JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

60th Anniversary

Article

FEASIBILITY STUDY ON THE USE OF VISIBLE-NEAR INFRARED SPECTROSCOPY FOR THE SCREENING OF INDIVIDUAL AND TOTAL GLUCOSINOLATE CONTENTS IN BROCCOLI

José Miguel Hernández-Hierro, Juan Valverde, Salvador Villacreces, Kim Reilly, Michael Gaffney, Maria Lourdes González-Miret, Francisco José Heredia, and Gerard Downey

J. Agric. Food Chem., **Just Accepted Manuscript** • DOI: 10.1021/jf3018113 • Publication Date (Web): 11 Jul 2012

Feasibility Study on the Use of Visible-Near Infrared Spectroscopy for the Screening of Individual and Total Glucosinolate Contents in Broccoli

José Miguel Hernández-Hierro[†], Juan Valverde[‡], Salvador Villacreces[‡], Kim Reilly[§], Michael Gaffney[§], Maria Lourdes González-Miret[†], Francisco José Heredia[†], Gerard Downey[‡,*]

[†] Food Colour & Quality Laboratory, Department of Nutrition & Food Science, Universidad de Sevilla, Facultad de Farmacia, 41012 Sevilla, Spain.

[‡] Teagasc, Ashtown Food Research Centre, Dublin 15, Dublin, Ireland.

[§] Horticulture Development Unit, Teagasc Research Centre, Kinsealy, Dublin 17, Ireland.

*** Corresponding author:** Gerard Downey

Phone: (+353 1) 805 9500

Fax: (+353 1) 805 9550

E-mail: Gerard.Downey@teagasc.ie

ABSTRACT

The potential of visible-near infrared spectroscopy to determine selected individual and total glucosinolates in broccoli has been evaluated. Modified partial least squares (MPLS) regression was used to develop quantitative models in order to predict glucosinolate contents. Both the whole spectrum and different spectral regions were separately evaluated to develop the quantitative models, in all cases the best results were obtained using the NIR zone between 2000 and 2498 nm. These models have been externally validated for the screening of glucoraphanin, glucobrassicin, 4-methoxyglucobrassicin, neoglucobrassicin and total glucosinolates contents. In addition, discriminant partial least squares was used to distinguish between two possible broccoli cultivars and showed a high degree of accuracy. In the case of the qualitative analysis, best results were obtained using the whole spectrum (i.e. 400-2498 nm) with a correct classification rate of 100% in external validation being obtained.

KEYWORDS: Glucosinolates; broccoli; visible spectroscopy; near infrared spectroscopy; chemometrics

INTRODUCTION

Plant bioactive compounds, commonly referred to as phytochemicals, are components that can affect cells in ways other than by provision of nutrients. Some of them may have functional properties, meaning that they provide health benefits beyond basic nutrition. It has been suggested that the health-promoting properties of plant foods are due to their content of bioactive components with health-promoting effects [1-3].

Broccoli (*Brassica oleracea* L. var *Italica*) has been produced and consumed in Europe since early times and is a characteristic constituent of European diets. This vegetable contains significant amounts of glucosinolates, which are small organic molecules not essential in the primary metabolism of the plant but with specific functions within its secondary metabolism. Although their role in plants is unclear, their potent odour and pronounced taste suggests a role in herbivore and microbial defence. Structurally, glucosinolates (β-thioglucoside-N-hydroxysulfates) (**Figure 1A**) are characterised by the presence of nitrogen and sulphur groups and they are derived from glucose and an amino acid [1]. Glucosinolates are not bioactive until they have been enzymatically hydrolysed to the associated isothiocyanate by an endogenous myrosinase enzyme that may be released by disruption of the plant cell through harvesting, processing or mastication. Epidemiological studies have consistently reported a reduction in incidence of chronic diseases such as cancer and myocardial infarction as a result of induction of detoxifying enzymes and reduction of oxidative stress, although the anti-nutritive effects of both glucosinolates and hydrolysis products have also been studied [1, 4, 5]. It is therefore important to characterise the content of bioactive compounds in new or re-introduced cultivars.

To this end, the use of rapid methods of analysis with minimal or no sample preparation could be important. Such a capability would significantly enhance the ability of primary

processors to select samples for the market based on a high content of bioactive compounds in a rapid and inexpensive manner.

Near infrared spectroscopy (NIRS) provides fast and non-destructive analysis. It allows qualitative and quantitative analysis to be performed in different matrices thereby reducing costs when compared to wet chemical analysis and without generating waste. In the recent past, the use of NIR spectroscopy in the determination of bioactive compounds in foods and plant natural products has increased considerably [6, 7]. With particular regard to glucosinolates, it has been used to determine these compounds in both seeds [8-13] and leaves [14, 15] from a wide range of *Brassica* species. The standard errors of the glucosinolate's predictions reported in the aforementioned studies ranged from 2.65 to 15.65 μmol g^{-1} DM for three cultivars of rapeseeds [13] and Indian mustard seeds [11] respectively. The results reported in all these works reveal that NIRS is able to predict glucosinolates in a wide range of matrices. Therefore, it may be expected that this technique coupled with chemometric tools could provide an alternative method to undertake the analysis of glucosinolates in broccoli.

The aim of this study was to evaluate the potential of visible-NIR technology to determine both selected individual (glucoraphanin, glucobrassicin, 4-methoxyglucobrassicin, neoglucobrassicin) and total glucosinolates in broccoli. In addition to this, its potential for the discrimination between two broccoli cultivars was evaluated. To our knowledge, this is the first time that vis-NIRS technology has been applied to broccoli for these purposes. The work is part of a wider study aimed at maximising the concentrations of phytochemicals in vegetables by adopting a farm to fork approach. The projected final output of the wider study is the production of vegetables with optimal levels of selected phytochemical groups.

