

Quantification of glucosinolates in leaves of leaf rape (*Brassica napus* ssp. *pabularia*) by near-infrared spectroscopy

Rafael Font ^{a,*}, Mercedes del Río-Celestino ^a, Elena Cartea ^b, Antonio de Haro-Bailón ^a

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

^b Misión Biológica de Galicia (CSIC), Apartado 28, E-36080 Pontevedra, Spain

Received 31 August 2004; received in revised form 12 November 2004

Available online 18 December 2004

Abstract

The potential of near-infrared spectroscopy (NIRS) for screening the total glucosinolate (t-GSL) content, and also, the aliphatic glucosinolates gluconapin (GNA), glucobrassicinapin (GBN), progoitrin (PRO), glucoalyssin (GAL), and the indole glucosinolate glucobrassicin (GBS) in the leaf rape (*Brassica napus* L. ssp. *pabularia* DC), was assessed. This crop is grown for edible leaves for both fodder and human consumption. In Galicia (northwestern Spain) it is highly appreciated for human nutrition and have the common name of “nabicol”. A collection of 36 local populations of nabicol was analysed by NIRS for glucosinolate composition. The reference values for glucosinolates, as they were obtained by high performance liquid chromatography on the leaf samples, were regressed against different spectral transformations by modified partial least-squares (MPLS) regression. The coefficients of determination in cross-validation (r^2) shown by the equations for t-GSL, GNA, GBN, PRO, GAL and GBS were, respectively, 0.88, 0.73, 0.81, 0.78, 0.37 and 0.41. The standard deviation to standard error of cross-validation ratio, were for these constituents, as follows: t-GSL, 2.96; GNA, 1.94; GBN, 2.31; PRO, 2.11; GAL, 1.27, and GBS, 1.29. These results show that the equations developed for total glucosinolates, as well as those for gluconapin, glucobrassicinapin and progoitrin, can be used for screening these compounds in the leaves of this species. In addition, the glucoalyssin and glucobrassicin equations obtained, can be used to identify those samples with low and high contents. From the study of the MPLS loadings of the first three terms of the different equations, it can be concluded that some major cell components as protein and cellulose, highly participated in modelling the equations for glucosinolates.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: *Brassica napus*; Brassicaceae; Rapeseed nabicol leaf; Near-infrared spectroscopy; Glucosinolates

1. Introduction

Over the past three decades, *Brassica* production has increased to become an important source of oil and protein of plant origin for human and animal nutrition, respectively. In addition, some species of the genus are highly consumed as vegetables all over the world. In the Iberian Peninsula, the high consumption of *Brassica* crops is reflected by a large use of flower buds and leaves of several of these species. However, the information available on the glucosinolate contents of these products is scarce (Rosa, 1997).

Abbreviations: GAL, glucoalyssin; GBS, glucobrassicin; GNA, gluconapin; GBN, glucobrassicinapin; MPLS, modified partial least squares; NIRS, near infrared spectroscopy; PRO, progoitrin; RPD, ratio of the standard deviation to standard error of prediction; SD, standard deviation; SEC, standard error of calibration; SECV, standard error of cross-validation; SEP, standard error of prediction; S-NV-DT, standard normal variate and detrending; R^2 , coefficient of determination in the calibration; r^2 , coefficient of determination in the cross-validation; t-GSL, total glucosinolates.

* Corresponding author. Tel.: +34 957 499211; fax: +34 957 499252.

E-mail address: font@cica.es (R. Font).

Glucosinolates have a well defined structure with a side-chain (R-group) and D-glucopyranose as β -thioglucoside attached to carbon atom no. 0 in (Z)-N-hydroximine sulphate esters (Sørensen, 1990) (Fig. 1). The structural diversity of glucosinolates is mainly due to the different substituents possible at the side-chain position R, which can be very variable (Rosa et al., 1997). The hydrolysis products of glucosinolates are of great concern because they are the compounds responsible for many of the beneficial and harmful properties of glucosinolate containing plants. Among the beneficial uses of glucosinolate hydrolysis products are their antibacterial and antifungal properties applied to biofumigation (Angus et al., 1994; Fahey et al., 2001, 2002), and as cancer-chemoprevention agents (Rosa et al., 1997; Shapiro et al., 2001). In addition to the above mentioned attributes, hydrolysis products of glucosinolates are also the molecules responsible for the pungent and hot flavors characteristic of mustard seeds. Both intact glucosinolates and their hydrolysis products variously act as stimulants or deterrents to insects and herbivores (Louda and Mole, 1991).

The toxic and anti-nutritive effects of glucosinolates have limited the use of *Brassica* species for human and animal feed (Sørensen, 1990). Among these negative aspects of glucosinolate-containing plants are the goitrogenic effects in animals (Fenwick et al., 1983a). The negative effects of glucosinolates have been the reason for research aimed at reducing glucosinolate contents in the seeds of some *Brassica* crops (Downey and Röbbelen, 1989).

Brassica crops are widely cultivated in Galicia (north-western Spain), in particular those corresponding to the species *B. napus* L., *B. oleracea* L. and *B. rapa* L. Crops of these species are usually consumed as fresh vegetables by local human populations, but sometimes they are manufactured and commercialised, as occurs in the case of a variety of *B. rapa* known as “grelos”.

Brassica napus is a variable species, divided into three groups or subspecies: *B. napus* ssp. *oleifera* (DC), includes oilseed rape and canola grown for the seeds from which vegetable oil is made or for edible leaves, as forage crops; *B. napus* ssp. *rapifera* L. includes the rutabaga,

which is grown for their enlarged swollen stems; and *B. napus* ssp. *pabularia* (DC) that includes rape kale, Siberian kale, and Hanover salad, grown for leafy greens like kales in northern Europe. Crops of *B. napus* ssp. *pabularia* are traditionally cultivated in the Northwest of Iberian Peninsula (South of Galicia and North of Portugal) where they are grown during the winter season and they are known as “nabicol” (Cartea et al., 2004).

The lack of commercial interest for this crop, has allowed the conservation of local populations of this form of *B. napus*. This fact is noteworthy, as they could serve as a base germplasm for future research in plant breeding programmes (Rodríguez et al., 2003), to enlarge the genetic variability of this species. Currently, those studies characterising the different varieties of “nabicol” in Galicia are minimal. There are previous studies on “nabicol” populations concerning their morphological and agronomic attributes (Rodríguez et al., 2003), and their genetic diversity (Cartea et al., 2004). However, this crop has not been studied for nutritive value and there is no information about the leaf composition. This is especially valid for the glucosinolate composition of the leaves, to which these compounds give a characteristic pungent taste which is highly appreciated.

