



Preliminary study on the use of near infrared hyperspectral imaging for quantitation and localisation of total glucosinolates in freeze-dried broccoli

José Miguel Hernández-Hierro, Carlos Esquerre, Juan Valverde, Salvador Villacreces, Kim Reilly, Michael Gaffney, M. Lourdes González-Miret, Francisco J. Heredia, Colm P. O'Donnell, Gerard Downey

This article is provided by the author(s) and Teagasc T-Stór in accordance with publisher policies.

Please cite the published version.

The correct citation is available in the T-Stór record for this article.

NOTICE: This is the author's version of a work that was accepted for publication in *Journal of Food Engineering*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Journal of Food Engineering*, Volume 126, April 2014, Pages 107-112. DOI 10.1016/j.jfoodeng.2013.11.005.

This item is made available to you under the Creative Commons Attribution-Non commercial-No Derivatives 3.0 License.



**Preliminary study on the use of near infrared hyperspectral imaging
for quantitation and localisation of total glucosinolates in freeze-dried
broccoli**

José Miguel Hernández-Hierro¹, Carlos Esquerre², Juan Valverde³, Salvador
Villacreces³, Kim Reilly⁴, Michael Gaffney⁴, M. Lourdes González-Miret¹, Francisco J.
Heredia¹, Colm P. O'Donnell², Gerard Downey^{3,*}

¹Food Colour & Quality Laboratory, Department of Nutrition & Food Science,
Universidad de Sevilla, Facultad de Farmacia, 41012 Sevilla, Spain.

²Biosystems Engineering, School of Agriculture, Food Science and Veterinary
Medicine, University College Dublin, Belfield, Dublin 4, Ireland.

³Teagasc Food Research Centre Ashtown, Dublin 15, Dublin, Ireland.

⁴Horticulture 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

1 **Abstract**

2 The use of hyperspectral imaging to (a) quantify and (b) localise total glucosinolates in
3 florets of a single broccoli species has been examined. Two different spectral regions
4 (vis-NIR and NIR), a number of spectral pre-treatments and different mask development
5 strategies were studied to develop the quantitative models. These models were then
6 applied to freeze-dried slices of broccoli to identify regions within individual florets
7 which were rich in glucosinolates. The procedure demonstrates potential for the
8 quantitative screening and localisation of total glucosinolates in broccoli using the 950-
9 1650 nm wavelength range. These compounds were mainly located in the external part
10 of florets.

11 **Keywords:** Glucosinolates, broccoli, hyperspectral imaging, near infrared, visible,
12 chemometrics.

13 **1. Introduction**

14 Glucosinolates are a class of about 120 chemicals distributed in only 16 plant families.
15 These compounds are well-known for their characteristic pungent smells and tastes
16 which are typical of some *Brassica* vegetables such as cabbage, mustard, cress,
17 cauliflower, broccoli, turnip, Brussel sprouts, radish and horseradish. Structurally,
18 glucosinolates (β -thioglucoside-N-hydroxysulfates) are characterised by the presence of
19 nitrogen and sulphur groups. Biosynthesis of glucosinolates is mainly carried out using
20 glucose and amino acids such as methionine, alanine, leucine and valine (aliphatic
21 glucosinolates) or tryptophan and phenylalanine (aromatic glucosinolates) (Crozier et
22 al., 2006). Epidemiological studies have consistently reported a reduced incidence of a
23 number of diseases in subjects consuming diets rich in these compounds although anti-
24 nutritive effects of both glucosinolates and their hydrolysis products have also been
25 reported (Crozier et al., 2006; Jeffery and Araya, 2009; Shahindi, 1990; Verkerk et al.,
26 2009). Broccoli (*Brassica oleracea* L. var *Italica*) contains significant amounts of these
27 potentially bioactive compounds (Vallejo et al., 2003; Wang et al., 2012a). This
28 vegetable is an economically important crop in a number of countries which may act as
29 a source not only of glucosinolates but also of vitamins, minerals and other beneficial
30 phytochemicals (Jeffery et al., 2003; Wang et al., 2012a). It is therefore important to
31 both characterise the content of bioactive compounds in broccoli and determine in what
32 parts of the plant these bioactive compounds are accumulated. Our previous work has
33 demonstrated the potential for near infrared spectroscopy to quantify total
34 glucosinolates in freeze-dried powders with acceptable accuracy for screening purposes
35 (Hernandez-Hierro et al., 2012). Information about the spatial distribution of the
36 aforementioned compounds would also be useful but near infrared spectroscopy does

37 not provide the capability to map the location of constituents. Hyperspectral imaging
38 may hold the answer to this problem.