MATERIAL AND METHODS

68 **Samples**

69 Samples were generated at Teagasc, Kinsealy Research Centre, located in the north of
70 County Dublin, Ireland (53° 25' 10" N Lat 6° 10' 45" W). The soil is a grey brown
71 podzolic and presents a loam to clay loam topsoil. The experimental design was 2^3
72 factorial split plot design with four plot replicates. Briefly, eight treatments in 5.5m x
73 3.4 m plots were applied with four replicates combining conventional or organic soil
74 management with organic or conventional pest control. The four combinations were
75 sown with Belstar or Fiesta broccoli cultivars. **Table 1** shows the different agricultural
76 managements applied in this study. The aforementioned factorial trial was carried out
77 during the 2009 and 2010 years and provided 64 samples.
78 Broccoli samples were harvested, frozen and freeze-dried in a large scale freeze drier
79 (Frozen in Time Ltd. York, UK). The specimen chamber was kept below 0 ºC during
80 the whole freeze-drying process. Once freeze-dried, samples were milled (Blixer 4,
81 Robotcoupe, France), vacuum packed in polypropylene bags and stored at -20 ºC prior
82 to analysis. Two aliquots were taken from each sample, one for the micellar
83 electrokinetic capillary chromatography (MEKC) analysis and the other for visible-near
84 infrared analysis.

85 **Chemical Analysis**

86 *Sulphatase Purification Procedure.* Sulphatase (Type H-1 from Helix pomatia, Sigma,
87 MO, USA) was purified by dissolving the sulphatase powder (70 mg) in deionised
88 water (3 mL) and adding ethanol (3 mL). This solution was centrifuged (12000 rpm, 10
89 min, room temperature) and ethanol (9 mL) was added to the supernatant after which
90 the solution was centrifuged again (12000 rpm, 10 min, room temperature). The
91 resulting pellet was dissolved in deionised water (2 mL) and this sulphatase solution
92 was subsequently passed through a 0.5 mL DEAE Sephadex A-25 and a 0.5 mL SP

Sephadex C-25 column. This solution was collected in a vial and kept at -20 °C until use.

Sampling and Extraction. Freeze-dried broccoli powder was extracted using pressurised liquid extraction with an ASE200 instrument (Dionex; Sunnyvale, CA,USA) with attached solvent controller. Extraction of 1.00 g freeze-dried broccoli was carried out in 22mL steel cartridges. Extraction conditions were slightly modified from Mohn *et al* [16]. In brief, preheat time: 1 min; static extraction per cycle: 5 min; flush: 60% of cell volume; purge: 60 s with nitrogen; pressure: 120 bar. Glucotropaeolin (Phytolab GmbH, Vestenbergsgreuth, Germany) was used as an internal standard - 100µL of a solution of 0.5 µmol of potassium salt of glucotropaeolin in Milli –Q water was spiked onto the sample packed in the ASE cell. The collected extract (ca 30 mL) was dried under a constant N_2 flow and re-dissolved in deionised water (7 mL). An aliquot (1 mL) of this crude extract was applied to a DEAE Sephadex A-25 column (0.5 mL) and the unbound material was removed by washing with deionised water (2 x 1 mL) and sodium acetate buffer (2 x 0.5 mL; 20 mM, pH 5.0). After washing, the purified sulphatase (75 µL) was added and the columns were incubated overnight at room temperature. After overnight incubation, the desulphoglucosinolates (dGLS) were eluted from the columns with deionised water (3 x 1 mL). The collected eluate was dried under constant N_2 flow, re-dissolved in deionised water (200 µL) and centrifuged prior to analysis. An aliquot of 50 µL was used for the MEKC analysis.

Micellar Electrokinetic Capillary Chromatography. Analyses were performed using a CE capillary electrophoresis system (Agilent, Waldbronn, Germany) equipped with diode array detector. All separations were performed on a fused silica capillary (Agilent, Stevens Creek, CA; 75 lm ID, 64.5 cm total length, 56 cm effective length). Samples were injected from the anodic end of the capillary (vacuum injection, 50 mbar, 1 s). The

separation buffer consisted of sodium chlorate (250 mM) and boric acid (200 mM) at pH 8.5; the separation was carried out at 12 kV and 60 °C. The capillary was conditioned between each run sequentially with 1.0 M NaOH (3 min), 0.1 M NaOH (1 min), water (1 min) and separation buffer (5 min). Detection was performed on column at 230 and 280 nm. Data processing was carried out with 3D-CE Chemstation software (Agilent Waldbronn, Germany). The quantity of the dGLS was estimated as the average of quantities calculated from the internal standards, taking into account the relative response factors revised by Clarke *et al* [17].

Near Infrared Spectroscopy Analysis

Each aliquot of freeze-dried broccoli powder was tempered at room temperature for 16 hours prior to spectroscopy analysis. A Foss NIRSystems 6500 spectrometer (NIR Systems Inc., Silver Spring MD) was used; samples were placed in small ring cups (3 cm diameter screen size) with a disposable backing disc and spectral measurements were made in reflectance mode over the 400-2498 nm wavelength range. Spectra were recorded in duplicate at intervals of 2 nm with 10 scans being performed for the reference tile and 25 for samples; the sample cell was rotated through 180 degrees between spectral collections. To minimise sampling error, all the samples were analysed in duplicate with re-filling of the ring cup. Therefore, the four recorded spectra of each sample were averaged to obtain each sample spectrum. Samples were scanned in random order. The spectrum of sinigrin standard powder ((−)-sinigrin hydrate from horseradish, Sigma-Aldrich Ireland Ltd, Arklow, Ireland) was also recorded in reflectance mode over the wavelength range 1100-2498 nm at intervals of 2 nm. The software used was Win ISI® (v1.50) (Infrasoft International, LLC, Port. Matilda, PA, USA). This software allowed not only the instrument control and spectral acquisition but also the data pre-treatment and development of qualitative and quantitative models.

From the whole data set, 25% of the samples (16 samples) were randomly allocated into the validation set and the remaining 75% (48 samples) into the calibration set.