For several decades, many authors have focused research in determining the glucosinolate composition of seeds, and to a lesser extent, of the leaves of *Brassica* species (Fahey et al., 2001, 2002). However, the determination of the glucosinolate content by the standard methods is expensive and time-consuming, and in addition, specialised personal is needed. The high cost and labour input required for obtaining the glucosinolate content in samples by HPLC are serious handicaps to analyse large sets of samples, which is usually necessary to identify the target genotypes in screening programs.

In contrast, the use of fast analytical techniques such as near-infrared spectroscopy (NIRS) results in many advantages, since analysis can be carried out with a considerable saving of time, at a low cost and without using hazardous chemicals (Gottlieb et al., 2004). NIRS has been widely used for decades for qualitative and quantitative analysis in environmental (Clark et al., 1989; Font et al., 2004a), agricultural and food research. Many authors have used this technique for determining the glucosinolate content in seeds of *Brassica* species (Biston et al., 1988; Daun et al., 1994; Velasco and Becker, 1998; Font et al., 2004b). However, to date no studies have been reported on the use of the NIRS technique to measure glucosinolates in plant leaves, where these compounds are present in significantly lower concentrations than those usually found in the seed. The purpose of this work was to test the potential of NIRS for predicting the total glucosinolate (t-GSL) content, as well as those major glucosinolates found in the leaves of *B. napus* ssp. *pabularia* (Table 1). In addition, we provide some knowledge about the mechanism used by NIRS for

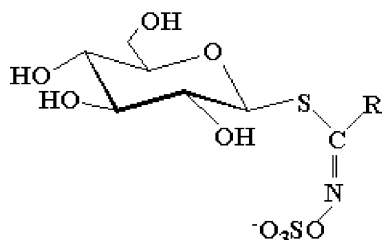
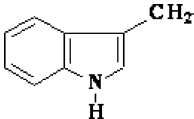


Fig. 1. Basic structure of glucosinolates. The R substituent may be an alkyl or alkenyl side-chain which itself may contain substituents such as hydroxyl groups or sulfur. Alternatively, the R substituent may be an aromatic or a hetero-aromatic group.

Table 1

Trivial names and side-chain structures of the glucosinolates considered in this work

Trivial name (abbreviation)	Chemical structure of side chain (R)
Gluconapin (GNA)	$\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_2-$
Glucobrassicinapin (GBN)	$\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_2-\text{CH}_2-$
Progoitrin (PRO)	$\text{CH}_2=\text{CH}-\text{CH}(\text{OH})-\text{CH}_2-$
Glucoalyssin (GLA)	$\text{CH}_3-\text{SO}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-$
Glucobrassicin (GBS)	

determining glucosinolates successfully in the leaves of this species.

2. Results

2.1. Reference analysis of glucosinolates in the samples

In previous screenings of natural populations of *B. n. ssp. pabularia* plants collected at different places in its natural cropping area, we have never found sinigrin among the different glucosinolates which are present in this species. By this reason, it was decided the use of this glucosinolate as an internal standard in the analyses. In addition to our findings, previous studies have demonstrated the lack of sinigrin in natural populations of *B. napus* plants (Rosa, 1997).

The major glucosinolates found in the leaf of *B. n. ssp. pabularia* were progoitrin (PRO) (2-hydroxybut-3-enylglucosinolate), glucoalyssin (GLA) [5-(methysulfi-

nyl)pentylglucosinolate)], gluconapin (GNA) (but-3-enylglucosinolate), glucobrassicinapin (GBN) (pent-4-enylglucosinolate) and glucobrassicin (GBS) (3-indolylmethylglucosinolate) (Table 1). Mean concentrations of all these glucosinolates were over $1 \mu\text{mol g}^{-1}$ dry wt (Table 2). Other minor glucosinolates that were present in the samples considered herein were napoleiferin (2-hydroxypent-4-enylglucosinolate), 4-hydroxyglucobrassicin (4-hydroxy-3-indolylmethylglucosinolate), 4-methoxyglucobrassicin (4-methoxy-3-indolylmethylglucosinolate) and neoglucobrassicin (1-methoxy-3-indolylmethylglucosinolate). The glucosinolates had mean concentrations that were below $0.1 \mu\text{mol g}^{-1}$ dry wt. Because of the very low concentrations and low standard deviation values exhibited by all these minor glucosinolates in the samples, they were not considered for the NIRS analysis. Tentative identifications of the glucosinolates shown in this work were based on comparison of retention times with those of certified standards.

Typical chromatograms of *B. n. ssp. pabularia* leaf samples in this work showed a profile that was similar to that of *B. napus ssp. midas* reported by Sang et al. (1984). However, a quantitative comparative study was not possible between data reported by Sang et al. (1984) and those obtained by us in the present work, because these authors did not publish the concentrations of glucosinolates found in the samples studied.

Frequency distributions of the different glucosinolates in the samples used in this work (Fig. 2) showed that the t-GSL (Fig. 2(a)) and also, the individual aliphatic glucosinolates GNA (Fig. 2(b)), GBN (Fig. 2(c)) and PRO (Fig. 2(d)) exhibited approximately normal distributions in their intervals. In contrast, GAL and the indole GBS showed skewed distributions with

Table 2

Calibration and cross-validation statistics for the different equations developed for glucosinolates ($n = 115$)

Glucosinolate	Calibration					Cross-validation		
	Range	Mean	SD ^g	SEC ^h	R^{2i}	SD SECV ^{-lj}	r^{2k}	nt ^l
t-GSL ^a	1.06–49.18	19.85	11.03	1.88	0.97	2.96	0.88	9
GNA ^b	0–6.88	2.76	1.54	0.61	0.84	1.94	0.73	5
GBN ^c	0.31–21.07	8.03	5.1	1.47	0.91	2.31	0.81	7
PRO ^d	0.10–13.03	5.66	3.28	0.98	0.91	2.11	0.78	8
GAL ^e	0–4.00	1.06	1.15	0.80	0.50	1.27	0.37	3
GBS ^f	0.06–3.45	1.01	0.83	0.59	0.50	1.29	0.41	3

^a Total glucosinolates.

^b Gluconapin.

^c Glucobrassicinapin.

^d Progoitrin.

^e Glucoalyssin.

^f Glucobrassicin.

^g Standard deviation of the reference data obtained by “wet chemistry”.

^h Standard error of calibration.

ⁱ Coefficient of determination of the calibration.

^j Ratio of the standard deviation of the reference data to the standard error of cross-validation.