39 Hyperspectral imaging is an emerging technique for non-destructive food analysis
40 which provides both spatial and spectral information about an object. Recorded images
41 consist of many thousands of pixels in a two-dimensional array, with each pixel
42 corresponding to a specific region on the surface of the sample; each pixel in a
43 hyperspectral image therefore contains a spectrum of the sample at that specific
44 position. Interrogation of these spectra makes possible the development of mathematical
45 models to predict the chemical composition or functional class of a sample at each
46 pixel. Reflectance imaging is the most common image acquisition mode and is usually
47 carried out in either the visible-near infrared (vis-NIR; 400-1000 nm) or near infrared
48 (NIR; 1000-1700 nm) spectral regions (Gowen et al., 2007). The use of multivariate
49 chemometric methods is required to handle the large quantities of spectral data collected
50 in each image and very many approaches are available for the development of
51 regression models to predict constituent concentrations in a sample at pixel level.

52 The number of research applications of hyperspectral analysis has risen considerably in
53 the food sector in the recent past (Burger and Gowen, 2011; Gowen et al., 2007;
54 Lorente et al., 2012; McGoverin et al., 2010; Sun, 2010). Hyperspectral image analysis
55 has been used to determine moisture, total soluble solids and pH in strawberries
56 (ElMasry et al., 2007), firmness and soluble solids in apples (Mendoza et al., 2011;
57 Wang et al., 2012b) anthocyanins in grape skins (Fernandes et al., 2011), chlorophyll
58 distribution in cucumber leaves (Ji-Yong et al., 2012) and maturity stage of bananas
59 (Rajkumar et al., 2012). Additionally, this analytical method has been used to determine
60 moisture in dehydrated prawns (Wu et al., 2012) and some quality parameters of both
61 lamb (Kamruzzaman et al., 2012) and pork (Barbin et al., 2012) meat. Results reported

62 in these studies have indicated that hyperspectral imaging is able to predict a number of
63 food components and quality parameters in a wide range of biological matrices.

64 The aim of this study was to evaluate the potential of hyperspectral imaging technology
65 for the quantitative screening and localisation of total glucosinolates in freeze-dried
66 broccoli. Since predictive models developed on freeze-dried powders by conventional
67 NIR spectrometers may not be transferred directly to hyperspectral imaging datasets, a
68 new predictive model must be generated using an actual hyperspectral imaging system
69 on homogeneous, freeze-dried broccoli powders after which it may be applied to
70 hyperspectral images of intact broccoli for localisation and quantitation of total
71 glucosinolates. To our knowledge, this is the first time that this analytical tool has been
72 applied to broccoli for these purposes.

73 **2. Material and methods**

74 *2.1. Samples and chemical analysis*

75 Sixty-four broccoli samples were grown at Teagasc (Kinsealy Research Centre, Dublin)
76 using different agricultural managements (see (Hernandez-Hierro et al., 2012)) ,
77 cultivars (Belstar and Fiesta) and years (2009 and 2010). Samples were harvested,
78 immediately frozen then freeze-dried. Once freeze-dried, samples were milled (Blixer 4,
79 Robot Coupe, France), vacuum-packed in polypropylene bags and stored at -20°C prior
80 to analysis. (Hernandez-Hierro et al., 2012). Two aliquots were taken from each sample,
81 one for the micellar electrokinetic capillary chromatography (MEKC) analysis and the
82 remaining aliquot for hyperspectral imaging analysis in order to develop the quantitative
83 model on powders. Separately, whole slices of broccoli were frozen and freeze-dried in
84 order to apply the constructed model for glucosinolate localisation studies.

85 Total glucosinolates were expressed as the sum of the individual glucosinolates
86 previously determined by a MEKC-DAD method and the results obtained in this earlier

87 work were used as reference data for the development of quantitative models
88 (Hernandez-Hierro et al., 2012). Total glucosinolate contents in the broccoli samples
89 ranged from 4.74 to 15.10 $\mu\text{mol g}^{-1}$ dry matter (DM) with mean value and standard
90 deviation values of 9.53 and 1.87 $\mu\text{mol g}^{-1}$ DM respectively.