Chemometric Techniques

A supervised pattern recognition technique-Discriminant Partial Least Squares (DPLS)- was used for qualitative analysis. The calibration was conducted by performing a regression on the spectral information and the associated dummy sample group values, in this case defined as 1 (variety 1) or 2 (variety 2). The regression method applied to this procedure was MPLS, which is a modification of a normal PLS 1 [18, 19].

Prior to quantitative analysis, an unsupervised pattern recognition technique, principal component analysis (PCA), was used in order to provide information about the latent structure of spectral data. This method provides not only information related to spectral outliers and the distribution of samples in the newly-created space but is also an important source of knowledge with which to create cross-validation groups used in the calibration process. PCA is also a useful tool to identify whether unknown samples belong to the spectral space created by the samples from which the equations were developed. Should this not be the case, the equations should not be used to make any prediction [20, 21]. Using the raw spectral data and testing different spectral pre-treatments and allocating the corresponding MEKC-DAD glucosinolate values to each sample, calibrations were performed by modified partial least squares regression (MPLS). In this method, the set of calibration samples is divided into a series of subsets in order to perform cross-validation to set the number of PLS factors, reduce the likelihood of overfitting [20] and remove chemical outliers. Using the $T \geq 2.5$ criterion, samples that presented a high residual value when they were predicted were eliminated from the set. Finally, validation errors are combined into a single figure, the standard error of cross-validation (SECV).

Spectral pre-treatments are usually applied to NIR raw data; scattering effects were removed using multiplicative scatter correction (MSC), standard normal variate (SNV), and detrending [22, 23]. Moreover, the effect of derivatisation and variations in spectral ranges were tested in the development of the NIRS calibrations.

RESULTS AND DISCUSSION

Chemical Analysis

Glucosinolates were determined by MEKC-DAD. The common glucosinolate skeleton structure is shown in **Figure 1A**, leading to the expectation that their spectral signatures should also be very similar. Sinigrin's reflectance spectrum (**Figure 1B**) shows considerable spectral detail in the wavelength range above about 2000 nm; it was the only pure glucosinolate available for spectral analysis but, given that its structure is similar to the others being analysed in these broccoli samples, it is expected that all glucosinolates will have similar information above 2000 nm. **Table 2** shows the range, mean value and standard deviation of the selected individual glucosinolates and the total glucosinolates in the broccoli samples analysed. These values were used in the development of the quantitative models as reference values. It is noticeable that the precision of the laboratory methods is quite poor with standard deviation between replicate figures of between 15.9% (neoglucobrassicin) and 27.2% (glucoraphanin) of mean contents.

Near Infrared Spectroscopy Analysis and Chemometric Techniques

Near Infrared Spectra. **Figure 1** shows the average and standard deviation spectra of broccoli powders over the 400-2498 nm range (**Figure 1C**) together with the sinigrin powder spectrum between 1100-2498 nm (**Figure 1B**).

Standard deviation spectra have been multiplied by a factor of 10 for display reasons. Spectral intensities were low and well within the linear response range of the instrument detector range; little noise was evident in these traces. A strong feature of the sample spectra was the absorbance pattern in the visible wavelength range i.e. 400-780 nm. This arose from the pigment remaining in the freeze-dried samples, all of which displayed a green tinge.

Figure 1D shows the second derivative spectra of broccoli and sinigrin and presents considerable spectral similarities over the 2000-2498 nm wavelength range. It may be expected that this range would be useful for quantitative purposes of glucuosinolates.

Qualitative analysis. An SNV 2,4,4,1 spectral pre-treatment was applied to whole vis-NIR spectra of samples from the calibration set and then a principal component analysis was carried out. **Figure 2** shows the scores of the broccoli samples in the space defined by the first and second principal components which described 52.3% (PC1) and 23.7% (PC2) of the variability in the data. In this plot, slight differences between the 2009 and 2010 harvest years (**Figure 2A**) and also between the Belstar and Fiesta cultivars (**Figure 2B**) are apparent. The main difference observed in these plots was between harvest years although the separation between 2009 and 2010 samples was not complete. This trend in harvest year locations was discernible on the basis of PC2. Cultivars were completely overlapped in this hyperspace.

D-PLS modelling of the groups was carried out initially using the complete vis-NIR raw spectra and one dummy variable, the values of which were defined as 1 (variety 1=Belstar) or 2 (variety 2=Fiesta). The model was developed using 6 PLS factors and presented an RSQ of 0.752, SEC of 0.27 and SECV 0.34. The spectral regions between 600-700 nm and 2000-2498 nm showed important contributions to the model loadings (**Figure 3A**) and are mainly related to green pigments such as chlorophyll [14, 15, 24] and C-

H, O-H and N-H overtones present in a number of bonds [25, 26]. The DPLS developed model predicted a dummy value for each sample and then the samples were allocated according to their predicted values ±0.5 in the corresponding cultivar with a dummy variable breakpoint of 1.5. The predicted dummy variable plot for the qualitative model is shown in **Figure 3B**. Samples of Belstar cultivar cluster around a dummy value of 1.0, while those belonging to Fiesta cultivar around 2.0. A prediction rate of 100% samples correctly classified was obtained in external validation. This demonstrates that vis-NIR spectroscopy coupled to discriminant DPLS analysis permit a clear differentiation between the two aforementioned broccoli cultivars. Regarding the studied compound, the classification of samples according to their cultivar could be useful since they present significant differences in the 4-methoxyglucobrassicin ($p<0.001$) and neoglucobrassicin ($p<0.001$) contents.

Quantitative analysis. Although different spectral ranges and the whole spectrum have been separately evaluated, in all cases the best results were obtained using the wavelength range 2000-2498 nm. As part of the quantitative analysis, a SNV (2,4,4,1) spectral pre-treatment was applied to this spectral range of samples in the calibration set and then principal component analysis was carried out in order to look for spectral outliers and create cross-validation groups. Overall, the spectral variability explained was 98% using 13 principal components and Mahalanobis distances for each sample were calculated. Samples were ranked in order of their H (Mahalanobis) distance from the mean spectrum of the entire sample set and the $H > 3$ criterion was applied. No H-outliers were found.

