Journal of Food and Drug Analysis

Volume 21 | Issue 3 Article 3

Quantification of bioactive gentiopicroside in the medicinal plant Gentiana scabra Bunge using Near infrared spectroscopy

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.jfda-online.com

Original Article

Quantification of bioactive gentiopicroside in the medicinal plant *Gentiana scabra* Bunge using near infrared spectroscopy

Yung-Kun Chuang [a,b], *Suming Chen* [a,c,*], *Yangming Martin Lo* [b], *I-Chang Yang* [b], *Yu-Fan Cheng* [a], *Ching-Yin Wang* [a], *Chao-Yin Tsai* [c], *Ruey-Min Hsieh* [d], *Kuo-Hsi Wang* [d], *Chuo-Chun Lai* [d], *Wen-Chung Chen* [d]

[a] *Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan, ROC*
[b] *Department of Nutrition and Food Science, University of Maryland, College Park, MD, USA*
[c] *Bioenergy Research Center, National Taiwan University, Taipei, Taiwan, ROC*
[d] *Taiwan Sugar Research Institute, Taiwan Sugar Corporation, Tainan, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 16 December 2012
Received in revised form
21 April 2013
Accepted 31 May 2013
Available online 7 August 2013

Keywords:
Calibration model
Gentiana scabra Bunge
Gentiopicroside
Near infrared spectroscopy

ABSTRACT

Near infrared (NIR) spectroscopy was employed to perform a quantitative analysis of gentiopicroside, the bioactive component of the medicinal plant *Gentiana scabra* Bunge. Modified partial least squares regression (MPLSR) and stepwise multiple linear regression (SMLR) calibration models were built using 94 plant tissue culture samples and 136 grown plant samples, respectively, over the full wavelength range (400–2498 nm) and the silicon charge-coupled-device (CCD) sensing band (400–1098 nm). For tissue culture, the smoothing, first-derivative MPLSR model can produce the best effect [calibration set (R_c) = 0.868, standard error of calibration (SEC) = 0.606%, standard error of validation (SEV) = 0.862%] in the wavelength ranges of 900–1000, 1200–1300, and 1600–1700 nm. By contrast, for grown plant samples, the smoothing, second-derivative MPLSR model can produce the best effect (R_c = 0.944, SEC = 0.502%, SEV = 0.685%) in the wavelength ranges of 400–500, 1100–1200, 1600–1800, and 2200–2300 nm. With the silicon CCD sensing band, the smoothing, second-derivative, four-wavelength (670, 786, 474, and 826 nm) SMLR model showed best predictability (R_c = 0.860, SEC = 0.775%, SEV = 0.848%). This study successfully built spectral calibration models for determining gentiopicroside content at different growth stages of *G. scabra* Bunge. The specific wavelengths selected within the silicon CCD sensing band can be used in combination with multispectral imaging as a powerful tool for monitoring or inspecting the quality of *G. scabra* Bunge during cultivation.

* *Corresponding author*. Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan.
E-mail address: schen@ntu.edu.tw (S. Chen).

1021-9498
http://dx.doi.org/10.1016/j.jfda.2013.07.011

1. Introduction

Dried root and rootstock of *Gentiana scabra* Bunge are commonly used as pharmaceutical raw materials because they are rich in many secoiridoid glycosides, such as gentiopicroside, swertiamarin, and sweroside [1]. In addition to its antimicrobial and anti-inflammatory effects, gentiopicroside, in particular, has been shown to protect liver, inhibit liver dysfunction, and promote gastric acid secretion, which make it a popular ingredient in Chinese herbal medicine and health products [2].

In the early days, *G. scabra* Bunge was mainly collected in the wild. With the increase in demand for *G. scabra* Bunge and depletion of the wild resources, restoration of *G. scabra* Bunge has become an important issue [3]. Studies in recent years used tissue culture technology to cultivate *G. scabra* Bunge artificially [4], by domesticating the tissue culture samples of *G. scabra* Bunge and then transplanting them to the greenhouse for cultivation. In order to monitor the changes in *G. scabra* Bunge during its growth process, it is necessary to measure its bioactive components. However, the commonly used methods, such as micellar electrokinetic capillary chromatography (MECC) [5], high-performance liquid chromatography (HPLC) [6–11], liquid chromatography–mass spectrometry [12,13], and ultraperformance liquid chromatography (UPLC) [14], are all time consuming and energy intense, and hence not applicable for daily quality inspection of *G. scabra* Bunge during cultivation.

Near infrared (NIR) spectroscopy is a nondestructive inspection method that can measure the target object rapidly. An NIR spectrum provides characteristic information on the internal constituents of the sample, so it has widely been used in dispensation, such as component analysis of Chinese herbal plants Angelicae gigantis Radix [15], *Rhubarb* [16], licorice [17], *Panax* species [18], and *Lonicera japonica* [19], as well as content detection of active pharmaceutical ingredients (APIs) in tablets [20–22]. However, it has not been employed to monitor the growth of *G. scabra* Bunge qualitatively.