91 *2.2. Hyperspectral imaging analysis*

92 Two different spectral regions using two different devices (HSI instrument DV Optics
93 Ltd, Padua, Italy) were used in this study. The main components of these systems,
94 including the illumination source, diffuser, moving base, optics (mirror and lens),
95 spectrograph, camera and computer are shown in **Fig. 1**. System 1 consisted of a high
96 performance CCD Basler A312f camera (580x580 pixels; Basler, Ahrensburg,
97 Germany), a spectrograph (Specim V10E) (Spectral Imaging Ltd., Oulu, Finland)
98 attached to the camera covering the spectral range between 400 and 1000 nm (spectral
99 resolution of 5 nm), a zoom lens (16 mm focal length), a halogen lamp (150W) light
100 source transmitted via fibre optics and a cylindrical diffuser in front of the fibre optic
101 line light to produce a diffuse light source, a moving table and a computer system.
102 System 2 comprised a SUI Goodrich SU320M-1.7RT InGaAs camera (320 × 240
103 pixels; Sensors Unlimited, Inc., Princeton, NJ, USA), a spectrograph (Specim
104 ImSpector N17E; Spectral Imaging Ltd., Oulu, Finland) covering the spectral range
105 between 900 and 1700 nm (spectral resolution of 7 nm), five halogen lamps (3 x 50W
106 and 2 x 20W) alternately mounted (50W-20W-50W-20W-50W) as source light and a
107 cylindrical diffuser in front of the lamps to produce a diffuse light source, a moving
108 table and a computer system.

109 A two point reflectance calibration was used. A white ceramic tile, the reflectance (R_w)
110 of which was calibrated against a tile of certified reflectance (Ceram Research Ltd,
111 UK), was used as a white reference while dark current was recorded by taking a

112 measurement after covering the spectrograph lens with a cup and closing the shutter.
113 Corrected reflectance values (**R**) were calculated taking into account the relationship
114 between sample (**S**), white standard (**W**) and dark current (**D**) absolute signal intensities
115 and the known reflectance of the ceramic tile (R_w) using the following formula:

$$116 \quad \mathbf{R} = \frac{(\mathbf{S} - \mathbf{D})}{(\mathbf{W} - \mathbf{D})} \cdot \mathbf{R}_w \quad (1)$$

117 Data were recorded in units of reflectance and saved in ENVI header format using the
118 instrument acquisition software (Spectral Scanner; DV Optics, Padua, Italy). When
119 using System 1, only spectral data in the 450 – 900 nm regions were used in data
120 analysis due to reduced efficiency of the light source and CCD in wavelength regions
121 outside this range. In the case of System 2, the spectral range was attenuated to 950 -
122 1650 nm for similar operational reasons.

123 *2.3. Data processing and analysis*

124 A flowchart of the data processing and analysis strategy employed in this study is
125 shown in **Fig. 2**. Data treatment and quantitative model development was carried out
126 using Matlab (R2010b; The Math Works, Inc. USA). For each hyperspectral image,
127 regions of interest (ROIs) of approx. 3 cm diameter were selected using an interactive
128 selection tool available in the acquisition software ('ROI tool') and 500 pixels were
129 randomly-selected within each ROI. Spectral information within each individual pixel
130 was used in subsequent data analysis. Spectral data were pre-treated using the standard
131 normal variate (SNV) transform to diminish the effects of light scatter. Prior to
132 quantitative model development, principal component analysis (PCA) was applied to
133 the data (32000 spectra = 64 samples x 500 pixels) in order to identify and, if necessary,
134 eliminate spectral outliers using the T^2 Hotelling value (Esquerre et al., 2012; Hotelling,
135 1931). Finally, quantitative calibrations were developed by partial least squares (PLS)

136 regression using total glucosinolates as the dependent (Y) variable and pixel spectra as
137 the independent (X) variables.

138 *2.4 Prediction map*

139 Slices of whole freeze-dried broccoli were scanned in the NIR zone (950-1650 nm) to
140 apply the previously constructed model and identify the glucosinolate allocation. Prior
141 to the quantitative analysis, a thresholding rule method was applied to the broccoli
142 images to isolate the broccoli from other parts of image. An image was generated using
143 the maximum reflectance value of each pixel spectrum in a raw image. A threshold of
144 0.45 reflectance units was set analysing the corresponding histogram and drawing a
145 tentative mask image in an iterative process. Pixels with the maximum reflectance
146 values larger than the threshold were classified as broccoli and the remaining pixels
147 were considered as background. SNV was applied to minimise the effects of scattering
148 in the mask created and then the PLS model was applied.