The average spectrum of the best of the different mathematically pre-treated spectra are shown in **Figure 4**. The statistical parameters of the final calibration equations are shown in **Table 3** where N is the number of samples used to obtain the calibration

241 equation after eliminating samples for chemical reasons (T criterion). The best of the
242 different mathematical treatments, concentration range, and standard deviations are also
243 shown. We checked the robustness of the method by applying NIRS technology to 16
244 samples that did not belong to the calibration group. **Table 3** also shows the results
245 obtained in the external validation and the SEP values are presented. The percentage of
246 error for total glucosinolates is similar to the errors previously reported for other
247 matrices such as Indian mustard seeds[11].

248 The magnitude of these errors indicate the potential of the NIR spectroscopy for
249 prediction of glucosinolates in freeze-dried broccoli and is satisfactory for the screening
250 of glucoraphanin, glucobrassicin, 4-methoxyglucobrassicin, neoglucobrassicin and total
251 glucosinolates. However, individual glucosinolates biosynthesis may follow essentially
252 the same course so inter-correlations among them could be expected. It is not possible
253 to ascertain if the results of NIRS models for predicting the composition of these
254 glucosinolates were due to their real absorbance or the correlation between them.
255 **Figure 5** shows the loading and the beta regression coefficients plots of the MPLS
256 model for total glucosinolates. The spectral regions around 2072 nm, 2220 nm, 2300 nm
257 and 2446 nm show important contributions to the model loadings. These could be
258 related to N-H bonds (2072 nm), to C-H bonds (2220 nm), to first overtones of O-H
259 alcohol functional group, to stretching-bending of CH-CH_2 bonds and C-O bonds (2300
260 nm) and to C-N bonds (2446 nm). These can be attributed to the chemical structure of
261 the compounds analyzed [25, 26]. This confirms previous studies that showed important
262 contributions in the aforementioned spectral zones for determining glucosinolates in
263 other matrices [11, 14, 15, 27].

264 The potential of vis-NIRS for determination of glucosinolates in freeze-dried broccoli
265 was examined. The procedure reported here presents a good potential for a fast and

reasonably inexpensive screening of almost all individual glucosinolates and total glucosinolates present in broccoli and also to distinguish between broccoli cultivars. Although different spectral ranges and the whole spectrum have been separately evaluated to develop the quantitative models, in all cases the best results (i.e. the highest RSQ and the lowest SECV and SEP values) were obtained using the NIR region between 2000 and 2498 nm. In the case of the qualitative analysis, best results were obtained used the whole spectrum (i.e. 400-2498 nm). Nonetheless, a comprehensive study should be made in order to evaluate other factors such as different production areas and varieties, in the complete development of these models.

ACKNOWLEDGEMENTS

Authors are grateful to Jens C. Sørensen and Hilmer Sørensen from the Department of Basic Sciences and Environment, Chemistry and Biochemistry, Faculty of Life Sciences, University of Copenhagen in Denmark for their useful help on setting up the Micellar Electrokinetic Capillary Chromatography method and for sharing their expert knowledge on the glucosinolate/myrosinase system.

REFERENCES

(1) Crozier, A.; Clifford, M. N.; Ashihara, H., *Plant secondary Metabolites. Occurrence, Structure and Role in the Human Diet*. Blacwell Publishing: Garsington Road, Oxford, UK, 2006.

(2) Sloan, A. E., Top 10 functional food trends. *Food Technol-Chicago* **2008,** *62*, 25-44.

(3) Wildman, R. E. C., *Handbook of Nutraceuticals and Functional Foods*. CRC Press LLC: Boca Raton, FL, USA, 2001.

(4) Shahindi, F., *Canola and rapeseed production, chemistry and processing technology*. Van Nostrand Reinhold: New York, USA, 1990.

(5) Verkerk, R.; Schreiner, M.; Krumbein, A.; Ciska, E.; Holst, B.; Rowland, I.; De Schrijver, R.; Hansen, M.; Gerhäuser, C.; Mithen, R.; Dekker, M., Glucosinolates in Brassica vegetables: The influence of the food supply chain on intake, bioavailability and human health. *Mol. Nutr. Food Res.* **2009,** *53*, S219-S219.

(6) McGoverin, C. M.; Weeranantanaphan, J.; Downey, G.; Manley, M., The application of near infrared spectroscopy to the measurement of bioactive compounds in food commodities. *J. Near Infrared Spectrosc.* **2010,** *18*, 87-111.

(7) Cozzolino, D., Near Infrared Spectroscopy in Natural Products Analysis. *Planta Med.* **2009,** *75*, 746-756.

(8) 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.

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

(10) Hom, N.; Becker, H.; Möllers, C., Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica* **2007,** *153*, 27-34.

(11) Font, R.; del Rio, M.; Fernandez-Martinez, J. M.; de Haro-Bailon, A., Use of Near-Infrared Spectroscopy for Screening the Individual and Total Glucosinolate Contents in Indian Mustard Seed (*Brassica juncea* L. *Czern. & Coss*.). *J. Agric. Food Chem.* **2004,** *52*, 3563-3569.

(12) Font, R.; del Río-Celestino, M.; de Haro-Bailon, A., The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crop. Prod.* **2006,** *24*, 307-313.

(13) Petisco, C.; García-Criado, B.; Vázquez-de-Aldana, B. R.; de Haro, A.; García-Ciudad, A., Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy. *Ind. Crop. Prod.* **2010,** *32*, 139-146.

(14) Font, R.; del Rio Celestino, M.; Rosa, E.; Aires, A.; de Haro-Bailon, A., Glucosinolate assessment in Brassica oleracea leaves by near-infrared spectroscopy. *J. Agric. Sci.* **2005,** *143*, 65-73.