The present study aimed at exploring the NIR features of gentiopicroside, the bioactive component of *G. scabra* Bunge, in order to build spectral calibration models. Moreover, the applicability of the silicon charge-coupled-device (CCD) sensing band when using multispectral imaging technology to inspect the quality of *G. scabra* Bunge was evaluated.

2. Methods

2.1. *G. scabra Bunge sample preparation*

Samples of *G. scabra* Bunge were provided by the Taiwan Sugar Research Institute (TSRI) (Tainan, Taiwan). A total of 94 tissue culture samples and 68 grown plant samples of different cultivation times were acquired. Shoots and roots of the grown plant samples were measured separately in order to compare their differences. The *G. scabra* Bunge samples were first dried for 48 hours in a dryer (50 °C) and then milled with a high-speed grinder (RT-02A; Sun-Great Technology Co., Ltd., New Taipei City, Taiwan). The dried powder was filtered with a 100-mesh sieve and stored in amber sample vials to avoid exposure to light.

2.2. *NIR spectra and HPLC measurement*

Dry powder of *G. scabra* Bunge was poured gently into a small ring cup [internal diameter (i.d.) 5 cm] and subjected to NIR measurements (NIRS 6500, FOSS NIRSystems, Inc., Laurel, MD, USA). The reflectance spectra of the samples were collected in the range of 400–2498 nm with 2 nm intervals, and the NIR spectrum of each sample was the average of 32 scans.

To attain the reference value of the bioactive component, gentiopicroside was measured by HPLC (DX 500 ion chromatograph; Dionex Corporation, Sunnyvale, CA, USA) equipped with a DIONEX C18 column (250 mm × 4.6 mm i.d.). The peak of gentiopicroside appeared at 250 nm when methanol–water (mixed at a ratio of 20:80) was used as the mobile phase at a flow rate of 1 mL/min. A high precision balance scale was used to measure the gentiopicroside standard powder, and diluted into 1000, 500, and 250 ppm with 70% methanol as the standard solutions for the three-point calibration of HPLC. A quantitative linear relationship was established between the standard concentration and the peak area.

2.3. *Data analysis*

In order to apply the specific wavelengths identified to multispectral imaging inspection of *G. scabra* Bunge, the spectra of the full wavelength range (400–2498 nm) and the silicon CCD sensing band (400–1098 nm) were analyzed. Modified partial least squares regression (MPLSR) and stepwise multiple linear regression (SMLR) methods were employed to build calibration models for determining gentiopicroside content.

2.4. *MPLSR*

MPLSR, an extension of partial least squares regression (PLSR), abides by the principle of normalization of the spectra and constituent values prior to PLSR, which is a standard tool in chemometrics and has widely been used in the pharmaceutical, chemical, and agricultural fields [23]. For a spectral analysis using PLSR, the spectra can be considered to be composed of several principal components (PCs) and expressed as a "factor" in the PLSR algorithm. The factors' sequence is determined by their influences, i.e., a more important factor is ranked earlier in the order. Because PLSR analysis uses information from spectral bands, analysis results can be improved by selecting appropriate number of factors and specific wavelength ranges.

2.5. *SMLR*

The SMLR analysis method selects specific wavelengths according to the F test ($F \geq 3$) of null hypothesis testing [24]. In order to build calibration models over numerous wavelengths, the SMLR algorithm chooses the most important specific wavelength from the major molecular bonding region of the objects, and the second most important specific wavelength is

usually chosen from the region located in the combination of related molecular bonding or the overtone of complementary bonding, and by analogy. When adding a new wavelength for training, the algorithm will base on the previously selected specific wavelengths to continue finding the wavelength, which can allow the highest multiple coefficient of determination (r2) and the minimum prediction error, and determine whether such a wavelength can replace the current specific wavelength or not. In case of poor competency of the newly added wavelength for training, the algorithm will stop training.

2.6. Spectral pretreatments

The purpose of spectral pretreatments was to eliminate the spectral variation that is not caused by chemical information contained in the samples [25,26]. Addition of light scattering effects into the spectra is unavoidable when using NIR spectroscopy to measure powder samples, especially when the particle size is not uniform, multiplicative scatter correction (MSC) was included to allow additive and multiplicative transformation of the spectra [Equation (1)]. It was conducted using the average spectrum of all samples as the reference value, and calculating parameters a and b with the least square method. After MSC treatment, not only the physical impact of nonuniform particles on the spectra of *G. scabra* Bunge powder was reduced [27,28], but also the linearity of the spectra was confirmed [29], which contributed to the subsequent linear regression analysis [30].

$$x_{ik}(\text{new}) = \frac{[x_{ik}(\text{old}) - a_i]}{b_i} \quad (1)$$

The spectra of *G. scabra* Bunge powder post MSC was subjected to three independent treatments, namely, (1) smoothing, (2) smoothing with first derivative, and (3) smoothing with second derivative, in order to choose the best pretreatment parameters, including the smoothing points and the gap ranging from 2 to 50, with the gap being greater than or equal to the smoothing points.