149 **3. Results and discussion**

150 *3.1. Spectra and model development*

151 **Fig. 3** shows the average and standard deviation (10 times amplified) spectra of broccoli
152 powders over the 450-900 nm (**Fig. 3a**) and 950-1650 nm (**Fig.3b**) ranges. Standard
153 deviation spectra have been multiplied by a factor of 10 for display reasons. A strong
154 feature of the sample spectra was the absorbance pattern in the 650-700 nm range. This
155 arose from pigment which displayed a green tinge, likely to be mainly chlorophyll (**Fig.**
156 **3a**). A marked absorbance pattern in the 1400-1500 nm range was clearly present in the
157 spectra (**Fig. 3b**) and this seems likely to arise from first overtone absorbances of O-H
158 functional groups. One likely cause of such absorbances is variations in moisture
159 content of the sample spectra; while all samples were lyophilised, the sensitivity of NIR
160 spectroscopy to water will disclose even small, inter-sample moisture content

161 variations. Absorbance by O-H groups in structural carbohydrates (e.g. cellulose) may,
162 of course, also be implicated.

163 In the case of the first hyperspectral device (450-900 nm), using 6 principal components
164 (98.7% of the variability in the spectral data) and the T^2 Hotelling value, 1382 spectra
165 were identified and eliminated as spectral outliers. The remaining 30618 spectra, which
166 represented about 96% of the spectral population, were retained and used to perform the
167 partial least squares regression. The optimum number of PLS terms used for the
168 calibration (PLS terms=8) was determined by cross-validation using 10 subsets.
169 Validation errors were combined into a single figure, the standard error of cross-
170 validation (SECV), which was calculated to be $1.80 \mu\text{mol g}^{-1}$ DM in this case.

171 When using the second hyperspectral device (950-1650 nm), using 11 principal
172 components (71.7% of the variability in the spectral data) and T^2 Hotelling value, 1863
173 spectra were identified and eliminated as spectral outliers, approximately 30% more
174 than for the first device reported above. The remaining 30137 spectra, which
175 represented about 94% of the spectral population, were selected and used to perform the
176 partial least squares regression. The optimum number of PLS terms used for the
177 calibration (PLS terms=7) was determined by cross-validation using 10 subsets.
178 Validation errors were combined into a single figure, the standard error of cross-
179 validation (SECV), which had a value of $1.75 \mu\text{mol g}^{-1}$ DM in this case.

180 Better results (SECV=1.75, 7 PLS terms) were obtained using the 950-1650 nm
181 wavelength range. This spectral region contains little information about colour
182 variations between samples, something which would have been a significant feature of
183 the shorter wavelength range models. Subsequent analyses were carried out using this
184 longer wavelength range. **Fig. 4** shows the loading (**Fig. 4a**) and beta regression
185 coefficient (**Fig. 4b**) plots of the PLS model developed for total glucosinolate

186 prediction. It is clear that spectral regions around 1100 nm and 1400 nm make the most
187 important contributions to the model loadings. In loading 1, three main features are
188 apparent - a minimum at 1110 nm which is opposed to a broad peak centred around
189 1200 nm and a reported sharp maximum at 1420 nm. Loading 2 consists simply of a
190 sloping baseline (from 1000 nm to a minimum around 1300 nm) and a single broad
191 maximum centred around 1440 nm. Loading 3 has two minima at 1110 and 1410 nm
192 while loading 4 exhibits a maximum at 1115 nm and a minimum around 1170 nm. In
193 the case of loadings 5, 6 and 7, the main features consist of minima around 1110 and
194 1470 nm (loading 5), maxima around 1080, 1155 and 1460nm (loading 6) and small
195 maxima at 1110 and 1200 nm (loading 7) respectively. Absorbance around 1110 nm
196 may reflect 2nd overtone –OH stretching although a combination –S=O stretch has been
197 previously reported around 1020-1060 nm (Stuart, 2004) which may be relevant given
198 the occurrence of sulphur-containing volatile compounds in broccoli (Jacobsson,
199 Nielsen and Sjöholm, 2004). Absorbances at 1200 nm have been attributed to 2nd
200 overtone of C-H stretch vibrations in -CH₃, -CH₂ or -CH structures (Maalouly and
201 Jaillais, 2006). Definitive attribution of absorbances around 1420-1440 nm is difficult
202 given that absorptions around this wavelength range originate from both –OH and –CH
203 structures. There is an undoubted difference between loading 1 and 2 in this regard,
204 since the relevant peak in loading 1 is sharp, very intense and has a maximum of 1420
205 nm while that in loading 2 is broad, less intense and is centred around 1440 nm. Given
206 the narrow range of moisture variation between samples since they were all lyophilised,
207 it seems reasonable to attribute the 1440 nm peak to water absorbance (1st harmonic -
208 OH stretch) and the 1420 nm feature to absorbances arising from –CH moieties. Given
209 the widespread occurrence of such bonds in biological structures, it is not possible to be

210 more specific about the molecular species involved; however, structural polysaccharides
211 comprising the bulk of broccoli composition are likely to be implicated.

