Research Article

Phytochemical Analysis

Received: 4 July 2009; Revised: 31 March 2010; Accepted: 1 April 2010 Published online in Wiley Online Library: 2 July 2010

(wileyonlinelibrary.com) DOI 10.1002/pca.1242

Discrimination of Geographical Origin and Adulteration of Radix Astragali using Fourier Transform Infrared Spectroscopy and Chemometric Methods

Lei Zhang and Lei Nie*

ABSTRACT:

Introduction – Radix Astragali, one of most widely used and important traditional Chinese medicines, is cultivated in different geographical regions. Because of varying growing conditions, the qualities of Radix Astragali vary, which can give rise to differences in clinical therapy. Detecting adulteration is a routine requirement in pharmaceutical practice.

Objective – To develop a simple and accurate approach to discriminate the geographical origin and potential adulteration of Radix Astragali, derived from the root of *Astragalus membranaceus* (Fischer) Bunge var. *mongholicus* (Bunge) Hsiao, using Fourier transform infrared (FT-IR) spectroscopy and chemometric methods.

Methodology – To obtain characteristic IR spectra for accurate discrimination, a one-solvent extraction method was utilised following a novel evaluation method for selecting appropriate solvents. Samples of Radix Astragali from different geographical origins were discriminated using FT-IR spectroscopy and discriminant partial least squares (DPLS) methods. FT-IR spectroscopy combined with Mahalanobis distance was employed to detect adulteration of Radix Astragali.

Results – In comparison with other solvents, butanone was more effective at extracting samples. Radix Astragali samples were accurately assigned to their corresponding geographical origins by using FT-IR spectroscopy and DPLS method. Most adulterated samples were detected accurately by application of FT-IR spectroscopy combined with Mahalanobis distance.

Conclusion – FT-IR spectroscopy combined with chemometric method was developed and demonstrated to be a useful tool to discriminate geographical origin and adulteration of Radix Astragali. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: Fourier transform infrared spectroscopy; chemometrics; Radix Astragali; discrimination; geographical origin; adulteration detection

Introduction

Traditional Chinese medicines (TCMs) have been used for thousands of years in Asian countries. They have been attracting more and more attention owing to their complementary therapeutic effects to Western medicines and low toxicities, leading to few or even no complications (Normile, 2003; Xue and Roy, 2003).

Radix Astragali, one of the most widely used TCMs, is known worldwide to reinforce "Qi". Two plant sources for Radix Astragali, described in the *Chinese Pharmacopoeia*, are *Astragalus membranaceus* (Fischer) Bunge and *A. membranaceus* (Fisch.) Bunge var. *mongholicus* (Bunge) Hsiao. Pharmacological studies and clinical practice have demonstrated that it possesses various bioactivities, including immunostimulant, tonic, antioxidant, hepatoprotective, diuretic, anti-diabetic, anti-cancer and expectorant properties (Lin *et al*., 2000; Ma *et al*., 2002). Chemical investigations into Radix Atragali resulted in the discovery of several kinds of bioactive components, such as isoflavonoids, triterpene saponins, polysaccharides, γ-amino butyric acids and various trace elements (Ma *et al*., 2002).

The physiological and pharmacological activities of Radix Astragali depend on the constituents it contains. There are more than 10 cultivation regions for Radix Astragali in China. The discrepancies in chemical composition and properties of Radix Astragali due to soil and climate differences as well as the collection and processing methods in different geographical origins (Yip and Kwan, 2006; Song *et al*., 2008), make it difficult to keep therapeutic potency consistent. Therefore, it is necessary to identify where the Radix Astragali is cultivated.

Since IR spectroscopy is fast, nondestructive and relatively accurate and dependable, it has been applied to discriminate Chinese herbal medicines according to geographical origins (Hua *et al*., 2003). Generally, because there are tens of constituents in a crude drug, the spectral bands are often overlapping and result in low resolution of the IR spectrum. To improve the resolution of the spectrum, data-processing approaches like second derivative and two-dimensional correlation techniques have been used (Hua *et al*., 2003). However, some common components (for example, starch and resin) in herbal medicines also contribute to the IR absorbance and can cause significant IR spectral band broadening, making it difficult to obtain a characteristic IR spectrum for each herb just using the crude drug powder. To resolve this problem, an appropriate solvent extraction is used to provide

* Correspondence to: Lei Nie, School of Pharmaceutical Sciences, Shandong University, No. 44 Wenhua Xi Road, Jinan, Shandong 250012, People's Republic of China. E-mail: nielei2004@sdu.edu.cn

School of Pharmaceutical Sciences, Shandong University, Jinan 250012, China

a 'cleaner' sample for IR spectroscopy analysis. (Wang and Yuan, 2004).