2.7. Model establishment

Spectral calibration models of MPLSR and SMLR were built using WinISI II chemometric software (Infrasoft International, LLC, Port the Matilda, PA, USA). The MPLSR analysis procedure included the following: (1) spectral pretreatments; (2) selecting specific wavelength regions; (3) selecting calibration and validation sets; and (4) determining the best calibration model. In steps 1 and 2, three-fold cross-validation was used to enable objective selection of the parameters. Samples were divided into calibration and validation sets at a ratio of 2:1, according to the gentiopicroside concentration in the sample. All samples were ranked according to their increasing gentiopicroside concentration, with this concentration being higher in the calibration set than in the validation set; yet distribution of gentiopicroside concentration was similar in both sets. When selecting the best calibration model, in order to avoid overfitting caused by the use of excessive factors, the following principles were adhered to: (1) the maximum number of factors should be one-tenth of the number of calibration sets plus two or three; (2) if addition of a new factor makes the standard error of validation (SEV) rise, addition of the new factor should be stopped; and (3) when the SEV is lower than the standard error of calibration (SEC), new factor should not be added. The SMLR analysis procedure had the following steps: (1) selecting calibration and validation sets; (2) spectral pretreatments; and (3) determining best calibration model and specific wavelengths. The same calibration and validation sets were used for both MPLSR and SMLR analyses.

After the respective spectral calibration models of MPLSR and SMLR were built, these models were then used to predict the gentiopicroside concentration of the calibration and validation sets. Predictability of the models was evaluated based on the statistical parameters, including coefficient of correlation of calibration set (R_c), SEC, SEV, bias, and the ratio of the standard error of performance to the standard deviation of the reference values (RPD), as defined in the following:

$$\text{SEC} = \left[\frac{1}{n_c}\sum_{i=1}^{n_c}(Yr - Yc)_i^2\right]^{1/2} \quad (2)$$

$$\text{SEV} = \left[\frac{1}{n_v}\sum_{i=1}^{n_v}[(Yr - Yv) - \text{Bias}]_i^2\right]^{1/2} \quad (3)$$

$$\text{Bias} = \frac{1}{n_v}\sum_{i=1}^{n_v}(Yr - Yv)_i \quad (4)$$

$$\text{RPD} = \text{SD}/\text{SEV} \quad (5)$$

where Yc and Yv represent the estimated gentiopicroside concentrations of the calibration set and the validation sets, respectively; Yr is the reference gentiopicroside concentration; n_c and n_v are the number of samples in the calibration set and validation set, respectively; and SD is the standard deviation of gentiopicroside concentration within the validation set.

3. Results and discussion

3.1. Gentiopicroside concentration and distribution in *G. scabra Bunge*

Gentiopicroside contents of samples consisting of different parts of *G. scabra* Bunge (94 tissue culture, 68 shoot, and 68 root

Table 1 – Gentiopicroside contents in tissue culture and grown plants of *Gentiana scabra* Bunge.

Sample	No.	Gentiopicroside content (%)				
		Min.	Max.	Mean	SD	CV
Tissue culture	94	2.69	8.18	5.35	1.29	0.24
Grown plants						
Shoot	68	1.34	5.90	3.26	0.91	0.28
Root	68	2.24	8.77	4.68	1.62	0.35

CV = coefficient of variation; SD = standard deviation.

samples) are shown in Table 1. Gentiopicroside content of grown plants of *G. scabra* Bunge (including shoots and roots) was found to increase after *G. scabra* Bunge tissue culture samples were transplanted into a greenhouse for cultivation. Within a grown plant, the gentiopicroside content was significantly higher in root than in shoot, indicating that during greenhouse cultivation, gentiopicroside was stored mainly in the root.

3.2. Correlation between NIR spectra and gentiopicroside content

NIR spectra of the 94 *G. scabra* Bunge tissue culture samples and 136 grown plant samples (68 shoots and 68 roots) were acquired after employing the MSC treatment. As shown in Fig. 1A and B, absorption peaks were observed in the visible region of both blue light (452 nm) and red light (666 nm) because, during photosynthesis, chlorophyll in *G. scabra* Bunge would absorb blue and red lights the most. Spectra of tissue culture samples and shoots were similar, which could be attributed to the fact that, during the domestication period, tissue is mainly composed of shoots of *G. scabra* Bunge, while root development is not obvious at that time. The root spectra in the visible region showed a significant difference, with high absorption from green to yellow light (from 492 nm to 586 nm) and low absorption (flat waveform) from orange to red light (from 606 nm to 700 nm). This could be due to lack of chlorophyll in the roots of *G. scabra* Bunge plant, which reduces absorption of blue and red light, and reflects green light.