212 It is hazardous to try to interpret regression coefficients at any significant level of detail
213 (Brown and Green, 2009) but a number of features of the regression vector shown in
214 Figure 4b merit mention. Wavelengths with the largest regression coefficients are found
215 at 1130, 1366, 1460, 1480, 1520 and 1540 nm; the latter four are considerably larger
216 than the first two. Coefficients associated with the 1460 and 1480 nm wavelengths are
217 in opposition to the others. The features at 1130 and around 1460-1480 nm are
218 somewhat consistent with the features observed in the individual loadings described
219 above but the emergence of large regression coefficients at 1520 and 1540 nm and a
220 smaller negative coefficient at 1366 nm have no parallel in the loadings vectors.
221 Broadly speaking, however, the regression vector indicates the importance of spectral
222 regions involving -CH vibrations and others which are non-identifiable in the prediction
223 of glucosinolate content in broccoli.

224 3.2. *Prediction map*

225 **Fig. 5** shows a flowchart of the process for performing the image segmentation
226 procedure and the glucosinolate prediction map in broccoli. The main feature of note
227 in this figure is that these compounds were almost exclusively located in the external
228 part of the broccoli florets. Predicted glucosinolate contents in the main stem and even
229 in peripheral structural features of the floret were very low ($<8 \mu\text{mol g}^{-1} \text{DM}$); given the
230 predictive accuracy of the model used for this measurement, it was not possible to
231 indicate whether these molecular species were absent or not from such structures.
232 Although their role in plants is unclear, their potent odour and pronounced taste
233 suggests a role in herbivore and microbial defence. Deposition in external plant parts,

234 confirmed spectroscopically in this work, would be the optimal location for these
235 purposes.

236 **4. Conclusion**

237 Two different spectral regions (vis-NIR and NIR) were studied to develop the
238 quantitative models. Better results were obtained using the 950-1650 nm wavelength
239 range and subsequent analyses were therefore carried out using this spectral zone. The
240 procedure demonstrates potential for the quantitative screening and location of total
241 glucosinolates in broccoli using the 950-1650 nm wavelength range. Nevertheless, a
242 comprehensive study should be made in order to evaluate all other relevant sources of
243 variability in the complete development of these models. It is well known that in the
244 case of agricultural products the range of the variability should be as large as that
245 expected in any future samples. Likely causes of variation in glucosinolates in broccoli
246 include effects of the environment, changes in farm soil, growing season, cultivar,
247 harvest conditions, disease incidence etc. Consequently, all these factors should be
248 taken into account in the development of future robust global models. Such a study
249 would entail several years work but the results reported herein suggest the viability of
250 obtaining useful results from such an undertaking. One noteworthy observation arising
251 from this preliminary study is that these glucosinolate compounds seem to be mainly
252 located in the external part of the vegetable florets.

253 **Acknowledgements**

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

259 Spanish MICINN for the Juan de la Cierva contract (JCI-2011-09201) and Universidad
260 de Sevilla for the mobility grant (Universidad de Sevilla Research Plan). Spanish
261 MICINN project AGL2011-30254-C02 and Junta de Andalucía PGC project AGR
262 6331.