(15) Font, R.; del Rio-Celestino, M.; Cartea, E.; de Haro-Bailon, A., Quantification of glucosinolates in leaves of leaf rape (Brassica napus ssp. pabularia) by near-infrared spectroscopy. *Phytochemistry* **2005,** *66*, 175-185.

(16) Mohn, T.; Cutting, B.; Ernst, B.; Hamburger, M., Extraction and analysis of intact glucosinolates--A validated pressurized liquid extraction/liquid chromatography-mass spectrometry protocol for Isatis tinctoria, and qualitative analysis of other cruciferous plants. *J. Chromatogr. A* **2007,** *1166*, 142-151.

(17) Clarke, D. B., Glucosinolates, structures and analysis in food. *Anal. Methods* **2010,** *2*, 310-325.

(18) Lars, S.; Svante, W., Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *J. Chemometr.* **1987,** *1*, 185-196.

(19) Massart, D. L.; Vandeginste, B. G.; Buydens, L. M.; Lewi, P. J.; Smeyers-Verbeke, J.; Jong, S. D., *Handbook of Chemometrics and Qualimetrics*. Elsevier Science Inc.: 1998; p 1616.

(20) Shenk, J. S.; Westerhaus, M. O., *Routine Operation,Calibration, Development and Network System Management Manual, NIRSystems*. Silver Spring: MD, USA, 1995; p 239.

(21) Brereton, R. G., *Chemometrics : data analysis for the laboratory and chemical plant*. J. Wiley: Chichester, West Sussex, England, 2003; p 489.

(22) Geladi, P.; MacDougall, D.; Martens, H., Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl. Spectrosc.* **1985,** *39*, 491-500.

(23) Dhanoa, M. S.; Lister, S. J.; Barnes, R. J., On the Scales Associated with Near-Infrared Reflectance Difference Spectra. *Appl. Spectrosc.* **1995,** *49*, 765-772.

(24) Tkachuk, R.; Kuzina, F. D., Chlorophyll analysis of whole rapeseed kernels by near infrared reflectance. *Can. J. Plant Sci.* **1982,** *62*, 875-884.

(25) Osborne, B. G.; Fearn, T.; Hindle, P. T., *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman Scientific & Technical ; Wiley: Harlow, Essex, England : New York :, 1993; p 224.

(26) Siesler, H. W.; Ozaky, Y.; Kawata, S.; Heise, H. M., *Near infrared spectroscopy: principles, instruments, applications*. Wiley-VCH: Weinheim, Germany, 2002; p 348.

(27) Tkachuk, R., Oil and Protein-Analysis of Whole Rapeseed Kernels by near-Infrared Reflectance Spectroscopy. *J. Am. Oil Chem. Soc.* **1981,** *58*, 819-822.

Acknowledgments for funding

J.M. Hernández-Hierro thanks the Spanish MICINN for the Juan de la Cierva contract (JCI-2011-09201) and Universidad de Sevilla for the mobility grant (Universidad de Sevilla Research Plan). Spanish MICINN project AGL2011-30254-C02 and Junta de Andalucia PGC project AGR 6331.

Figure captions

Figure 1. (A) Glucosinolete skeleton. (B) Sinigrin powder spectrum (1100-2498 nm). (C) Average and standard deviation spectra (400-2498 nm) of broccoli powders. (D) 2nd derivative spectra of the broccoli and sinigrin (1100-2000 nm).

Figure 2. Score plot of broccoli samples (calibration set) in the space defined by PC1 and PC2.

Figure 3. (A) Loading plot of the DPLS model. (B) Dummy variable plot of broccoli samples in the external validation of the DPLS method.

Figure 4. Average spectrum of different mathematical pre-treatments applied to broccoli powder spectra (2000-2498 nm).

Figure 5. (A) Loading and (B) beta regression coefficients plots of the MPLS model for total glucosinolate prediction.

Table 1. Different Agricultural Managements Applied in this Study.

	Organic	Conventional
Soil treatment	• Four year rotation ley: red clover-broccoli-onion-carrot • Additional organic fertilisation as indicated by soil test • Winter cover crop	• No rotation, plots, randomly allocated each year • Mineral fertilisers as indicated by soil test • No ley crop • No winter crop
Pest control	• Certified organic seed • Refuge area • *Brassica* collars on broccoli • Certified organic chemicals (eg. garlic spray) • Weed control by mechanical methods	• Chemically treated seed • Chemical weed control–herbicides • Chemical pest control–fungicides and insecticides

Table 2. Statistical Descriptors of Glucosinolate Contents Determined by MEKC-DAD (μmol g^{-1} DM).

	GBS[a]	GRA[b]	4MGBS[c]	NGBS[d]	TGS[e]
Min (n=64)	1.60	1.09	0.09	0.74	4.74
Max (n=64)	6.83	4.82	0.55	4.45	15.10
Mean (n=64)	3.92	2.83	0.28	1.97	9.53
SD (n=64)	1.09	0.87	0.10	0.86	1.87
SD replicates	0.88	1.01	0.08	0.59	1.80

[a]GBS: glucobrassicin; [b]GRA: glucoraphanin; [c]4MGBS: 4-methoxyglucobrassicin; [d]NGBS: neoglucobrassicin; [e]TGS: total glucosinolates

Table 3. Calibration Statistical Descriptors for the Models Developed in the NIR Zone Close to 2000-2498 nm.