After MSC treatment, the spectra of *G. scabra* Bunge tissue culture and grown plant samples were analyzed using the following pretreatments: (1) smoothing; (2) smoothing with first derivative; and (3) smoothing with second derivative. The best pretreatment parameters of the tissue culture spectra (smoothing points/gap) were (1/0), (6/6), and (8/8), whereas those of the grown plant spectra were (1/0), (2/2), and (3/3); both the smoothing points and the gap were less than 10, indicating that the NIRS 6500 spectrophotometer was stable and the spectra of *G. scabra* Bunge powder exhibited minimal noise.

The correlation between the spectra and gentiopicroside content of *G. scabra* Bunge powder was analyzed prior to selecting the specific wavelength regions. Distributions of gentiopicroside correlation coefficients of *G. scabra* Bunge tissue culture and grown plant samples were compared using the original, first derivative, and second derivative spectra, and the threshold value ($|r| > 0.55$) was set to determine the degree of correlation. It is unnecessary to avoid the absorption bands of O–H bond around 1450 and 1900 nm because the influence of water absorption on the spectra of *G. scabra* Bunge powder has already been eliminated. Fig. 2A shows that the bands of high correlation between the spectra and gentiopicroside content of tissue culture were mainly distributed in the NIR region, with only a few in the visible region. Absorption bands of the original spectra were located in the first overtone of the C–H and C–C bonds, whereas those of the first derivative spectra were located in the orange light and the combination of the first overtone of C–H bond. Moreover, absorption bands of the second derivative spectra were found to be located in the second overtone of C=O bond stretch.

Correlation coefficient distributions between absorbance values of the spectra and gentiopicroside content of the *G. scabra* Bunge grown plants were also compared using the original, first derivative, and second derivative spectra (Fig. 2B). Highly correlated bands were found in both visible and NIR regions. Absorption bands in the original spectra were located between the yellow and orange light, as well as in the combination of two C–H bonds. Absorption bands in the first derivative spectra were located between the orange and red light, and in the fourth overtone of C–H bond, the third overtone of C–H bond, the first overtone of C–H bond, and the combination of two C–H bonds, whereas absorption bands of the second derivative spectra were located in the blue and red light, the third overtone of N–H bond, and the combination of two C–H bonds. Because the spectra of shoots and roots showed obvious differences in the visible region, the correlation of blue and red light to gentiopicroside content was improved, indicating that the amount of chlorophyll contained in different parts of a grown plant also affects the performance of the specific wavelength regions. The specific wavelengths of both tissue culture and grown plant samples in the NIR region were located in the combination of two C–H bonds and the overtones of C–H bond, indicating that C–H bonds are the main absorbers of NIR light. Fig. 2 shows that the wavelength ranges of 900–1300 nm, 1500–1800 nm, and 2200–2300 nm were the major absorption bands (according to the absorption bands of C–H bonds in the spectrum), and

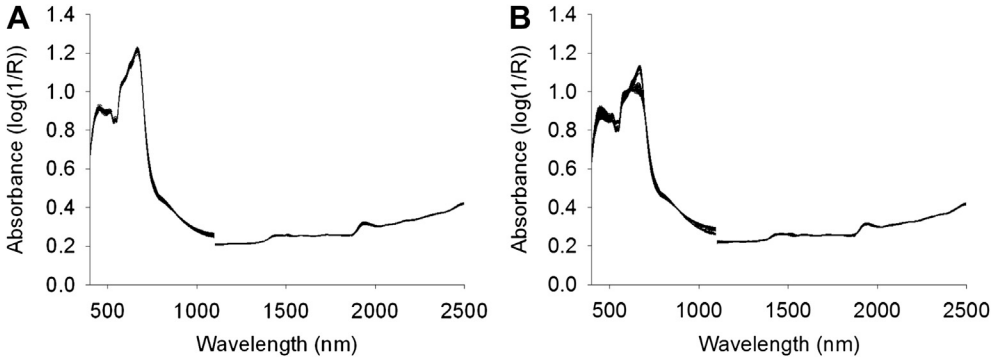

Fig. 1 – Spectra of *Gentiana scabra* Bunge powder post-multiplicative scatter correction (MSC): (A) tissue culture and (B) grown plants.