If traditional Chinese medicines are adulterated, their clinical efficacy and compatibility with prescription medicines may be adversely affected, posing a hazard to health. In commercial herb markets, shredded slices of Radix Astragali mixed with adulterants are often found (Jiang and Ye, 2006). Accidental contamination during transport, storage or production processes are possible. Chromatographic approaches (Ma *et al*., 2002) and molecular genetic methods (Chen *et al*., 2005) have been developed in the past for identification of specific adulterants but not for detection of adulteration. These methods are time-consuming, labour-intensive and expensive. In the present work, to detect adulteration of Radix Astragali, a red substitute, *Hedysarum polybotrys* Handel–Mazzetti, which sometimes carries the name of Radix Astragali on the commercial market, was chosen as an adulterant (Ma *et al*., 2002). The possibility of using IR spectroscopy combined with chemometric methods to detect adulteration of Radix Astragali was explored.

Radix Astragali contains many components leading to extensively overlapped bands in IR spectra of samples. Dissimilarities between samples may cause slight differences that are very difficult to distinguish with the naked eye and, therefore, chemometric methods are required for differentiating Radix Astragali from various geographical origins and for detecting adulteration. Discriminant partial least squares (DPLS) has been frequently used for classification due to its powerful discriminating ability and ease of understanding classification models (Alsberg *et al*., 1998). Mahalanobis distance could be a candidate method appropriate for detection of adulteration because it can determine whether unknown samples belong to an existing class or not according to the probabilities calculated by Hotelling's statistics (Gemperline and Boyer, 1995).

The objective of this study was to develop a simple and accurate approach that could easily discriminate Radix Astragali based on geographical origin and detect its adulteration using FT-IR spectroscopy and chemometric methods.

Experimental

Plant materials

The Radix Astragali used was derived from dried root of *Astragalus membranaceus* (Fischer) Bunge var. *mongholicus* (Bunge) Hsiao. All samples of Radix Astragali were obtained from the north of China (Table 1). Samples of *H. polybotrys* Handel–Mazzetti were all collected from Min country, Gansu Province. All samples were authenticated by Professor Y. L. Chen of the Department of Pharmacognosy of Shandong University. Voucher specimens were deposited in the Department of Pharmaceutical Analysis, School of Pharmaceutical Sciences, Shandong University.

Chemicals and reagents

n-Hexane, ether, chloroform, ethyl acetate and butanone of analytical grade were purchased from Guangcheng chemical company (Tianjin, China).

Apparatus and FT-IR spectra

A Nicolet Nexus 470 FT-IR spectrometer (Thermo Nicolet, USA) equipped with a deuterated triglycine sulphate (DTGS) detector was used. FT-IR spectra were acquired over 4000–400 cm^{-1} with the resolution of 4 cm^{-1}. Each sample spectrum was averaged from a total of 64 scans. All spectra were processed with automatic baseline correction and *y*-scale normalisation (the lowest and the highest absorbances were set to 0 and 1, respectively) to minimise the problem arising from baseline drift.

Data sets

IR spectra of samples form two data sets. Data set 1 includes IR spectra of 86 samples of Radix Astragali, which were divided into training sets and test sets by a random selection shown in Table 1. There are 15 spectra of samples of Radix Astragali from Inner Mongolia province (nos 1–15) and nine spectra of samples of Radix Astragali adulterated with different levels (g/g) of *H. polybotrys* (nos 16–24) in data set 2 (Table 2).

Sample preparation

The dried roots of Radix Astragali were milled to pass through a no. 40 mesh sieve. An accurately weighed (ca. 2.0 g) aliquot of milled sample

Table 1. Data set 1

	Geographic origins of Radix Astragali				
	Min county, Gansu province	Jian county, Jilin province	Damao county, Inner Mongolia province	Guangling county, Shanxi province[a]	Xinzhou county, Shanxi province
Number of spectra in training set	11	12	9	11	10
Number of spectra in test set	7	8	6	6	6

[a] Because Guangling and Xinzhou are different counties in Shanxi province, they can be regarded as different geographic origins.