Spectral pre-treatments	Compounds	T outliers	PLS factors	N[a]	Mean	Est. Min	SD[b]	Est. Max	SEC[c]	RSQ[d]	SECV[e]	SEP[f]
					(μmol g^{-1} DM)						(μmol g^{-1} DM)	
Detrend 1,4,4,1	GBS[g]	2	7	46	3.93	0.68	1.09	7.18	0.35	0.89	0.52	0.74
Detrend 1,4,4,1	GRA[h]	1	2	47	2.86	0.26	0.87	5.46	0.67	0.40	0.71	0.75
SNV 1,4,4,1	4MGBS[i]	0	5	48	0.28	0.00	0.11	0.62	0.06	0.69	0.07	0.05
MSC 2,8,6,1	NGBS[j]	0	5	48	1.93	0.00	0.86	4.51	0.48	0.68	0.66	0.74
None 2,10,10,1	TGS[k]	3	5	45	9.40	4.04	1.79	14.76	0.93	0.73	1.11	1.44

[a]N: number of samples (calibration set); [b]SD: standard deviation; [c]SEC: standard error of calibration; [d]RSQ: coefficient of determination; [e]SECV: standard error of cross-validation; [f]SEP: standard error of prediction; [g]GBS: glucobrassicin; [h]GRA: glucoraphanin; [i]4MGBS: 4-methoxyglucobrassicin; [j]NGBS: neoglucobrassicin; [k]TGS: total glucosinolates

Figure 1.

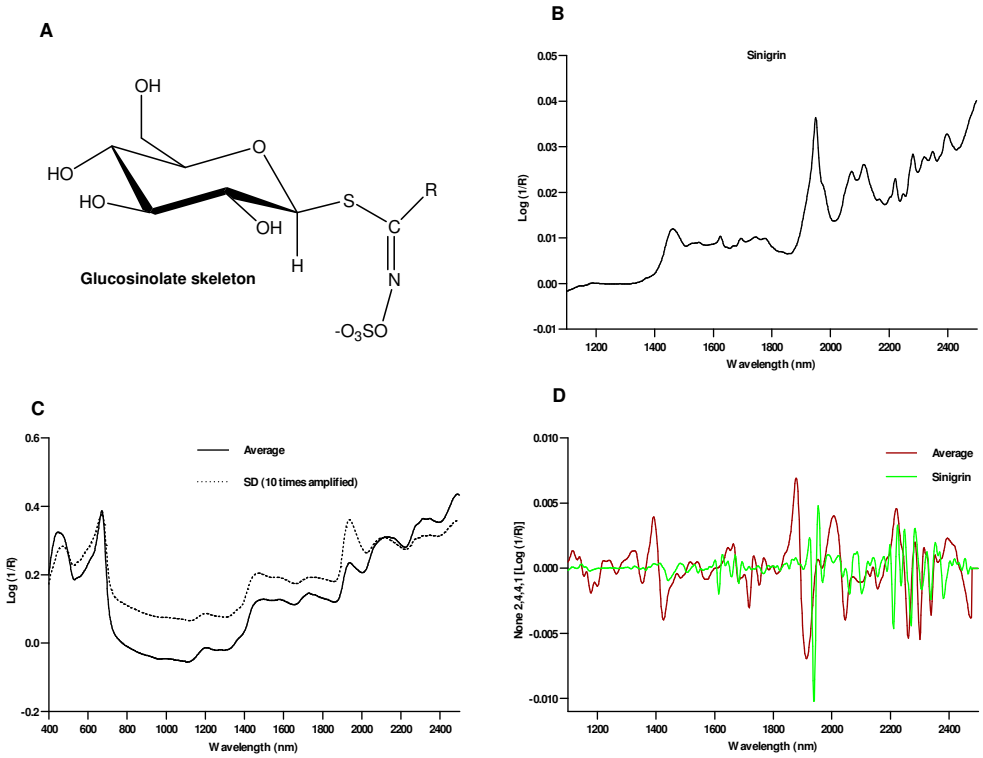

A
OH
HO
HO
O
OH
H
S
R
C
N
-O3SO
Glucosinolate skeleton
B
Sinigrin
Log (1/R)
Wavelength (nm)
C
Average
SD (10 times amplified)
Log (1/R)
Wavelength (nm)
D
Average
Sinigrin
None 2,4,4,1 [Log (1/R)]
Wavelength (nm)

Figure 2.