^k Coefficient of determination of the cross-validation.

^l Number of terms of the equation selected in cross-validation.

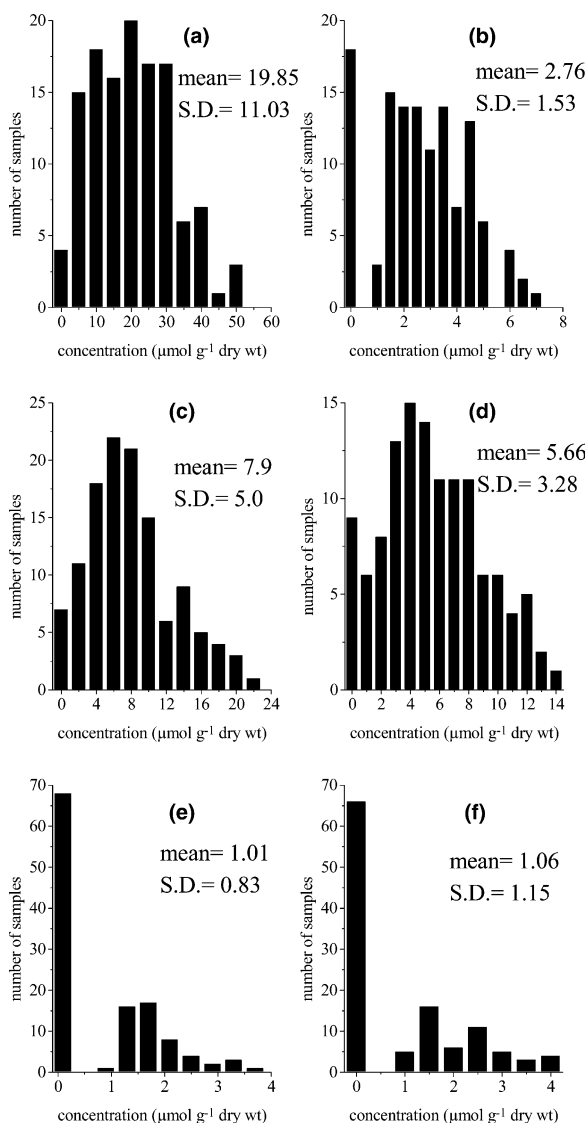


Fig. 2. Frequency distribution plots of total glucosinolates (a), gluconapin (b), glucobrassicinapin (c), progoitrin (d), glucobrassicin (e) and glucoalyssin (f) for the samples used in this work ($n = 115$) ($\mu\text{mol g}^{-1}$ dry wt).

a greater number of samples with contents below $0.50 \mu\text{mol g}^{-1}$ dry wt.

The ranges, means and standard deviations of the total and individual glucosinolates used in this study are summarised in Table 2. Individual plants exhibited t-GSL concentrations that ranged from 1.06 to $49.18 \mu\text{mol g}^{-1}$ dry wt, and a mean value of $19.85 \mu\text{mol g}^{-1}$ dry wt. These concentrations are similar to those contents previously reported in *B. napus* (Fenwick et al., 1983a,b). GBN was the glucosinolate that showed the higher mean content of all them, representing 40% of the t-GSL, followed by PRO and GNA. The aliphatic GAL and the indole GBS glucosinolates showed ranges from 0 to $4 \mu\text{mol g}^{-1}$ dry wt, with mean contents of $1 \mu\text{mol g}^{-1}$ dry wt.

It is important to stress that some samples showed t-GSL values that were over the recommended intake levels for use in animal feeds ($\geq 30 \mu\text{mol g}^{-1}$ dry wt) (Bell, 1995).

2.2. Spectral data pre-treatments and equation performances

The application of the second derivative and standard normal variate and de-trending algorithms to the raw spectra ($\log 1/R$) (Fig. 3), resulted in substantial correction (Fig. 4) of the baseline shift caused by differences in particle size and path length. Peaks and troughs in Fig. 4 correspond to the points of maximum curvature in the raw spectrum, and it has a trough corresponding to each peak in the original. The increase in the complexity of the derivative spectra resulted in a clear separation between peaks which overlap in the raw spectra.

The use of the mathematical approach described above yielded the equations with the highest prediction abilities when it was applied over the near-infrared segment (1100–2500 nm). The visible segment of the spectrum was left out of the calibrations as it only contributed with noise to the MPLS models for gluco-

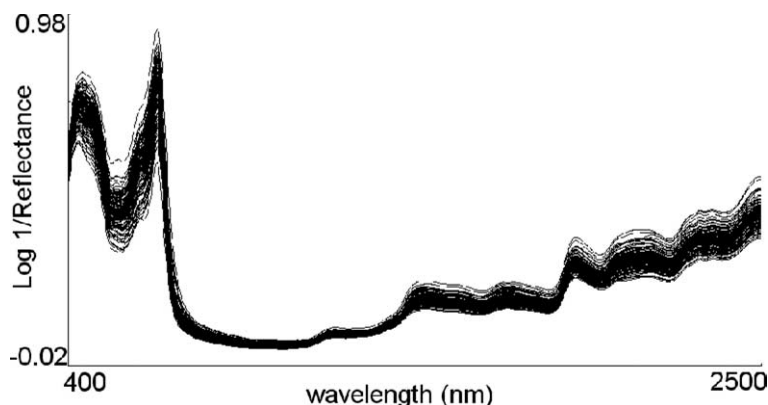


Fig. 3. Raw spectra ($\log 1/R$) of the leaf samples of *Brassica napus* ssp. *pabularia* used in this work ($n = 115$), in the range from 400 to 2500 nm.

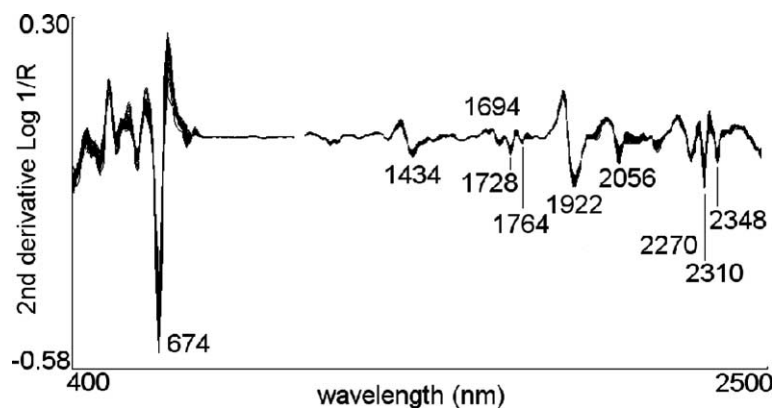


Fig. 4. Second derivative spectra (2, 5, 5, 2; SNV + DT) of the raw optical data in the range from 400 to 2500 nm.

sinolate concentrations, a phenomenon that has been previously reported for the estimation of other components (Gislum et al., 2004).