Table 2. Data set 2

	Radix Astragali[a]	The percentage of samples of Radix Astragali adulterated with *H. polybotrys*[b]						
	0	1%	5%	10%	30%	50%	80%	100%
Number of spectra	15	1	1	1	1	1	1	3
Sample	1–15	16	17	18	19	20	21	22–24

[a] Samples of Radix Astragali from Damao county, Inner Mongolia province.
[b] The percentage is weight ratio of the adulterant to Radix Astragali.

was added to a conical flask and mixed with solvent (30 mL): *n*-hexane, ether, chloroform, ethyl acetate or butanone. The mixture was then ultrasonicated at room temperature for 30 min. The solvent extract was filtered and evaporated to dryness using a rotary evaporator. About 5 mg of the residue was uniformly smeared onto a potassium bromide plate and the corresponding spectrum was collected.

To detect adulteration of Radix Astragali, dried roots of three samples of *H. polybotrys* were milled and the powders were passed through a no. 40 mesh sieve. A small portion of powder of each sample of *H. polybotrys* was taken out and mixed well as an adulterant. The same process was also applied to each sample of Radix Astragali from Inner Mongolia province. Afterward, the powders of Radix Astragali and its adulterant were accurately weighed according to the ratios shown in Table 2 and then were mixed and extracted as described for Radix Astragali.

Data analysis

Automatic baseline correction and *y*-scale normalisation for all IR spectra of samples were performed with OMNIC software (Thermo Electron Corporation). DPLS and Mahalanobis distance were carried out using MATLAB (The Math Work Inc.).

Results and Discussion

Solvent selection

Three samples were chosen randomly from each geographic origin i.e. a total of 15 samples from five origins were selected. Five solvents with varying polarities, i.e. *n*-hexane, ether, chloroform, ethyl acetate and butanone were tested. Fifteen FT-IR spectra were recorded for extracts using each solvent and classified by hierarchical clustering analysis using the ward linkage method, which is relatively powerful compared with other methods (Morey *et al*., 1983). The B_k measure (Fowlkes and Mallows, 1983) was used to evaluate the similarity between two hierarchical clusterings, which were labeled C_1 and C_2, respectively. In this study, C_1 represents 15 FT-IR spectra for each solvent, and C_2 is based on real classes according to geographical origins. The B_k measure is defined as:

$$B_k = F_k \Big/ \sqrt{G_k H_k} \tag{1}$$

where

$$F_k = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij}^2 - n \tag{2}$$

$$G_k = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{i\bullet}^2 - n \tag{3}$$

$$H_k = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{\bullet j}^2 - n \tag{4}$$

$$m_{i\bullet} = \sum_{j=1}^{k} m_{ij} \tag{5}$$

$$m_{\bullet j} = \sum_{i=1}^{k} m_{ij} \tag{6}$$

Table 3. Results of solvent selection

	Solvents				
	Butanone	Chloroform	Ether	Ethyl acetate	*n*-Hexane
Similarity	0.592	0.385	0.520	0.377	0.520

$$n = \sum_{i=1}^{k}\sum_{j=1}^{k} m_{ij} \tag{7}$$

and m_{ij} refers to the element in matching matrix, which is $M = [m_{ij}]$ ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, k$). The quantity m_{ij} is the number of samples in common between the *i*th cluster of C_1 and the *j*th cluster of C_2; *n*, the number of samples, is equal to 15; and *k*, the number of clusters, is set to 5, corresponding to five geographic origins. The greater the value of B_k measure is, the more similar the two hierarchical clusterings are. The B_k measure has an upper bound of 1 when the two clusterings are identical and a lower bound of 0.

If C_1 based on one solvent is more similar to C_2 than a C_1 based on other solvents, it indicates that IR spectra obtained from this solvent could have more distinct characteristics. This enables better differentiation of samples of Radix Astragali from different geographic origins, and suggests that the solvent could be more appropriate for extraction. Table 3 summarises the results of solvent selection. Thus, cluster analysis based on butanone has the best similarity to the real class categorised by geographical origin; therefore, butanone was selected as an appropriate solvent for extraction.

The same experiment of solvent selection was also carried out for detection of adulteration. Three samples of Radix Astragali from Inner Mongolia province and three samples of its adulterant from Gansu province were used. The same five solvents were tested. The value of B_k measured for each solvent is equal to 1, suggesting that corresponding IR spectra based on these solvents all have clear characteristics for classifying Radix Astragali and *H. polybotrys*. Butanone was still selected as the extraction solvent for detection of adulteration.