263 **References**

- 264 Barbin, D.F., ElMasry, G., Sun, D.-W., Allen, P., (2012). Predicting quality and sensory
265 attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta*
266 719, 30-42.
- 267 Brown, C. D. and Green, R.L. (2009). Critical factors limiting the interpretation of
268 regression factors in multivariate calibration. *Trends in Analytical Chemistry* 28(4),
269 506-514.
- 270 Burger, J., Gowen, A.A., (2011). Data handling in hyperspectral image analysis.
271 *Chemometrics and Intelligent Laboratory Systems* 108(1), 13-22.
- 272 Crozier, A., Clifford, M.N., Ashihara, H., (2006). *Plant Secondary Metabolites.*
273 *Occurrence, Structure and Role in the Human Diet.* Blackwell Publishing, Garsington
274 Road, Oxford, UK.
- 275 ElMasry, G., Wang, N., ElSayed, A., Ngadi, M., (2007). Hyperspectral imaging for
276 nondestructive determination of some quality attributes for strawberry. *Journal of Food*
277 *Engineering* 81(1), 98-107.
- 278 Esquerre, C., Gowen, A.A., Downey, G., O' Donnell, C.P., (2012). Wavelength
279 selection for development of a near infrared imaging system for early detection of
280 bruise damage in mushrooms (*Agaricus bisporus*). *Journal of Near Infrared*
281 *Spectroscopy* 20, 537-546.
- 282 Fernandes, A.M., Oliveira, P., Moura, J.P., Oliveira, A.A., Falco, V., Correia, M.J.,
283 Melo-Pinto, P., (2011). Determination of anthocyanin concentration in whole grape
284 skins using hyperspectral imaging and adaptive boosting neural networks. *Journal of*
285 *Food Engineering* 105(2), 216-226.
- 286 Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, J.M., (2007).
287 Hyperspectral imaging -an emerging process analytical tool for food quality and safety
288 control. *Trends in Food Science and Technology* 18(12), 590-598.
- 289 Hernandez-Hierro, J.M., Valverde, J., Villacreces, S., Reilly, K., Gaffney, M.,
290 Gonzalez-Miret, M.L., Heredia, F.J., Downey, G., (2012). Feasibility Study on the Use
291 of Visible-Near-Infrared Spectroscopy for the Screening of Individual and Total
292 Glucosinolate Contents in Broccoli. *Journal of Agricultural and Food Chemistry* 60(30),
293 7352-7358.
- 294 Hotelling, H., (1931). The Generalization of Student's Ratio. *Annals of Mathematical*
295 *Statistics* 2(3), 360-378.
- 296 Jacobsson, J., Nielsen, T., Sjöholm, I. (2004). Influence of Temperature, Modified
297 Atmosphere Packaging, and Heat Treatment on Aroma Compounds in Broccoli. *Journal*
298 *of Agricultural and Food Chemistry* 52, 1607-1614.
- 299 Jeffery, E., Araya, M., (2009). Physiological effects of broccoli consumption.
300 *Phytochemistry Reviews* 8(1), 283-298.

- 301 Jeffery, E.H., Brown, A.F., Kurilich, A.C., Keck, A.S., Matusheski, N., Klein, B.P.,
302 Juvik, J.A., (2003). Variation in content of bioactive components in broccoli. *Journal of*
303 *Food Composition and Analysis* 16(3), 323-330.
- 304 Ji-Yong, S., Xiao-Bo, Z., Jie-Wen, Z., Kai-Liang, W., Zheng-Wei, C., Xiao-Wei, H.,
305 De-Tao, Z., Holmes, M., (2012). Nondestructive diagnostics of nitrogen deficiency by
306 cucumber leaf chlorophyll distribution map based on near infrared hyperspectral
307 imaging. *Scientia Horticulturae* 138, 190-197.
- 308 Kamruzzaman, M., ElMasry, G., Sun, D.-W., Allen, P., (2012). Prediction of some
309 quality attributes of lamb meat using near-infrared hyperspectral imaging and
310 multivariate analysis. *Analytica Chimica Acta* 714, 57-67.
- 311 Lorente, D., Aleixos, N., Gómez-Sanchis, J., Cubero, S., García-Navarrete, O., Blasco,
312 J., (2012). Recent Advances and Applications of Hyperspectral Imaging for Fruit and
313 Vegetable Quality Assessment. *Food and Bioprocess Technology* 5(4), 1121-1142.
- 314 Maalooly, J., Jaillais, B. (2006). Glucides. In: *La spectroscopie infrarouge et ses*
315 *applications analytiques*. (pp 175-232). Lavoisier Paris, France.
- 316 McGoverin, C.M., Weeranantanaphan, J., Downey, G., Manley, M., (2010). The
317 application of near infrared spectroscopy to the measurement of bioactive compounds in
318 food commodities. *Journal of near Infrared Spectroscopy* 18(2), 87-111.
- 319 Mendoza, F., Lu, R., Ariana, D., Cen, H., Bailey, B., (2011). Integrated spectral and
320 image analysis of hyperspectral scattering data for prediction of apple fruit firmness and
321 soluble solids content. *Postharvest Biology and Technology* 62(2), 149-160.
- 322 Rajkumar, P., Wang, N., ElMasry, G., Raghavan, G.S.V., Garipey, Y., (2012). Studies
323 on banana fruit quality and maturity stages using hyperspectral imaging. *Journal of*
324 *Food Engineering* 108(1), 194-200.
- 325 Shahindi, F., (1990). *Canola and rapeseed production, chemistry and processing*
326 *technology*. Van Nostrand Reinhold, New York, USA.
- 327 Sun, D.W., (2010). *Hyperspectral Imaging for Food Quality Analysis and Control*.
328 Elsevier Science & Technology, San Diego, CA, USA.
- 329 Stuart, B.H. (2004). *Infrared Spectroscopy: Fundamentals and Applications*. John
330 Wiley & Sons, New York, USA (2004).
- 331 Vallejo, F., Tomás-Barberán, F.A., Benavente-García, A.G., García-Viguera, C.,
332 (2003). Total and individual glucosinolate contents in inflorescences of eight broccoli
333 cultivars grown under various climatic and fertilization conditions. *Journal of the*
334 *Science of Food and Agriculture* 83, 307–313.
- 335 Verkerk, R., Schreiner, M., Krumbein, A., Ciska, E., Holst, B., Rowland, I., De
336 Schrijver, R., Hansen, M., Gerhäuser, C., Mithen, R., Dekker, M., (2009).
337 Glucosinolates in *Brassica* vegetables: The influence of the food supply chain on intake,
338 bioavailability and human health. *Molecular Nutrition Food Research* 53(S2), S219-
339 S219.