2.2.1. Total glucosinolates

The t-GSL equation showed a high coefficient of determination in the calibration ($R^2 = 0.97$) and low

standard error of calibration ($SEC = 1.88 \mu\text{mol g}^{-1}$ dry wt) (Table 2). Nine terms were selected in cross-validation as the optimum number to model the equation. The final model for t-GSL showed the highest r^2 (0.88) and also standard deviation (SD) $SECv^{-1}$ ratio (2.96) (Fig. 5(a)) shown by any of the equations for individual glucosinolates.

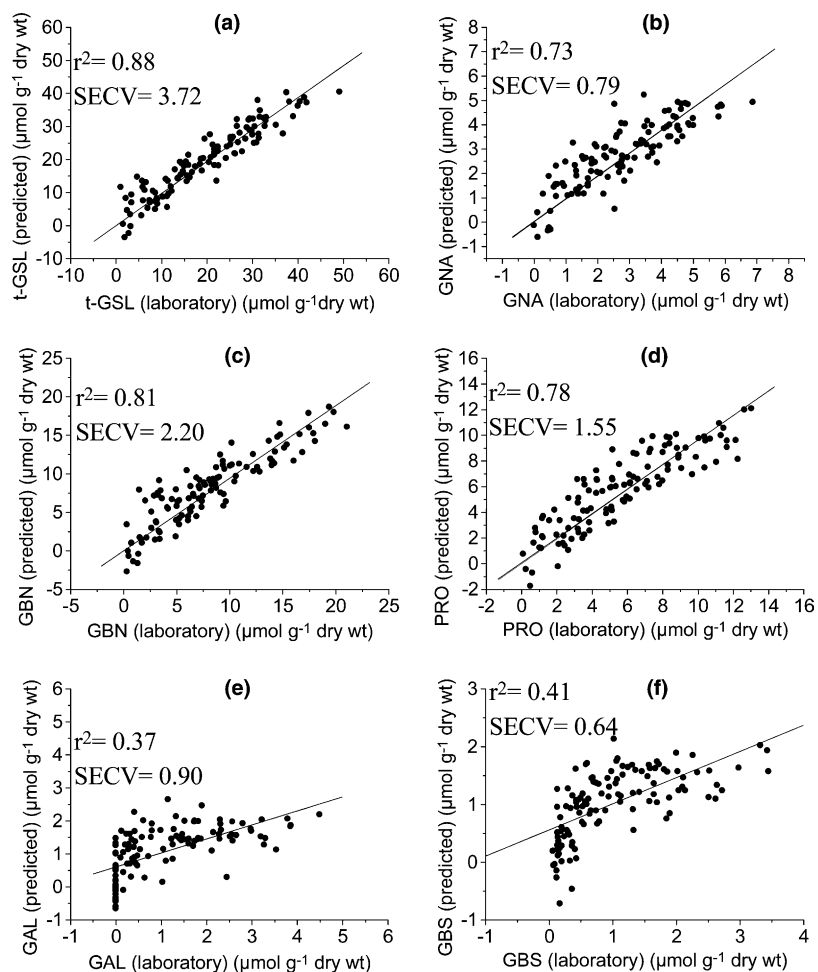


Fig. 5. Cross-validation scatter plots of laboratory vs. predicted values by NIRS for total glucosinolates (a), gluconapin (b), glucobrassicinapin (c), progoitrin (d), glucoalyssin (e) and glucobrassicin (f) ($n = 115$) ($\mu\text{mol g}^{-1}$ dry wt).

2.2.2. *Gluconapin*

The equation for GNA showed a lower SEC ($0.61 \mu\text{mol g}^{-1}$ dry wt) and higher R^2 (0.84) in calibration than raw data or first derivative. In cross-validation, the second derivative equation was modelled with five terms, and also showed the highest r^2 (0.73) (Fig. 5(b)) and SD SECV^{-1} (1.94) ratio of those shown by the other mathematical transformations.

2.2.3. *Glucobrassicinapin*

The second derivative equation resulted in a high R^2 (0.91) and low SEC ($1.47 \mu\text{mol g}^{-1}$ dry wt) in the calibration for GBN (Table 2). Seven terms were selected in cross-validation as the optimum number to fit the model. The r^2 (0.81) (Fig. 5(c)) and SD SECV^{-1} (2.31) values obtained for this glucosinolate, were higher than those shown by the equation for GNA.

2.2.4. *Progoitrin*

Similar prediction ability to that shown by GBN was exhibited by the second derivative equation for PRO (Table 2). As previously occurred for the other glucosinolates, the second derivative equation (2, 5, 5, 2; SNV + DT) performed over the infrared segment exhibited the highest prediction ability of the different equations for this glucosinolate. This equation showed coefficients of determination in calibration and cross-validation of 0.91 and 0.78, respectively (Fig. 5(d)), and a SD SECV^{-1} ratio of 2.11, which was close to those exhibited by the GNA and GBN equations. The equation for PRO was modelled with eight terms.

2.2.5. *Glucoalyssin and glucobrassicin*

Equations for GAL and GBS showed similar prediction abilities, as it can be concluded from the data reported in Table 2. Both equations showed R^2 values of 0.50, and also similar SD SECV^{-1} ratios, which were close to 1.30. These ratios were the lowest of the different glucosinolates analysed, and also the r^2 values, which were close to 0.4 (Fig. 5(e) and (f)). Both equations were modelled with three terms, as being the optimum number selected in cross-validation.

2.3. Second derivative spectra of leaf and modified partial least square loadings

Wavelengths noted on Fig. 4 as being those of most relevance in the spectra, correspond to those wavelengths of maximum absorbance shown by the average spectrum of the freeze-dried leaf. The conspicuous band in the visible region at 674 nm is due to electronic transitions in the red and has been assigned to absorption by chlorophyll (Tkachuk and Kuzina, 1982). In the NIR segment of the spectrum, the main absorption bands were displayed at 1922 nm, which has been attributed to O–H stretch plus O–H deformation; 2056 nm related

to N–H stretch of amides; 2270 nm which has been assigned to O–H plus C–C stretch groups (Osborne et al., 1993) of cellulose, and at 2310 and 2348 nm related to C–H stretching and combination bands of the methylene groups (Murray and Williams, 1987). Other minor absorptions were due to the first overtone of O–H stretching (1434 nm), S–H stretch first overtone or C–H stretch first overtone of CH_3 groups (1694 nm), and C–H stretching by methylene groups (1728 and 1764 nm).