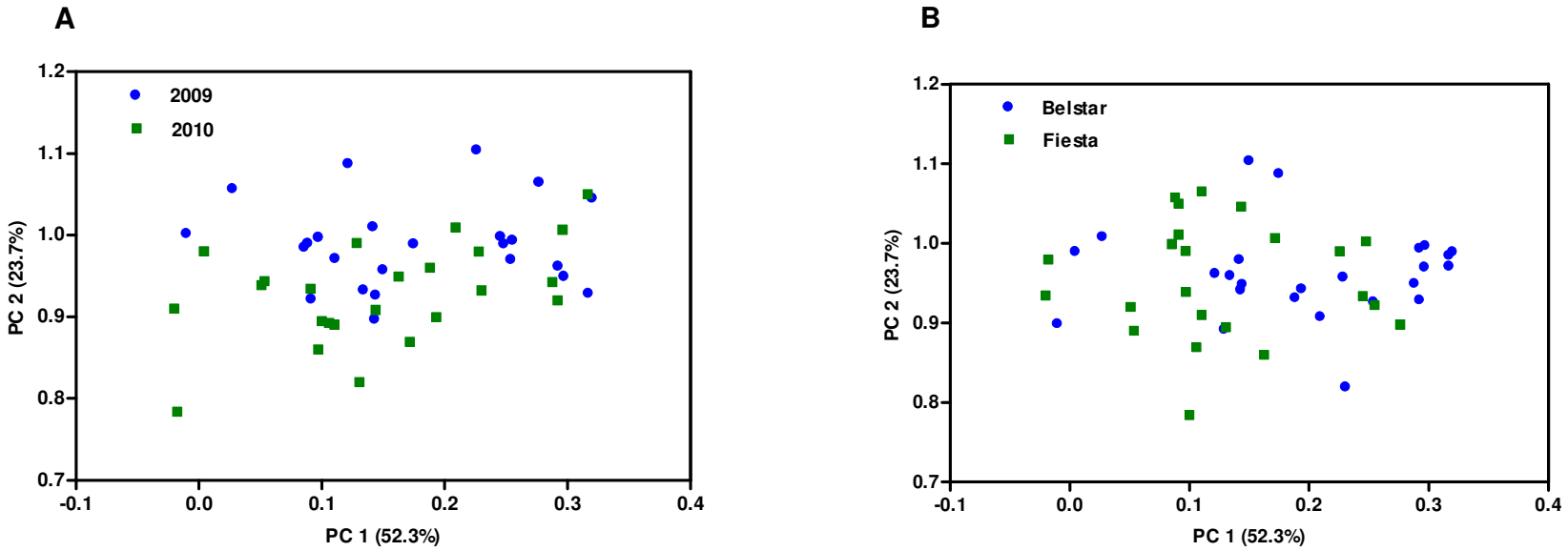

A
2009
2010
PC 2 (23.7%)
PC 1 (52.3%)
1.2
1.1
1.0
0.9
0.8
0.7
-0.1
0.0
0.1
0.2
0.3
0.4
B
Belstar
Fiesta
PC 2 (23.7%)
PC 1 (52.3%)

Figure 3.

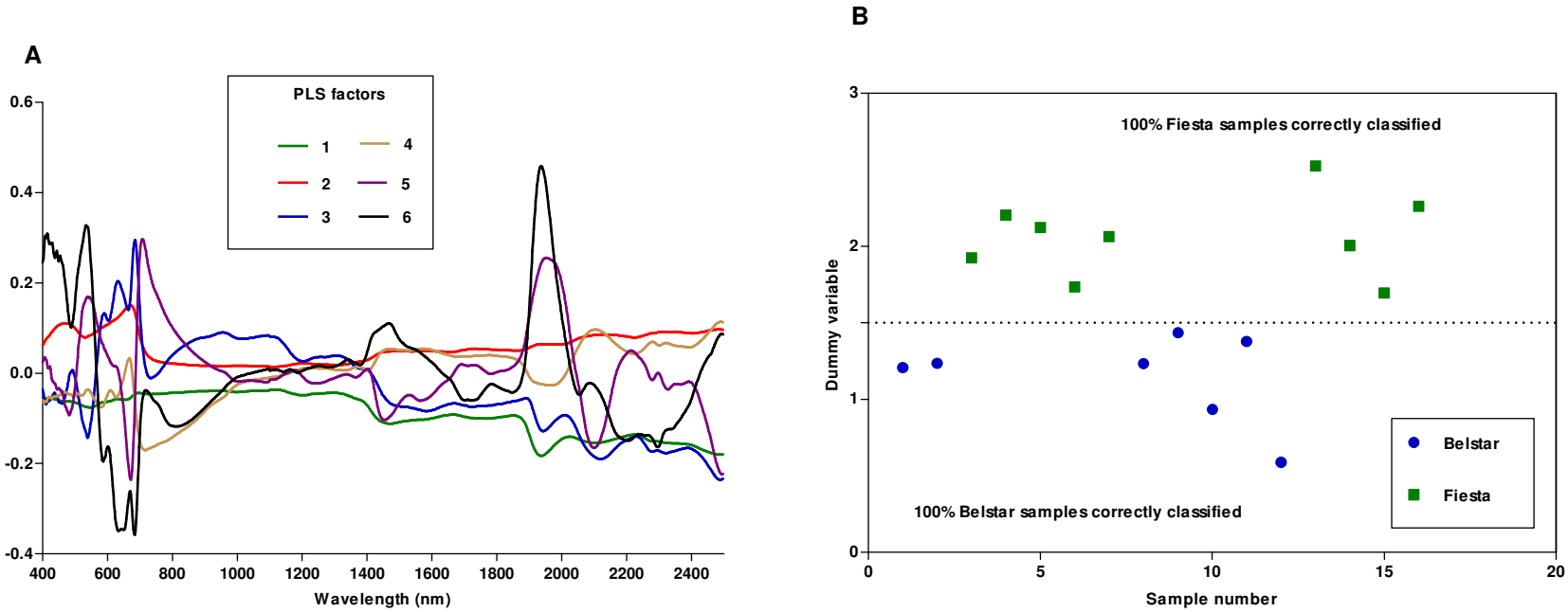

A
PLS factors
1
2
3
4
5
6
0.6
0.4
0.2
0.0
-0.2
-0.4
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
Wavelength (nm)
B
100% Fiesta samples correctly classified
100% Belstar samples correctly classified
Belstar
Fiesta
Dummy variable
Sample number
0
1
2
3
0
5
10
15
20

Figure 4.

