIATIONS USED

FDTE, first-derivative transformation equation; MPLS, modified partial least-squares; NIRS, near-infrared spectroscopy; PCA, principal component analysis; RER, ratio of the range to standard error of prediction; RPD, ratio of the standard deviation to standard error of prediction; SD, standard deviation; SDTE, second-derivative transformation equation; SEC, standard error of calibration; SECV, standard error of cross-validation; SEL, standard error of laboratory; SEP, standard error of prediction; SNV-DT, standard normal variate and detrending; R^2, coefficient of determination in the calibration; r^2, coefficient of determination in the external validation; 1-VR, coefficient of determination in the cross-validation.

ACKNOWLEDGMENT

We thank Dr. Ian Murray (Scottish Agricultural College, Aberdeen, Scotland) for providing useful suggestions on the manuscript and also Dr. Jean Paul Wathelet (University of Gembloux, Gembloux, Belgium), Dr. Alain Quinsac (CETIOM, Grignon, France), and Gloria Fernández Marín (IAS, CSIC, Córdoba, Spain) for the performance of the HPLC analyses.

LITERATURE CITED

(1) Hashem, F. A.; Saleh, M. M. Antimicrobial components of some Cruciferae plants (*Diplotaxis harra* Forsk. and *Erucaria microcarpa* Boiss.). *Phytother. Res.* **1999**, *13*, 329−332.

(2) Fahey, J. W.; Zalcmann, A. T.; Talalay, P. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **2001**, *56*, 5−51.

(3) Rosa, E. A. S.; Heancy, R. K.; Fenwick, G. R.; Portas, C. A. M. Glucosinolates in crop plants. *Hortic. Rev.* **1997**, *19*, 99−215.

(4) Sorensen, H. Glucosinolates: structure, properties, function. In *Canola and Rapeseed. Production, Chemistry, Nutrition and Processing Technology*; Shahidi, F., Ed.; van Nostrand Reinhold: New York, 1990; pp 149−172.

(5) Downey, R. K.; Röbbelen, G. *Brassica* species. In *Oil Crops of the World*; Röbbelen, G., Downey, R. K., Ashri, A., Eds.; McGraw-Hill: New York, 1989; pp 339−362.

(6) Biston, R.; Dardenne, P.; Cwikowski, M.; Marlier, M.; Severin, M.; Wathelet, J. P. Fast analysis of rapeseed glucosinolates by near infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.* **1988**, *65*, 1599−1600.

(7) Michalski, K.; Krzymanski, J. Application of NIR method for screening of oilseed rape on glucosinolate content. *Cruciferae Newsl.* **1988**, *13*, 54−55.

(8) Salgó, A.; Weinbrenner-Varga, Zs.; Fabián, Z.; Ungár, E. Glucosinolates, determination without any wet chemical analysis. In *Making Light Work: Advances in Near Infrared Spectroscopy*; Murray, I., Cowe, I. A., Eds.; VCH: Weinheim, Germany, 1992; pp 328−332.

(9) Williams, P. C.; Sobering, D. C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* **1993**, *1*, 25−32.

(10) Daun, J. K.; Clear, K. M.; Williams, P. Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. *J. Am. Oil Chem. Soc.* **1994**, *71*, 1063−1608.

(11) Michalski, K.; Kolodziej, K. Measurement of glucosinolate content in intact seeds of rapeseed with NIR reflectance spectrometry. In *International Consultative Group of Rapeseed (GCIRC)*; GCIRC web publications, 2003; http://195.101.239.21/GCIRCFR.html.

(12) Velasco, L.; Becker, H. C. Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. *Plant Breed.* **1998**, *117*, 97−102.

(13) Font, R.; Del Río, M.; Domínguez, J.; Fernández-Martínez, J. M.; De Haro, A. Using of NIRS for determining glucosinolate content in *Brassica juncea* seed. In *Proceedings of the 10th International Rapeseed Congress*; Wratten, N., Salisbury, P. A., Eds.; The Regional Institute, Ltd.: Canberra, Australia, 1999.