Those wavelengths corresponding to absorptions by CH_2 stretching and combination, N–H stretching by amides and O–H stretch/OH deformation hydroxyl (Murray and Williams, 1987) influenced highly the first two factors of the t-GSL equation (Fig. 6), which in addition were the factors most correlated to the total content. The third factor was mainly modelled with those wavelengths previously used in the other factors, but with an increase in the participation of hydroxyl groups.

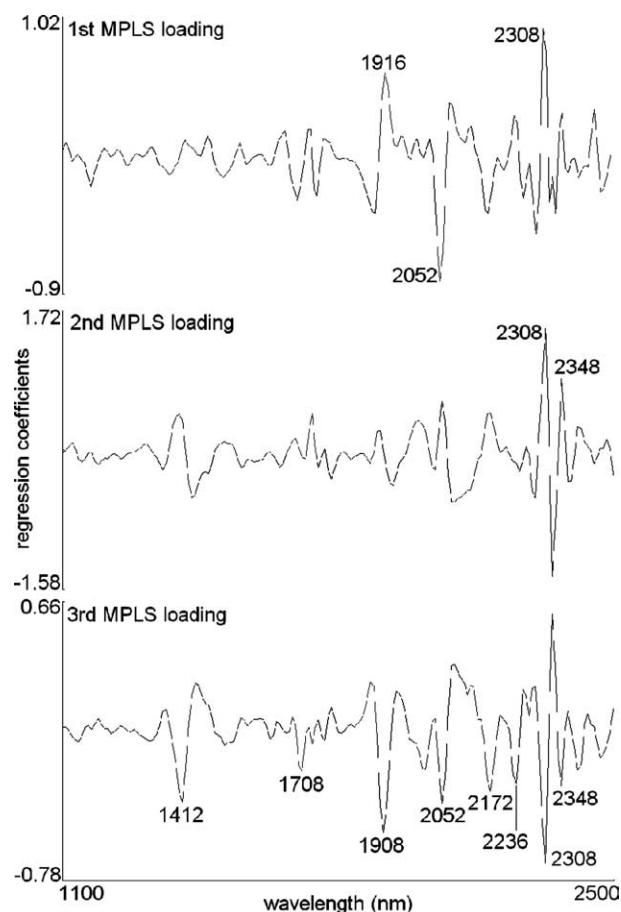


Fig. 6. Plots of the first, second and third loadings obtained by modified partial least squares regression for the total glucosinolates equation. These loadings represent factors (terms) 1, 2 and 3 of the equation, respectively. The level of participation of the different wavelengths to each term of the equation is given by its regression coefficient value (y axis).

High similarities were found among the first three MPLS loadings of the different glucosinolate equations and that for t-GSL equation. Wavelengths used in modelling the first three terms of the t-GSL equation were systematically used in modelling the different factors for aliphatic, aromatic and indole glucosinolate equations.

3. Discussion

The validity of cross-validation to evaluate the performance of an NIR equation has been supported by different researchers (Shenk and Westerhaus, 1996; Williams and Sobering, 1996), having been applied successfully by the authors of this work to the analysis of glucosinolates in a previous report (Font et al., 2004b). Following the considerations reported by Shenk and Westerhaus (1996) about the estimation of the accuracy of a calibration equation from cross-validation, the r^2 obtained for t-GSL, and also for the aliphatic GNA, GBN and PRO glucosinolates, was indicative of equations with good quantitative information. For these components, the mathematical models developed explained from 0.73% to 0.88% of the variance contained in the chemistry data (Table 2). In contrast, GAL and the indole glucosinolate GBS showed coefficients of determination that were characteristic of equations useful for good separation of samples into high and low groups.

The number of MPLS terms of the different equations selected as optimum in cross-validation was for all them, in the limits recommended to avoid overfitting, i.e., one term by each ten samples in the calibration file (Shenk and Westerhaus, 1995). From data shown in Table 2, it can be concluded that the variance in the reference chemistry values explained by the different equations, was related to the mean concentration of each glucosinolate in the samples. The accuracy of the HPLC analysis is highly dependent on the concentration of the component in the sample, and accuracy in the reference analysis is essential to setting up efficient NIR calibrations (Williams, 1987). The extremely low concentrations exhibited by GBS and GAL in the leaf samples would make them more prone to errors in obtaining the reference values, thus decreasing the correlation with the spectral information.

Cross-validation resulted in SD SECV^{-1} ratios that ranged from 1.27 (GAL) to 2.96 (t-GSL) (Table 2). The differences shown by these ratios for the different glucosinolates, are explained by the fact that the SECV value is limited by the degree of correlation between chemistry reference data and NIR predictions (Williams, 1987). The higher r^2 shown by the t-GSL equation with respect to those displayed by the individual glucosinolates, would lead to a lower SECV, thus increasing the

value of the ratio. In contrast, the low coefficients of determination displayed by those minor glucosinolates, lead to higher SECVs. In addition, when the range, and therefore, the variance in reference data are low, the values for r^2 and also the SD SECV^{-1} ratio, cannot be very high, which is the case for the GAL and GBS glucosinolates in the leaf of “nabicol”.

Previous studies reporting NIRS calibrations of glucosinolates in *Brassica* species have been performed mainly on rapeseed seed, because of the commercial interest of this species. Thus, Biston et al. (1988) reported r^2 values of 0.99 for total glucosinolates, independently of the different reference method they used (palladium, glucose, gas–liquid chromatography or HPLC). Williams and Sobering (1993) and Daun et al. (1994) also reported predictions for total glucosinolates with coefficients of determination that varied from 0.74 to 0.82, and ratios of the standard error of prediction (SEP) performed on external validation to the SD of the reference values, which is known as RPD, that ranged from 1.36 to 2.29. Other authors (Velasco and Becker, 1998) developed multi-product calibrations for individual and total glucosinolates considering simultaneously different *Brassica* species. These authors demonstrated the validity of the technique in approaches like that, reporting high r^2 values for individual (gluconapin = 0.89; sinigrin = 0.90; progoitrin = 0.86) and total glucosinolates (0.99) predictions. In a previous work performed on *B. juncea* (L. Czern. & Coss.) seed (Font et al., 2004b), we obtained r^2 values that ranged from 0.82 to 0.95 for total and individual glucosinolates. It has to be noted that all the above referenced works, showed prediction accuracies that were similar or slightly higher than those reported for leaves in this work, in spite of the lower concentrations of these compounds in *Brassica* leaves. The fact of the success of the NIR calibration at these low concentrations of glucosinolates found in the “nabicol” leaves was investigated, by studying those wavelengths mainly used in the first three terms of the calibration equation for t-GSL.