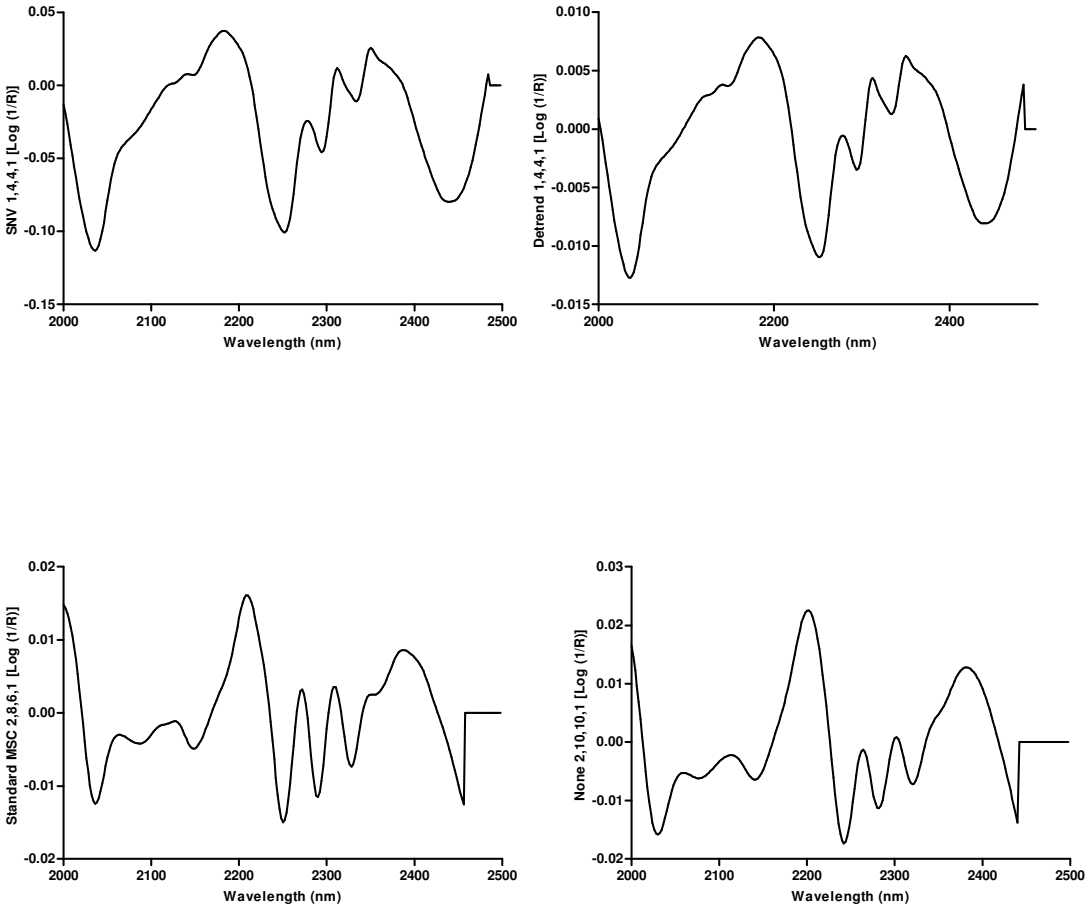
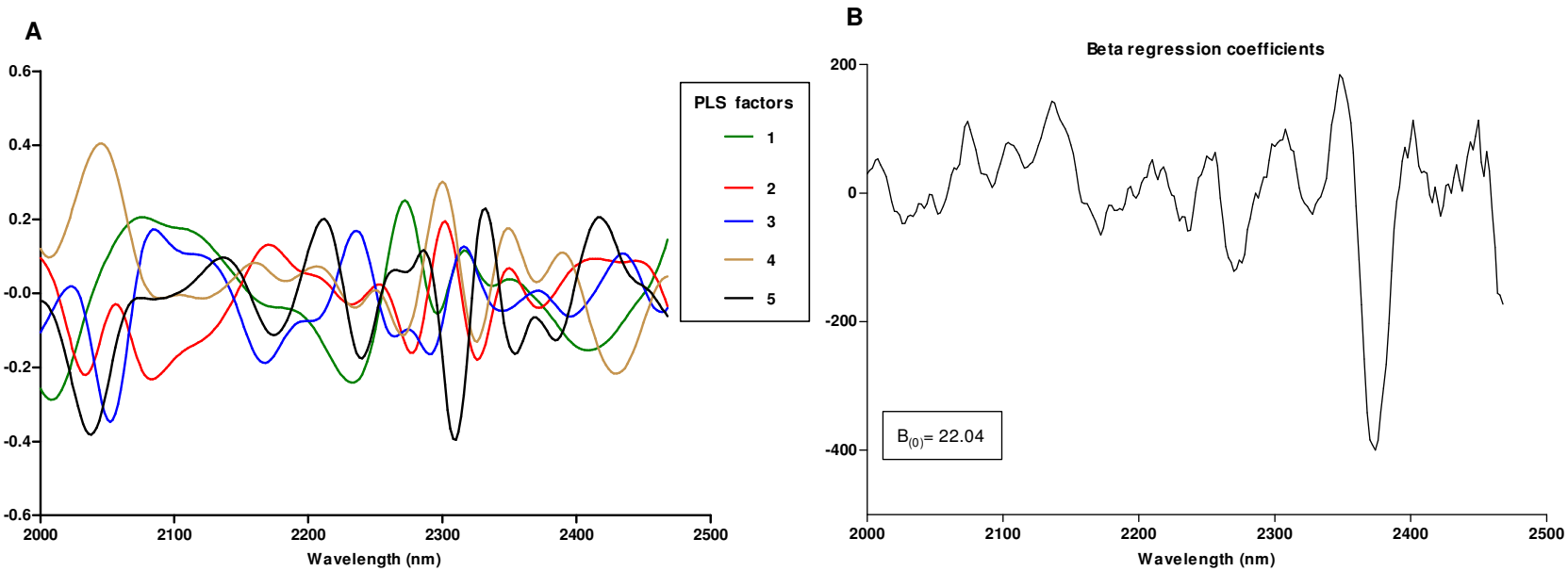

SNV 1,4,4,1 [Log (1/R)]
0.05
0.00
-0.05
-0.10
-0.15
2000
2100
2200
2300
2400
2500
Wavelength (nm)
Detrend 1,4,4,1 [Log (1/R)]
0.010
0.005
0.000
-0.005
-0.010
-0.015
2000
2200
2400
Wavelength (nm)
Standard MSC 2,8,6,1 [Log (1/R)]
0.02
0.01
0.00
-0.01
-0.02
2000
2100
2200
2300
2400
2500
Wavelength (nm)
None 2,10,10,1 [Log (1/R)]
0.03
0.02
0.01
0.00
-0.01
-0.02
2000
2100
2200
2300
2400
2500
Wavelength (nm)

Figure 5.


A
0.6
0.4
0.2
-0.0
-0.2
-0.4
-0.6
2000
2100
2200
2300
2400
2500
Wavelength (nm)
PLS factors
1
2
3
4
5
B
Beta regression coefficients
200
0
-200
-400
B(0)= 22.04
2000
2100
2200
2300
2400
2500
Wavelength (nm)

Graphic for table of contents