(14) Uppström, B. Seed chemistry. In *Brassica Oilseeds. Production and Utilization*; Kimber, D., McGregor, D. I., Eds.; CAB International: Wallingford, U.K., 1995; pp 217−242.
(15) Fereres, E.; Fernández-Martínez, J. M.; Minguez, Y.; Domínguez, J. Productivity of *Brassica juncea* and *Brassica carinata* in relation to rapeseed, *B. napus*. I. Agronomic studies. In *Proceedings of the 6th International Rapeseed Congress*; International Consultative Research Group on Rapeseed (GCIRC): Paris, France, 1983; pp 293−298.
(16) Soil Survey Staff. *Soil Survey Manual, Handbook 18*; U.S. Department of Agriculture: Washington, DC, 1951.
(17) Font, R.; del Río, M.; Fernández-Martínez, J. M.; De Haro, A. Using near infrared spectroscopy (NIRS) for the determination of bulk density in Indian mustard seed. *Cruciferae Newsl.* **1999**, *21*, 75−76.
(18) Giese, A. T.; French, C. S. The analysis of overlapping spectral absorption bands by derivative spectrophotometry. *Appl. Spectrophotom.* **1955**, *9*, 78−96.
(19) Shenk, J. S.; Workman, Jr., J. J.; Westerhaus, M. O. Application of NIR spectroscopy to agricultural products. In *Handbook of Near-Infrared Analysis*; Burns, D. A., Ciurczak, E. W., Eds.; Dekker Inc.: New York, 1992; pp 383−431.
(20) Shenk, J. S.; Westerhaus, M. O. Population structuring of near infrared spectra and modified partial least squares regression. *Crop Sci.* **1991**, *31*, 1548−1555.
(21) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772−777.
(22) Shenk, J. S.; Westerhaus, M. O. Calibration the ISI way. In *Near Infrared Spectroscopy: The Future Waves*; Davies, A. M. C., Williams, P. C., Eds.; Nir Publications: Chichester, U.K., 1996; pp 198−202.
(23) Shenk, J. S.; Westerhaus, M. O. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.* **1991**, *31*, 469−474.
(24) Williams, P. C.; Sobering, D. C. How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In *Near Infrared Spectroscopy: The Future Waves*; Davies, A. M. C., Williams, P. C., Eds.; Nir Publications: Chichester, U.K., 1996; pp 185−188.
(25) Thies, W. Isolation of sinigrin and glucotropaeolin from cruciferous seeds. *Fat Sci. Technol.* **1988**, *90*, 311−314.
(26) Martens, H.; Næs, T. Outlier detection. In *Multivariate Calibration*; Martens, H., Næs, T., Eds.; Wiley: Chichester, U.K., 1991; pp 267−295.
(27) Röbbelen, G.; Thies, W. Variation in rapeseed glucosinolates and breeding for improved meal quality. In *Brassica Crops and Wild Allies*; Tsunoda, S., Hinata, K., Gómez-Campo, C., Eds.; Japan Scientific Societies Press: Tokyo, Japan, 1980; pp 285−299.
(28) Gland, A.; Röbbelen, G.; Thies, W. Variation of alkenyl glucosinolates in seeds of *Brassica* species. *Z. Pflanzenzuechtg.* **1981**, *87*, 96−110.
(29) Josefsson, E. Variation of pattern and content of glucosinolates in seed of some cultivated Cruciferae. *Z. Pflanzenzuechtg.* **1972**, *68*, 113−123.
(30) Velasco, L.; Becker, H. C. Variability for seed glucosinolates in a germplasm collection of the genus *Brassica*. *Genet. Resour. Crop Evol.* **2000**, *47*, 231−238.
(31) Batten, G. D. Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Aust. J. Exp. Agric.* **1998**, *38*, 697−706.
(32) 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., Eds.; American Association of Cereal Chemists: St. Paul, MN, 1987; pp 17−34.
(33) Murray, I. The NIR spectra of homologous series of organic compounds. In *Proceedings of the International NIR/NIT Conference*; Hollo, J., Kaffka, K. J., Gonczy, J. L., Eds.; Akademiai Kiado: Budapest, Hungary, 1986; pp 13−28.
(34) Williams, P. C. Variables affecting near-infrared reflectance spectroscopy analysis. In *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P. C., Norris, K., Eds.; American Association of Cereal Chemists: St. Paul, MN, 1987; pp 143−167.

Received for review November 17, 2003. Revised manuscript received March 10, 2004. Accepted March 11, 2004. This work has been supported by Project CICYT AGF98-0917 of the Spanish government.

JF0307649