Those wavelengths highly participating in modelling the first three terms of the equations for glucosinolates (Fig. 6), were, to some extent, similar to those previously reported for determining glucosinolates in *B. juncea* seed (Font et al., 2004b). Due to the fact that glucosinolates derive from natural amino acids (Ettlinger and Kjaer, 1968; Mikkelsen et al., 2002), one might think that those features in the spectral bands related to protein absorption be used in the equations for glucosinolates. This is the case of the band at 2052 nm, which participated mainly in the construction of the first and third loadings of all the glucosinolates equations. Other inherent correlations between glucosinolates and major cell components as it is cellulose influenced such calibrations, as it is concluded from the bands at 2308 and 2348 nm. However, pure glucosinolates absorb also at these wave-

lengths (Font et al., 2004b), and thus, it cannot be ruled out that, to some extent, the spectrum of the “nabicol” leaf may be directly caused by the contained glucosinolates.

From the data reported in this work it is concluded that t-GSL, and also those major glucosinolates in the leaves of “nabicol”, i.e., GNA, GBN and PRO, can be predicted with sufficient accuracy for screening purposes. GAL and GBS equations can be used for a correct separation of the samples into “low” and “high” groups. After NIRS screening, more accurate analyses can be carried out by HPLC of those samples of interest. Thus, a considerable saving of the labour input, time and cost of analysis for these compounds is achieved. The development of these calibrations will allow researchers in the fields of plant breeding, biofumigation or medical applications, to quickly identify “nabicol” individuals of interest without the need of doing HPLC analysis.

4. Experimental

4.1. Plant material and crop management

This work is based on 115 individual plants belonging to 36 different accessions of the species *B. napus* ssp. *pabularia*, which are a part of the germplasm bank of Brassica at the Misión Biológica de Galicia (MBG) (CSIC, Spain). These accessions represent the whole variability of this variety of *B. napus* in its natural cropping area.

The plant material was grown in the years 2002 and 2003 in Pontevedra (Spain). Seeds of each accession were sown in greenhouse on August, in small pots containing sterile commercial potting mixture, under controlled conditions of temperature, light and irrigation.

Thirty days after sowing, the seedlings were transplanted to field at the five leaf stage. Before transplanting, the experimental field was fertilised with 64 kg ha^{-1} of N, 90 kg ha^{-1} of P_2O_5 , and 90 kg ha^{-1} of K_2O . Populations were sown in a 6×6 lattice design with three replications. Each experimental plot consisted of two rows with 10 plants per row. Rows were spaced 0.9 m apart and plants between rows 0.6 m apart (Cochran and Cox, 1957).

The season of harvest comprised from two to five months after planting corresponding to winter season because leaf flavor is enhanced by frost. From three to five mature individual plants of each one of the 36 accessions used in this study were collected for analysis at four months after planting. For each plant, three or four leaves were frozen “in situ” in liquid N_2 , to avoid glucosinolate enzymatic hydrolysis by the enzyme myrosinase (thioglucoside glucohydrolase, E.C. 3.2.3.1) contained in the cell vacuoles. Plant samples were then

transported to the laboratory. The green material were ground to a powder in liquid N_2 , kept at -80°C and freeze-dried in a Telstar freeze-dryer model Lioalfa-6, until analysis.

4.2. HPLC analysis

Leaves were analysed by HPLC at the Department of Agronomy and Plant Breeding (DAPB) at the Institute of Sustainable Agriculture (IAS, CSIC, Spain). About 100 mg dry wt of leaves was ground in a Janke and Kunkel, Model A10 mill (IKA-Labortechnik) for about 20 s and a two-step glucosinolate extraction was carried out in a waterbath at 75°C to inactivate myrosinase. In the first step the sample was heated for 15 min in 2.5 ml 70% aqueous methanol and 200 μl 10 mM sinigrin (2-propenyl glucosinolate) as an internal standard. A second extraction was applied after centrifugation (5 min, 5000g) by using 2 ml of 70% aqueous methanol. One ml of the combined glucosinolate extracts was pipetted onto the top of an ion-exchange column containing 1 ml Sephadex DEAE-A25 in the formate form. Desulphation was carried out by the addition of 75 μl of purified sulphatase (E.C. 3.1.6.1, type H-1 from *Helix pomatia*) (Sigma) solution. Sulphatase was purified according to the ISO protocol (ISO 9167-1, 1992). Desulphated glucosinolates were eluted with 2.5 ml ($0.5 \text{ ml} \times 5$) Milli-Q (Millipore) ultra-pure water and analysed 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). HPLC solvents and gradient were according to the ISO protocol (ISO 9167-1, 1992). The HPLC chromatogram was compared to the desulpho-glucosinolate profile of three certified reference materials recommended by U.E. and ISO (CRMs 366, 190 and 367) (Wathelet et al., 1991). The amount of each individual glucosinolate present in the sample was calculated by mean of the internal standard, and expressed as $\mu\text{mol g}^{-1}$ of dry wt. The total glucosinolate content was computed as the sum of all the individual glucosinolates present in the sample. Data were corrected for UV response factors for different types of glucosinolates (ISO 9167-1, 1992).

4.3. NIRS procedure: recording of spectra and processing of data

All the samples which were previously analysed by HPLC, were then analysed by NIRS. Near-infrared spectra were recorded on an NIRS spectrometer model 6500 (Foss-NIRSystems, Inc., Silver Spring, MD, USA) in reflectance mode equipped with a transport module. Ground freeze-dried of crude leaf samples were placed in a 3 cm diameter round cell sample holder, and

their spectra were registered as an individual file, in the range from 400 to 2500 nm, at 2 nm intervals.

The monochromator 6500 consists of a tungsten bulb and a rapid scanning holographic grating with detectors positioned for transmission or reflectance measurements. To produce a reflectance spectrum, a ceramic standard is placed in the radiant beam, and the diffusely reflected energy is measured at each wavelength. The actual absorbance of the ceramic is very consistent across wavelengths. In this work, each spectrum was recorded once from each sample, and was obtained as an average of 32 scans over the sample, plus 16 scans over the standard ceramic before and after scanning the sample. The ceramic and the sample spectra were used to generate the final $\log(1/R)$ spectrum. The whole time of analysis took about 2 min per sample, approximately.