- 340 Wang, J., Gu, H., Yu, H., Zhao, Z., Sheng, X., Zhang, X., (2012a). Genotypic variation
341 of glucosinolates in broccoli (*Brassica oleracea* var. *italica*) florets from China. Food
342 Chemistry 133(3), 735-741.
- 343 Wang, S., Huang, M., Zhu, Q., (2012b). Model fusion for prediction of apple firmness
344 using hyperspectral scattering image. Computers and Electronics in Agriculture 80, 1-7.
- 345 Wu, D., Shi, H., Wang, S., He, Y., Bao, Y., Liu, K., (2012). Rapid prediction of
346 moisture content of dehydrated prawns using online hyperspectral imaging system.
347 Analytica Chimica Acta 726, 57-66.
348
349

350 **Figure captions**

351 **Fig. 1.** Imaging system line drawing: illumination source, diffuser, moving base, optics
352 (mirror and lens), spectrograph, camera and computer.

353 **Fig. 2.** Flow chart of the data processing and analysis strategy employed in this study.

354 **Fig.3.** Average and standard deviation (10 times amplified) spectra of broccoli powders
355 over the 450-900 nm (a) and 950-1650 nm (b) wavelength ranges.

356 **Fig 4.** PLS loadings (a) and beta regression coefficients (b) of the PLS model for total
357 glucosinolate prediction. For separate plots of each PLS loading, see
358 Supplementary data (**Supplementary Fig. 1.**).

359 **Fig 5.** Flowchart of the process for performing the image segmentation procedure and
360 glucosinolate prediction map in broccoli.