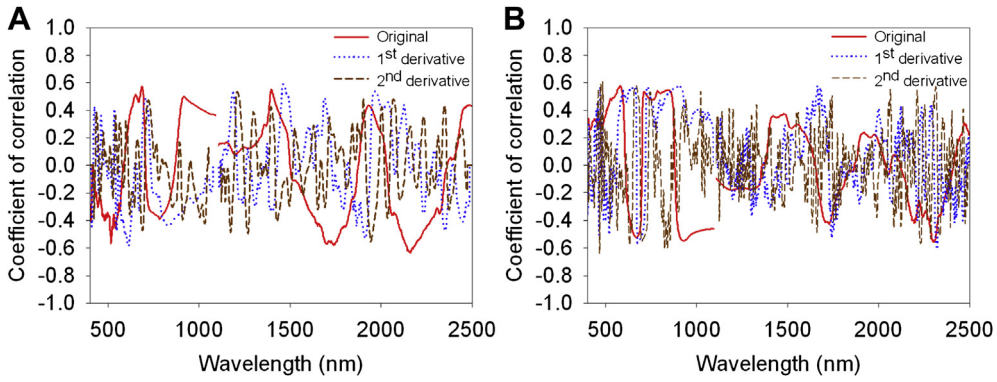

Fig. 2 – Correlation coefficient distributions between absorbance values of the spectra and gentiopicroside contents of the *Gentiana scabra* Bunge powder: (A) tissue culture and (B) grown plants.

these wavelengths can be used to provide a basis for selecting the appropriate specific wavelength regions when conducting MPLSR analysis. As for the spectral band 400–650 nm, which belonged to the absorption band of blue to red light, color information was also reflected in the spectra.

3.3. *Gentiopicroside quantification using specific wavelength ranges*

Out of the valid *G. scabra* Bunge samples, 89 and 126 tissue culture and grown plant samples, respectively, were retained for statistical calibration and validation of the gentiopicroside content (Table 2). No significant difference was found in the mean, standard deviation, and coefficient of variation (CV) of the effective samples, calibration set, and validation set, indicating that the distributions of gentiopicroside contents of the two sample groups were consistent.

MPLSR analysis results of full wavelength range spectra (400–2498 nm) are shown in Table 3. The best calibration model of *G. scabra* Bunge tissue culture was found with the first derivative spectra and six factors, with both smoothing points and gap set at 6, using the wavelength ranges of 900–1000 nm, 1200–1300 nm, and 1600–1700 nm, and resulting in R_c = 0.868, SEC = 0.606%, SEV = 0.862%, bias = −0.215%, and RPD = 1.32. Due to the spectral difference between the calibration and validation sets, prediction result of the validation set was a little worse than that of the calibration set when using the calibration model. The best calibration model of a *G. scabra* Bunge grown plant was identified using the second derivative spectra and five factors, with both smoothing points and gap set at 3, using the wavelength ranges of 400–500 nm, 1100–1200 nm, 1600–1800 nm, and 2200–2300 nm. The results were R_c = 0.944, SEC = 0.502%, SEV = 0.685%, bias = −0.162%, and RPD = 2.19. The calibration models built based on the first and second derivative spectra were both better than those based on the original spectra, confirming that heterogeneous particles of *G. scabra* Bunge powder affect the spectral absorption. Calibration models of grown plants were all better than those of the tissue culture, even with fewer spectral pretreatments, because more grown plant samples can build more stable calibration models. The specific wavelength regions of tissue culture and grown plants were distributed mainly in 900–1300 nm and 1600–1800 nm, and the calibration models of grown plants also incorporated the spectral information within 400–500 nm and 2200–2300 nm, indicating that the NIR region contained more information about gentiopicroside. The differences in absorption of shoots and roots in the visible region also qualified 400–500 nm to be employed as a specific wavelength region.

Table 2 – Gentiopicroside contents of effective samples, calibration set, and validation set in tissue culture and grown plants of *Gentiana scabra*.

Sample	No.	Gentiopicroside content (%)				
		Min.	Max.	Mean	SD	CV
Tissue culture						
Effective samples	89	2.69	7.83	5.26	1.19	0.23
Calibration set	60	2.69	7.83	5.26	1.22	0.23
Validation set	29	3.12	7.35	5.26	1.14	0.22
Grown plants						
Effective samples	126	1.34	8.77	4.01	1.51	0.38
Calibration set	84	1.34	8.77	4.01	1.52	0.38
Validation set	42	1.59	8.19	4.01	1.50	0.37

CV = coefficient of variation; SD = standard deviation.