In the second step, the calibration file was formed by adding the reference chemistry values for all glucosinolates, as they were obtained by HPLC, to the file of spectra, thus forming a new file, each spectrum having an associated value for each glucosinolate.

4.4. Developing calibration equations

Using the application GLOBAL v. 1.50 (WINISI II, Infrasoft International, LLC, Port Matilda, PA, USA), different calibration equations for t-GSL, GNA, PRO, GBN, GAL, and GBS were developed on the whole set ($n = 115$). 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); (1, 10, 10, 1); (2, 5, 5, 2); (2, 20, 20, 2)]. To correlate the spectral information (raw optical data or derivative spectra) and the chemistry values of the different glucosinolates, as they were determined by the reference method, modified partial least squares (MPLS) was used as a regression method, by using wavelengths from 400 to 2500 nm every 8 nm. In addition, the algorithms termed standard normal variate (SNV) and de-trending (DT) (Barnes et al., 1989) were used to correct baseline offset due to scattering effects produced by differences in particle size and path length variation among samples.

4.5. Validation of the equations

The performances of the different calibration equations obtained in the calibration were determined from cross-validation. Cross-validation is an internal validation method that like the external validation approach seeks to validate the calibration model on independent test data, but it does not waste data for testing only, as occurs in external validation. This procedure is useful because all available chemical analyses for all individu-

als can be used to determine the calibration model without the need to maintain separate validation and calibration sets. The method is carried out by splitting the calibration set into M segments and then calibrating M times, each time testing about a $(1/M)$ part of the calibration set (Martens and Naes, 1989). In this work, the different calibration equations were validated with 5 cross-validation segments (groups), as this was the optimum number of terms automatically selected by the software as a function of the number of samples employed.

The prediction ability of the equations obtained was determined on the basis of their coefficient of determination in the cross-validation (r^2) (Shenk and Westerhaus, 1996) (Eq. (1)) and SD to standard error of cross-validation (SECV) ratio (Williams and Sobering, 1996) (Eq. (2)).

$$r^2 = \left(\sum_{i=1}^n (\hat{y} - \bar{y})^2 \right) \left(\sum_{i=1}^n (y_i - \bar{y})^2 \right)^{-1}, \quad (1)$$

where \hat{y} = NIR measured value; \bar{y} = mean “ y ” value for all samples; y_i = lab reference value for the i th sample.

$$\text{SD SECV}^{-1} = \text{SD} \left\langle \left[\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) (N - K - 1)^{-1} \right]^{1/2} \right\rangle^{-1}, \quad (2)$$

where y_i = lab reference value for the i th sample; \hat{y} = NIR measured value; N = number of samples; K = number of wavelengths used in an equation; SD = standard deviation.

The statistics shown in Eqs. (1) and (2) give a more realistic estimate of the applicability of NIRS to the analysis than those of the external validation, as cross-validation avoids the bias produced when a low number of samples representing the full range are selected as validation set (Shenk and Westerhaus, 1996; Williams and Sobering, 1996). The SECV method is based on an iterative algorithm which selects samples from a sample set population to develop the calibration equation and then predicts on the remaining unselected samples. This statistic indicates an estimate of the standard error of prediction (SEP) that may have been found in an external validation (Workman, 1992), and as occurred with SEP is calculated as the square root of the mean square of the residuals for $N-1$ degrees of freedom, where the residual equals the actual minus the predicted value.

4.6. Spectrum of *B. napus* leaf and modified partial least square regression loadings

The MPLS loading plots of the first three factors generated from the MPLS regression (2, 5, 5, 2; SNV + DT) for t-GSL and also the individual glucosinolates, were obtained. MPLS regression constructs its factors

capturing as much of the variation in the spectral data as possible by using the reference values actively during the decomposition of the spectral data. 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 more variation and better correlated to the compound in the calibration set.

The derivative spectrum of the *B. napus* leaf used in this work was obtained by transforming the original absorbance values of all samples at each wavelength (raw optical data from 400 to 2500 nm, every 2 nm) to their second derivative. In addition, the SNV plus DT algorithms were applied. In the second step, the average spectrum was calculated. The second order derivative transformation of the original spectrum resulted in a spectral pattern display of absorption peaks pointing downward.

The information given by the second derivative of the spectrum, together with that information shown by the loadings for the factors of the different equations, were used to identify some of the absorbers employed in modelling these components. In this work we used band assignments from literature, to relate some major absorption bands in the spectrum of “nabicol” leaves with the main wavelengths used by MPLS to construct the first three MPLS terms of the glucosinolate equations.

Acknowledgements

We thank Gloria Fernández Marín, IAS, CSIC, Córdoba, Spain for the performance of the HPLC analyses. This work has been supported by the Project MCYT No. AGL 2003-01366 of the Spanish Government.