361

Fig. 1

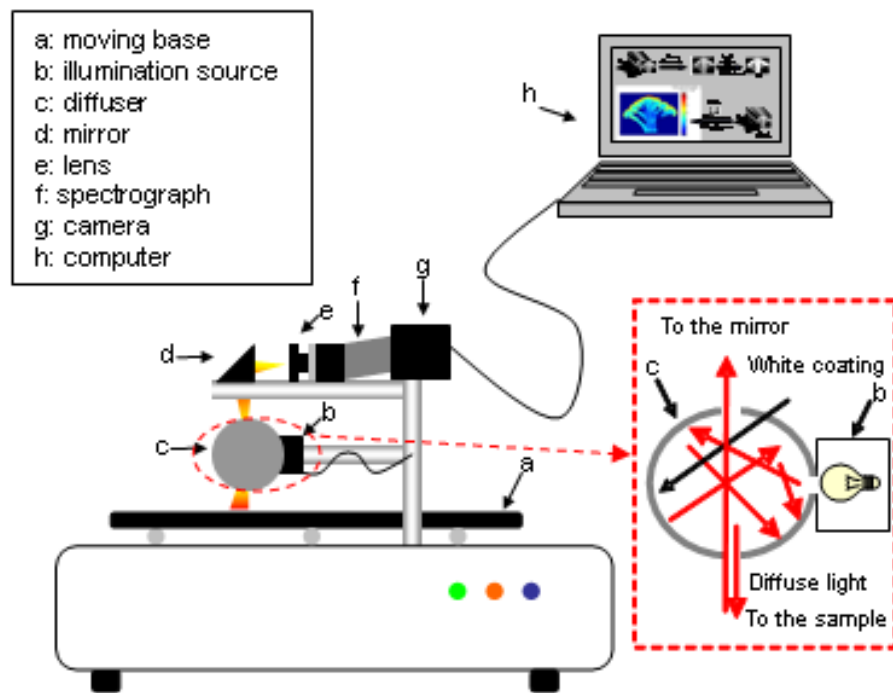


Fig. 2

Modelling strategy

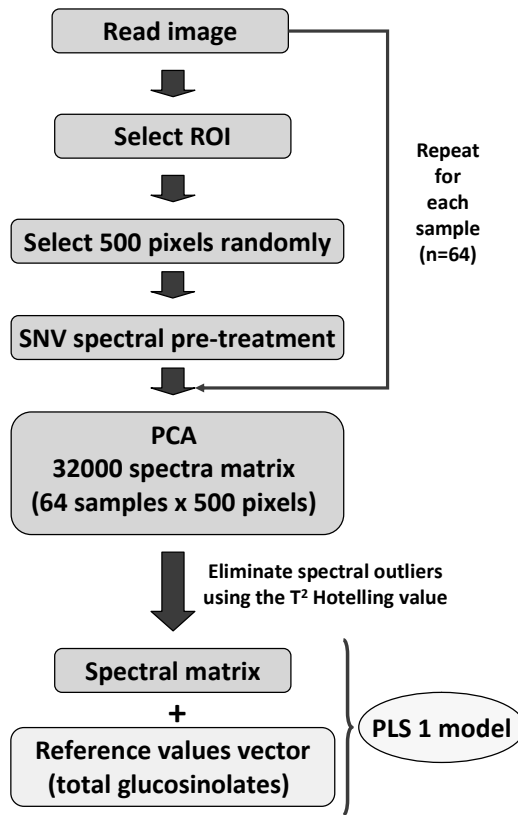


Fig. 3

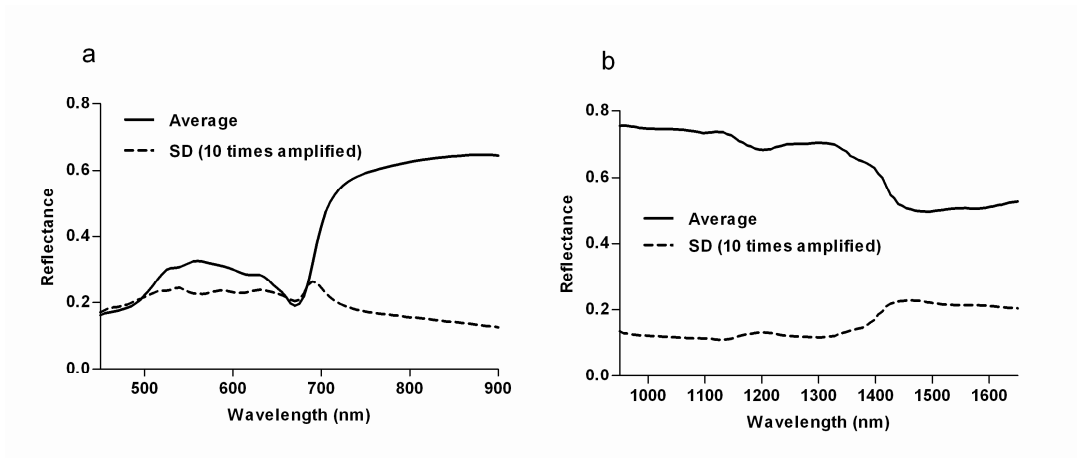


Fig. 4

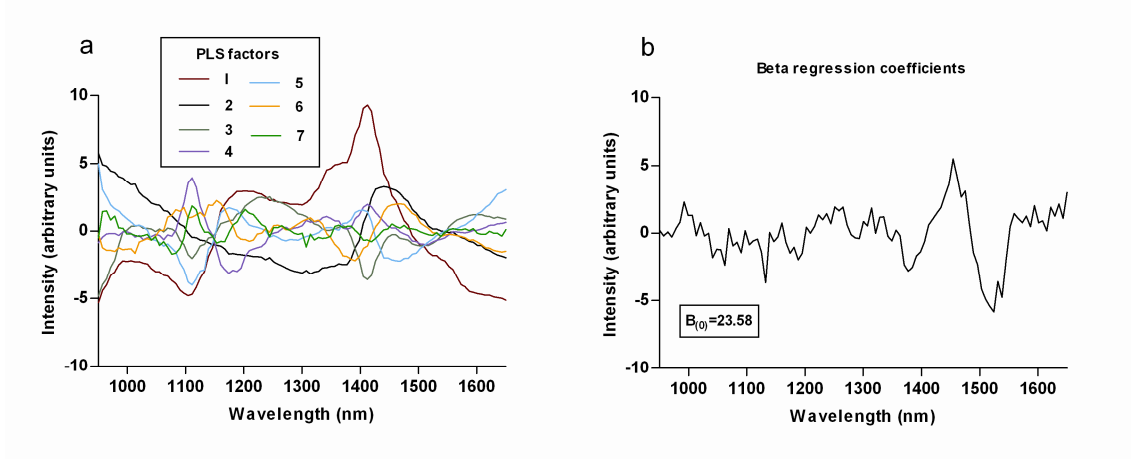


Fig. 5