IR spectra of samples from different geographical origins

The averaged spectra of samples from each geographical origin were compared and are shown in Fig. 1. Although the whole IR spectra of the five different origins are rather similar, noticeable differences can still be observed in some regions of IR spectra (Fig. 1A). For example, the ratios of intensities between two peaks near 1725 cm^{-1} are different. The magnitude of peak 1 is higher than that of peak 2 in IR spectra of samples from Guangling county (Shanxi Province), Jian county (Jilin Province) and Min county (Gansu province), while the opposite order of the two-peak intensities can be found in IR spectra of samples from Xinzhou county (Shanxi province) and Daomao county (Inner Mongolia province). Additionally, apparent dissimilarities are also observed in the region ranging from 1550 to 850 cm^{-1}, which was expanded and plotted at the same baseline level for clarification and convenient comparison (Fig. 1B). The band around 1070 cm^{-1} is clearly different. Since all of the constituents extracted in butanone contribute to the total absorbance of IR, the dissimilarities between corresponding IR spectra indicate that samples of Radix Astragali from the five geographical origins vary in their composition to some extent.

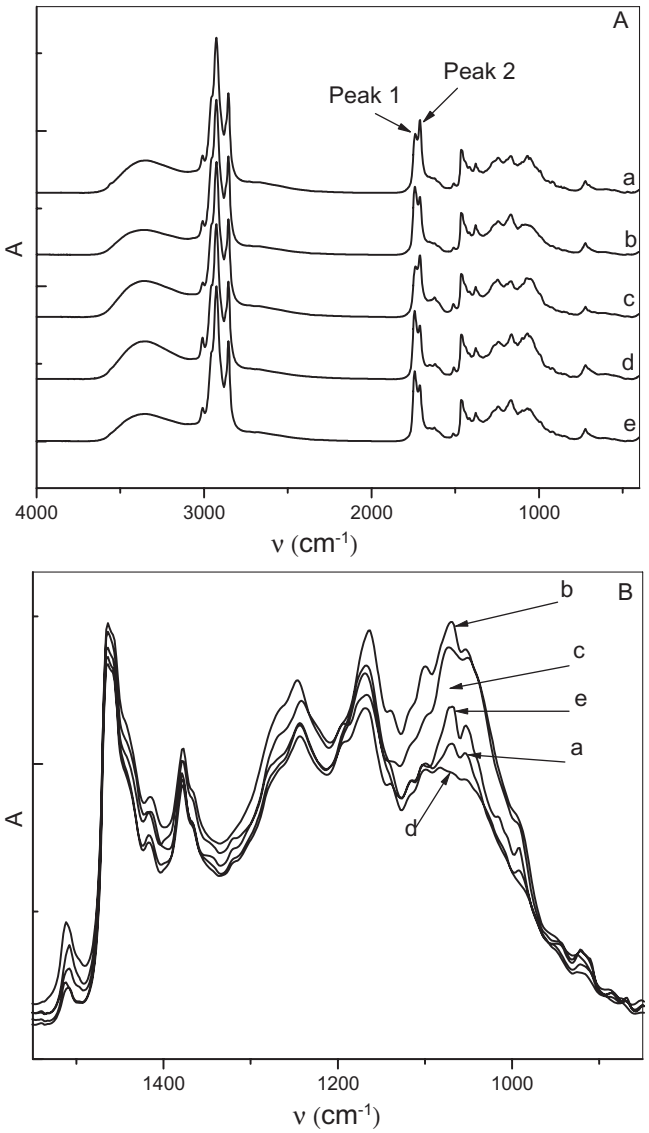

Figure 1. Representative IR spectra of Radix Astragali samples from different geographical regions, including the whole IR spectra (A) and expansions plotted at the same baseline for clarification and convenient comparison (B). The letters a–e refer to Xinzhou county (Shanxi province), Guangling county (Shanxi province), Daomao county (Inner Mongolia province), Jian county (Jilin Province) and Min county (Gansu province), respectively.