3.4. *Gentiopicroside quantification using CCD camera wavelength spectra*

An MPLSR analysis of the silicon CCD sensing band (400–1098 nm) is shown in Table 4. The best calibration model of *G. scabra* Bunge tissue culture was acquired when the second derivative spectra and three factors were employed, where both smoothing points and gap were at 2, with wavelength ranges of 400–500 nm and 800–1000 nm; the results were R_c = 0.865, SEC = 0.611%, SEV = 0.772%, bias = 0.025%, and RPD = 1.47. The best calibration model of a *G. scabra* Bunge grown plant was found with the first derivative spectra and five factors, with smoothing points and gap being at 2, and wavelengths in the range of 400–600 nm and 900–1098 nm; the results were as follows: R_c = 0.904, SEC = 0.649%, SEV = 0.724%, bias = −0.089%, and RPD = 2.08. Regardless of

Table 3 – Prediction of the gentiopicroside content in tissue culture and grown plants of *Gentiana scabra* Bunge by MPLSR models in the wavelength range of 400–2498 nm.

Sample	Spectrum	Wavelength range (nm)*	Smoothing points/gap	Factors	Calibration set		Validation set		
					R_c	SEC (%)	SEV (%)	Bias (%)	RPD
Tissue culture	Original	900–1000 1600–1700	1/0	5	0.752	0.804	0.943	−0.137	1.21
	First derivative	900–1000 1200–1300 1600–1700	6/6	6	0.868	0.606	0.862	−0.215	1.32
	Second derivative	500–600 1050–1098 1100–1300 1550–1750	8/8	4	0.852	0.638	0.830	−0.123	1.37
Grown plant	Original	400–500 1600–1700 2200–2300	1/0	7	0.881	0.717	0.775	−0.054	1.94
	First derivative	400–500 1100–1200 1600–1700 2200–2300	2/2	5	0.919	0.597	0.726	−0.141	2.07
	Second derivative	400–500 1100–1200 1600–1800 2200–2300	3/3	5	0.944	0.502	0.685	−0.162	2.19

MPLSR = modified partial least squares regression; R_c = calibration set; RPD = ratio of the standard error of performance to the standard deviation of the reference values; SEC = standard error of calibration; SEV = standard error of validation.
* Interval is 2 nm.

the samples being tissue culture or grown plants, the calibration models built based on the first and second derivative spectra were better than those based on the original spectra, indicating that spectral pretreatments indeed enhanced the predictability of the calibration models. Spectral calibration models of grown plants were all better than those of the tissue culture, with fewer spectral pretreatments, which was consistent with the results shown in Table 3. The specific wavelength regions of tissue culture and grown plants were mainly distributed in 400–600 nm (blue and red light) and 800–1098 nm (the second and third overtones of C–H bond). Absorption capacity of these bands was slightly inferior to the combination and the first overtone of C–H bond, producing fewer spectral absorption performances of gentiopicroside; hence, the predictability declined slightly when the silicon CCD sensing band was used to build the calibration models.

In addition, the SMLR analysis results of the silicon CCD sensing band (400–1098 nm) are shown in Table 5. The best

Table 4 – Prediction of the gentiopicroside content in tissue culture and grown plants of *Gentiana scabra* Bunge by MPLSR models in the wavelength range of 400–1098 nm.

Sample	Spectrum	Wavelength range (nm)*	Smoothing points/gap	Factors	Calibration set		Validation set		
					R_c	SEC (%)	SEV (%)	Bias (%)	RPD
Tissue culture	Original	550–650 900–1050	9/0	4	0.704	0.866	1.084	−0.047	1.05
	First derivative	600–700 900–1000	6/6	6	0.764	0.786	0.906	−0.061	1.26
	Second derivative	400–500 800–1000	2/2	3	0.865	0.611	0.772	0.025	1.47
Grown plant	Original	400–600 950–1050	1/0	5	0.840	0.823	1.089	−0.015	1.38
	First derivative	400–600 900–1098	2/2	5	0.904	0.649	0.724	−0.089	2.08
	Second Derivative	400–650 950–1098	3/3	3	0.888	0.697	0.750	−0.100	2.00

MPLSR = modified partial least squares regression; R_c = calibration set; RPD = ratio of the standard error of performance to the standard deviation of the reference values; SEC = standard error of calibration; SEV = standard error of validation.
* Interval is 2 nm.

Table 5 – Prediction of the gentiopicroside content in tissue culture and grown plants of *Gentiana scabra* Bunge by SMLR models in the wavelength range of 400–1098 nm.