References

- Angus, J.F., Gardner, P.A., Kirkegaard, J.A., Desmarchelier, J.M., 1994. Biofumigation: isothiocyanates released from *Brassica* roots inhibit growth of the take all fungus. *Plant Soil* 162, 107–112.
- Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43, 772–777.
- Bell, J.M., 1995. Meal and by-products utilization in animal nutrition. In: Kimber, D.S., McGregor, D.I. (Eds.), *Brassica Oilseeds. Production and Utilization*. CAB International, Wallingford, pp. 301–337.
- Biston, R., Dardenne, P., Cwikowski, M., Marlier, M., Severin, M., Wathélet, J.P., 1988. Fast analysis of rapeseed glucosinolates by near infrared reflectance spectroscopy. *J. Am. Oil. Chem. Soc.* 65, 1599–1600.
- Cartea, M.E., Soengas, P., Picoaga, A., Ordás, A., 2004. Relationships among *Brassica napus* germplasm from Spain and Great Britain as determined by RAPD Markers. *Gen. Res. Crop Evolution*, in press.
- Clark, D.H., Cary, E.E., Mayland, H.F., 1989. Analysis of trace elements in forages by near infrared reflectance spectroscopy. *Agron. J.* 81, 91–95.
- Cochran, W.G., Cox, G.M., 1957. *Experimental Designs*. John Wiley & Sons, New York.
- Daun, J.K., Clear, K.M., Williams, P., 1994. Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed. *J. Am. Oil. Chem. Soc.* 71, 1063–1608.
- Downey, R.K., Röbbelen, G., 1989. *Brassica* species. In: Röbbelen, G., Downey, R.K., Ashri, A. (Eds.), *Oil Crops of the World*. McGraw-Hill, New York, pp. 339–362.
- Ettlinger, M.G., Kjaer, A., 1968. Sulphur compounds in plants. In: Mabry, T.J., Alston, R.E., Runeckles, V.C. (Eds.), *Recent Advances in Phytochemistry*, vol. I. Appleton-Century-Crofts, New York, pp. 89–144.
- Fahey, J.W., Zalcmann, A.T., Talalay, P., 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56, 5–51.
- Fahey, J.W., Zalcmann, A.T., Talalay, P., 2002. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Corrigendum. Phytochemistry* 59, 237.
- Fenwick, R.G., Heaney, R.K., Mullin, W.J., 1983a. Glucosinolates and their breakdown products in food and food plants. *CRC Crit. Rev. Food Sci. Nutr.* 18, 123–201.
- Fenwick, R.G., Griffiths, N.M., Heaney, R.K., 1983b. Bitterness in Brussels sprouts (*Brassica oleracea* L. var. *gemmifera*): the role of glucosinolates and their breakdown products. *J. Sci. Food Agr.* 34, 73–80.
- Font, R., Del Río-Celestino, M., Vélez, D., De Haro-Bailón, A., Montoro, R., 2004a. Visible and near-infrared spectroscopy as a technique for screening the inorganic arsenic content in the red crayfish (*Procambarus clarkii* Girard). *Anal. Chem.* 76, 3893–3898.
- Font, R., Del Río, M., Fernández-Martínez, J.M., De Haro, A., 2004b. Use of near-infrared spectroscopy for screening the individual and total glucosinolate content in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). *J. Agri. Food Chem.* 52, 3563–3569.
- Gislum, R., Micklander, E., Nielsen, J.P., 2004. Quantification of nitrogen concentration in perennial ryegrass and red fescue using near-infrared reflectance spectroscopy (NIRS) and chemometrics. *Field Crops Res.* 88, 269–277.
- Gottlieb, D.M., Schultz, J., Bruun, S.W., Jacobsen, S., Søndergaard, I., 2004. Multivariate approaches in plant science. *Phytochemistry* 65, 1531–1548.
- ISO norm, 1992. Rapessed- Determination of glucosinolates content – Part 1: method using high-performance liquid chromatography. ISO 9167-1, 1–9.
- Louda, S., Mole, S., 1991. Glucosinolates: chemistry and ecology. In: Rosenthal, G.A., Berenbaum, M.R. (Eds.), *Herbivores: Their Interactions with Secondary Plant Metabolites*, vol. 1. Academic Press, New York, pp. 123–164.
- Martens, H., Naes, T., 1989. *Multivariate Calibration*. John Wiley & Sons, New York.
- Mikkelsen, M.D., Petersen, B.L., Olsen, C.E., Halkier, B.A., 2002. Biosynthesis and metabolic engineering of glucosinolates. *Amino Acids* 22, 279–295.
- Murray, I., Williams, P.C., 1987. Chemical principles of near-infrared technology. In: Williams, P., Norris, K. (Eds.), *Near-Infrared Technology in the Agricultural and Food Industries*. American Association of Cereal Chemists, Inc., St. Paul, pp. 17–34.
- Osborne, B.G., Fearn, T., Hindle, P.H., 1993. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific & Technical, Essex.
- Rodríguez, V.M., Padilla, G., Cartea, M.E., Ordás, A., 2003. Evaluación de variedades gallegas de nabicol (*B. napus* var. *pubularia*) en siembra precoz. *Actas de Horticul.* 39, 120–121.

- Rosa, E., 1997. Glucosinolates from flower buds of Portuguese *Brassica* crops. *Phytochemistry* 44, 1415–1419.
- Rosa, E.A.S., Heaney, R.K., Fenwick, G.R., Portas, C.A.M., 1997. Glucosinolates in crop plants. *Hort. Rev.* 19, 99–215.
- Sang, J.P., Minchinton, I.R., Johnstone, P.K., Truscott, R.J.W., 1984. Glucosinolate profiles in the seed, root and leaf tissue of cabbage, mustard, rapeseed, radish and swede. *Can. J. Plant Sci.* 64, 77–93.
- Shapiro, T.A., Fahey, J.W., Wade, K.L., Stephenson, K.K., Talalay, P., 2001. Chemoprotective glucosinolates and isothiocyanates of broccoli sprouts: metabolism and excretion in humans. *Cancer Epidemiol., Biomed. & Prev.* 10, 501–508.
- Shenk, J.S., Westerhaus, M.O., 1995. Analysis of Agricultural and Food Products by Near Infrared Reflectance Spectroscopy. Infracore International, Port Matilda.
- Shenk, J.S., Westerhaus, M.O., 1996. Calibration the ISI way. In: Davies, A.M.C., Williams, P.C. (Eds.), *Near Infrared Spectroscopy: The Future Waves*. Nir Publications, Chichester, pp. 198–202.
- Sørensen, H., 1990. Glucosinolates: structure, properties, function. In: Shahidi, F. (Ed.), *Canola and Rapeseed. Production, Chemistry, Nutrition and Processing Technology*. Van Nostrand Reinhold, New York, pp. 149–172.
- Tkachuk, R., Kuzina, F.D., 1982. Chlorophyll analysis of whole rapeseed kernels by near infrared reflectance. *Can. J. Plant Sci.* 62, 875–884.
- Velasco, L., Becker, H.C., 1998. Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. *Plant Breed.* 117, 97–102.
- Wathelet, J.-P., Wagstaffe, P., Boenke, A., 1991. The certification of the total glucosinolate and sulphur contents of three rapeseeds (colza), CRMs 190, 366 and 367. Commission of the European Communities, report EUR 13339 EN, 1–75.
- Williams, P., 1987. Variables affecting near-infrared reflectance spectroscopy analysis. In: Williams, P., Norris, K. (Eds.), *Near-Infrared Technology in the Agricultural and Food Industries*. American Association of Cereal Chemists, Inc., St. Paul, pp. 143–167.
- Williams, P.C., Sobering, D.C., 1993. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *JNIRS* 1, 25–32.
- Williams, P.C., Sobering, D.C., 1996. How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In: Davies, A.M.C., Williams, P.C. (Eds.), *Near Infrared Spectroscopy: The Future Waves*. Nir Publications, Chichester, pp. 185–188.
- Workman Jr., J.J., 1992. Nir spectroscopy calibration basics. In: Burns, D.A., Ciurczak, E.W. (Eds.), *Handbook of Near-Infrared Analysis*. Dekker Inc., New York, pp. 247–280.