Discrimination of Radix Astragali according to geographical origins with DPLS

Although some differences in the IR spectra were observed, the IR spectra corresponding to Radix Astragali from different origins are so similar that it is very hard to distinguish them. Discrimination of Radix Astragali according to geographical origins could, in essence, be regarded as the problem of pattern recognition. It would be impossible or impractical to determine the 'similarities' and "differences" systematically among the IR spectra of samples from different origins without the aid of a chemometric method for classification, especially for differentiating large numbers of samples. The DPLS method was used to discriminate the samples based on geographical origins. The DPLS calibration model can be simply described as a regression analysis model. The training set (Table 1) was utilised to construct the DPLS model and acquire regression coefficients. These coefficients were then used to predict class membership of samples in test set (Table 1). The performance of the DPLS calibration model can be validated according to the prediction results.

Since IR spectrum variables (each wavenumber regarded as one variable) are numerous and correlated, there is substantial risk of "over-fitting" (Wold *et al.*, 2001), indicating that model fitted well has poor or no predictive power. Therefore, it is necessary to test the significance of each PLS component. In addition, not all of IR spectrum variables contribute greatly to discrimination of Radix Astragali based on geographical origins. Some variables containing irrelevant classification information may be redundant. Removal of these variables and maintenance of the important ones can not only improve the predictive ability but also simplify the established model. Some approaches related to variable selection have been proposed (Centner *et al.*,1996; Cai *et al.*, 2008). In the present work, the PLS-VIP scores method based on the DPLS calibration model was used for IR spectrum variable selection (i.e. wavenumber selection) due to its excellent performance in variable selection (Chong and Jun, 2005). Two parameters, the number of PLS components and the selected variables, are required to construct the optimal DPLS model, which is evaluated by predictive residual sum of squares (PRESS) acquired by leave-one-out cross-validation (Hastie *et al.*, 2001). The lower the value of PRESS, the better the predictive performance of the DPLS model is. When the minimum value of PRESS is achieved, the optimal number of PLS components and the selected variables are both determined.

Figure 2 shows the selected variables in the IR spectra (upper part) by using the PLS-VIP scores method (greater than 1 rule used as a criterion for variable selection (Chong and Jun, 2005) based on an optimal DPLS model. Original variables of IR spectra (i.e. original IR spectra) are also shown in Fig. 2 (lower part) for convenient comparison. The selected variables containing relevant information on classification are important for discrimination of Radix Astragali according to their origins.

Figure 3 shows the PRESS values as a function of the number of PLS components. The number of PLS components is equal to 24 when PRESS reaches its minimum value. The trend shown in Fig. 3 exhibits a steep decrease followed by a bend and then a plateau. This gives an indication that the number of PLS components can be 6, 11 or 14, whereas 24 seems too big. Because the value of PRESS corresponding to 14 components is very close to the minimum value of PRESS, 14 components are appropriate for the model. When the optimal DPLS model is used for prediction of class membership of samples in test set, the predicted results are acquired from a matrix of dependent variables, **Y**, containing information about class membership of samples. Ideally, if K is the number of classes (i.e. five geographic origins), each row, y^T, in the **Y** matrix has the following structure: $y_k^T = 1$ if sample belongs to class k, or $y_k^T = 0$ otherwise; y_k refers to the kth column in **Y**, k is the class number, and we have $k = 1, 2, \ldots, K$. The superscript T represents transpose. However, the estimated matrix of **Y** predicted by the optimal DPLS model does not have such a perfect structure. The predicted values are real numbers. Each row in the estimated matrix of **Y** represents each sample in the test set. The class membership of each sample is given by the column index corresponding to the element with the maximum of predicted absolute values in each row.

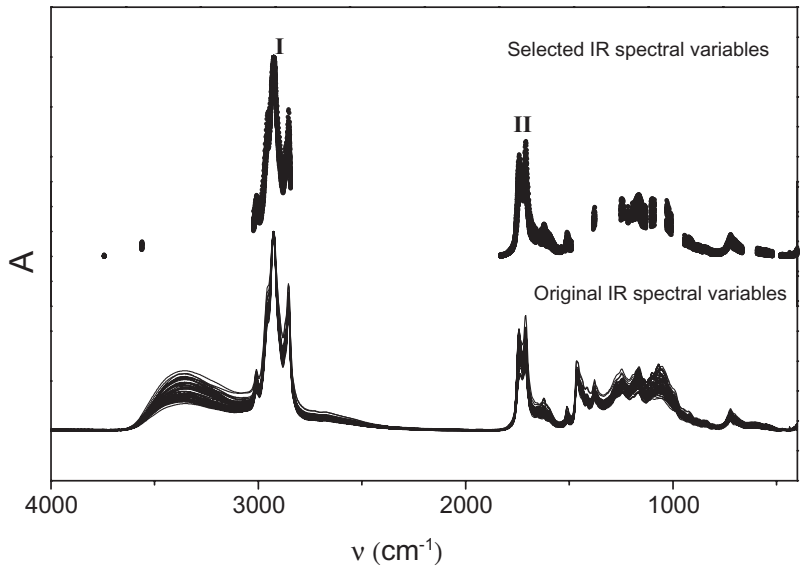

Figure 2. Original IR spectral variables and the variables selected by the PLS-VIP scores method.