Sample	Spectrum	Specific wavelength (nm)	Smoothing points/gap	Calibration set		Validation set		
				R_c	SEC (%)	SEV (%)	Bias (%)	RPD
Tissue culture	Original	684	4/0	0.613	0.963	1.028	−0.064	1.11
		910, 512		0.643	0.934	0.999	0.000	1.14
	First derivative	612	2/2	0.654	0.922	1.060	−0.116	1.07
	Second derivative	848	3/3	0.632	0.946	1.016	0.376	1.12
		846, 932		0.750	0.806	0.990	0.270	1.15
Grown plant	Original	580	2/0	0.588	1.227	1.249	−0.076	1.20
		690, 480		0.689	1.099	1.329	−0.112	1.13
		436, 690, 420		0.759	0.988	1.284	−0.154	1.17
		966, 420, 408, 436		0.802	0.906	1.186	−0.178	1.27
	First derivative	730	2/2	0.590	1.225	1.265	−0.074	1.19
		462, 676		0.725	1.044	0.889	−0.041	1.69
		684, 780, 462		0.806	0.897	0.936	0.072	1.61
		650, 780, 462, 512		0.850	0.799	0.823	0.008	1.83
	Second derivative	468	3/3	0.626	1.182	1.122	0.001	1.34
		460, 634		0.736	1.027	1.011	−0.250	1.49
		666, 788, 474		0.834	0.838	0.897	−0.144	1.67
		670, 786, 474, 826		0.860	0.775	0.848	−0.134	1.77

R_c = calibration set; RPD = ratio of the standard error of performance to the standard deviation of the reference values; SEC = standard error of calibration; SEV = standard error of validation; SMLR = stepwise multiple linear regression.

calibration model of *G. scabra* Bunge tissue culture was found when the second derivative spectra were used. Both smoothing points and gap were at 3, with the specific wavelengths of 846 nm and 932 nm, which yielded R_c = 0.750, SEC = 0.806%, SEV = 0.990%, bias = 0.270%, and RPD = 1.15. The best calibration model of a grown plant was attained when the second derivative spectra were employed, where both smoothing points and gap were set at 3, in the combination of four wavelengths (670 nm, 786 nm, 474 nm, and 826 nm); the results were R_c = 0.860, SEC = 0.775%, SEV = 0.848%, and bias = −0.134%, RPD = 1.77. Calibration models built based on the first and second derivative spectra were all better than those based on the original spectra, indicating that spectral pretreatments reduced the noise influence and made the combination of selected wavelengths more consistent when the number of wavelengths increased. The selected specific wavelengths given in Table 5 were similar to those listed in Tables 3 and 4, with only a small number of specific wavelengths beyond those selected through the MPLSR analysis. Because the silicon CCD sensing band contains less information on gentiopicroside content, and the spectral calibration model built using SMLR was based on the combination of a small number of wavelengths, giving less spectral information than MPLSR, the analysis results seemed slightly inferior to those given in Tables 3 and 4. Compared to the tissue culture that can apply only two wavelengths at most for inspection, a grown plant can apply four wavelengths to build the calibration model, consequently improving its predictability.

4. Conclusion

This study applied NIR spectroscopy for a quantitative analysis of gentiopicroside present in the medicinal plant *G. scabra* Bunge. Spectral pretreatments of MSC in combination with the second derivative were found to reduce the spectral noise caused by the heterogeneous particle of *G. scabra* Bunge powder. The specific wavelength regions or specific wavelengths selected based on their characteristic response to gentiopicroside could improve the predictability of calibration models effectively. This study successfully built spectral calibration models for *G. scabra* Bunge tissue culture and grown plants, enabling a quantitative inspection of the bioactive component gentiopicroside in *G. scabra* Bunge during its different growth stages. The specific wavelengths selected in the silicon CCD sensing band can be used as the foundation to establish a nondestructive and rapid method to assess the quality of *G. scabra* Bunge using multispectral imaging.

Acknowledgments

This work was supported by the Industrial Development Bureau, Ministry of Economic Affairs (09611101087-9601). We thank Mr Cheng-Wei Huang, Mr Yu-Song Chen, and Mr Chun-Chi Chen for their assistance.

REFERENCES

[1] Kakuda R, Iijima T, Yaoita Y, et al. Secoiridoid glycosides from *Gentiana scabra*. J Nat Prod 2001;64:1574–5.
[2] Kim JA, Son NS, Son JK, et al. Two new secoiridoid glycosides from the rhizomes of *Gentiana scabra* Bunge. Arch Pharm Res 2009;32:863–7.
[3] Zhang HL, Xue SH, Pu F, et al. Establishment of hairy root lines and analysis of gentiopicroside in the medicinal plant *Gentiana macrophylla*. Russ J Plant Physiol 2010;57:110–7.

[4] Cai Y, Liu Y, Liu Z, et al. High-frequency embryogenesis and regeneration of plants with high content of gentiopicroside from the Chinese medicinal plant *Gentiana straminea* Maxim. In Vitro Cell Dev Biol Plant 2009;45:730–9.
[5] Glatz Z, Pospísilová J, Musil P. Determination of gentiopicroside in extracts of *Centaurium erythreae* and *Gentiana lutea* by micellar electrokinetic capillary chromatography. J Liq Chromatogr Relat Technol 2000;23:1831–9.
[6] Szücs Z, Dános B, Nyiredy SZ. Comparative analysis of the underground parts of *Gentiana* species by HPLC with diode-array and mass spectrometric detection. Chromatographia 2002;56:S19–23.
[7] Kikuchi M, Kakuda R, Kikuchi M, et al. Secoiridoid glycosides from *Gentiana scabra*. J Nat Prod 2005;68:751–3.
[8] Carnat A, Fraisse D, Carnat AP, et al. Influence of drying mode on iridoid bitter constituent levels in gentian root. J Sci Food Agric 2005;85:598–602.
[9] Kušar A, Šircelj H, Baričevič D. Determination of seco-iridoid and 4-pyrone compounds in hydro-alcoholic extracts of *Gentiana lutea* L. subsp. *symphyandra* Murb. leaves and roots by using high performance liquid chromatography. Isr J Plant Sci 2010;58:291–6.
[10] Hayta S, Akgun IH, Ganzera M, et al. Shoot proliferation and HPLC-determination of iridoid glycosides in clones of *Gentiana cruciata* L. Plant Cell Tissue Organ Cult 2011;107:175–80.
[11] Hayta S, Gurel A, Akgun IH, et al. Induction of *Gentiana cruciata* hairy roots and their secondary metabolites. Biologia 2011;66:618–25.
[12] Aberham A, Schwaiger S, Stuppner H, et al. Quantitative analysis of iridoids, secoiridoids, xanthones and xanthone glycosides in *Gentiana lutea* L. roots by RP-HPLC and LC-MS. J Pharm Biomed Anal 2007;45:437–42.
[13] Aberham A, Pieri V, Croom Jr EM, et al. Analysis of iridoids, secoiridoids and xanthones in *Centaurium erythraea, Frasera caroliniensis* and *Gentiana lutea* using LC-MS and RP-HPLC. J Pharm Biomed Anal 2011;54:517–25.
[14] Nastasijević B, Lazarević-Pašti T, Dimitrijević-Branković S, et al. Inhibition of myeloperoxidase and antioxidative activity of *Gentiana lutea* extracts. J Pharm Biomed Anal 2012;66:191–6.
[15] Woo YA, Kim HJ, Ze KR, et al. Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of Angelicae gigantis Radix. J Pharm Biomed Anal 2005;36:955–9.
[16] Zhang Z, Tang Y. Identification of rhubarb samples by using NIR spectrometry and Takagi–Sugeno fuzzy system. Spectr Lett 2005;38:447–57.
[17] Wang L, Lee FSC, Wang X. Near-infrared spectroscopy for classification of licorice (*Glycyrrhizia uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content. LWT Food Sci Technol 2007;40:83–8.
[18] Chen X, Wu D, He Y, et al. Nondestructive differentiation of *Panax* species using visible and shortwave near-infrared spectroscopy. Food Bioprocess Technol 2011;4:753–61.
[19] Wu Z, Xu B, Du M, et al. Validation of a NIR quantification method for the determination of chlorogenic acid in *Lonicera japonica* solution in ethanol precipitation process. J Pharm Biomed Anal 2012;62:1–6.
[20] Paris I, Janoly-Dumenil A, Paci A, et al. Near infrared spectroscopy and process analytical technology to master the process of busulfan paediatric capsules in a university hospital. J Pharm Biomed Anal 2006;41:1171–8.
[21] Jamrógiewicz M. Application of the near-infrared spectroscopy in the pharmaceutical technology. J Pharm Biomed Anal 2012;66:1–10.
[22] Porfire A, Rus L, Vonica AL, et al. High-throughput NIR-chemometric methods for determination of drug content and pharmaceutical properties of indapamide powder blends for tabletting. J Pharm Biomed Anal 2012;70:301–9.
[23] Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 2001;58:109–30.
[24] Chang WH, Chen S, Tsai CC. Development of a universal algorithm for use of NIR in estimation of soluble solids in fruit juices. Trans ASAE 1998;41:1739–45.
[25] de Noord OE. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. Chemom Intell Lab Syst 1994;23:65–70.
[26] Fearn T. Standardisation and calibration transfer for near infrared instruments: a review. J Near Infrared Spec 2001;9:229–44.
[27] Helland IS, Næs T, Isaksson T. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. Chemom Intell Lab Syst 1995;29:233–41.
[28] Maleki MR, Mouazen AM, Ramon H, et al. Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. Biosyst Eng 2007;96:427–33.
[29] Isaksson T, Næs T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. Appl Spectrosc 1988;42:1273–84.
[30] Thennadil SN, Martens H, Kohler A. Physics-based multiplicative scatter correction approaches for improving the performance of calibration models. Appl Spectrosc 2006;60:315–21.