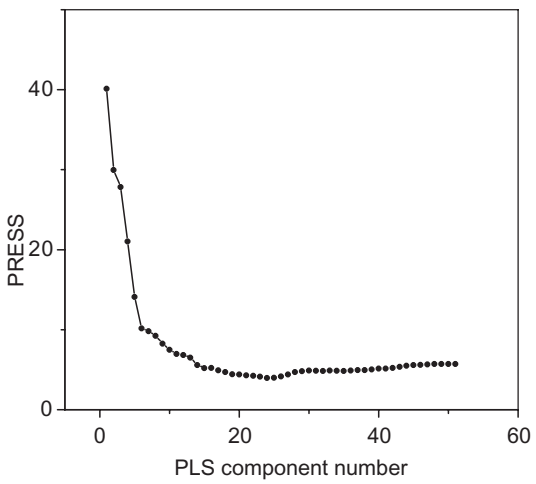

Figure 3. PLS component number determined by PRESS values from leave-one-out cross-validation.

Figure 4 illustrates the predicted results based on the optimal DPLS model. Each column in the estimated matrix of **Y** is shown in Fig. 4(a–e), respectively. The class membership for each sample relies on the maxima of predicted absolute values among different categories, i.e. a–e. For example, the maxima of samples from 16 to 21 all lie in Fig. 4(c), so they all belong to Damao county, Inner Mongolia province. As shown in Fig. 4, all samples in the test set are accurately assigned to their corresponding geographical origins. According to the selected variables, some useful information on chemical components can be obtained. For instance, the selected variables falling in the region from 3040 to 2830 cm^{-1} (around peaks I in Fig. 2) are associated with the bands corresponding to C–H stretching vibration, indicating that constituents containing these hydrogen groups may be essential to classification. Similarly, the region of 1815 –1490 cm^{-1} (around peaks II shown in Fig. 2) exhibits IR absorptions from a wide variety of double-bonded functional groups, such as C=O, C=N and C=C (Settle, 1997), suggesting that constituents including these functional groups may be helpful for differentiation of Radix Astragali based on various origins.

Radix Astragali adulteration detected by Mahalanobis distance

Evaluating the degree of homogenisation of various components in mixtures of samples of Radix Astragali with different levels of adulterants is essential for detection of adulteration. The one-solvent extraction method can make samples of Radix Astragali that are mixed well with adulterants because components contained in Radix Astragali and its adulterant distribute more homogeneously in the liquid phase than in the solid phase.

For detection of adulteration, the Mahalanobis distance (Mahalanobis, 1936) was used. It is a useful way to determine whether the unknown samples belong to the known set according to the confidence level calculated by Hotelling's statistics. The *H. polybotrys* samples and the samples of Radix Astragali adulterated with different percentages of *H. polybotrys* can be regarded as the unknown samples, and the 15 unadulterated samples of Radix Astragali (Table 2) form the known set. Figure 5 shows the averaged spectra taken from samples of Radix Astragali or *H. polybotrys* as representative known samples or adulterants, respectively. The IR spectrum of Radix Astragali is similar to that of its adulterant, while observable differences in the relative peak intensities in some spectral regions are still found.

Mahalanobis distance was used to classify the adulterated samples as acceptable or unacceptable for the known set based on a threshold of probability level. Here, the threshold value is set to 5%. If samples have a probability level in the range from 0.05 to 1.0, they are classified as members of the known set. Samples having a probability level in the range from 0 to 0.05 are classified as nonmembers. In this paper, principal components analysis is employed to calculate Mahalanobis distance (Gemperline and Boyer, 1995; Aldridge *et al*., 1996). Critical values at probability level α are obtained by *F* distribution (Gemperline and Boyer, 1995; Aldridge *et al*., 1996). The sample's Mahalanobis distance, *D*, is compared with critical values of *F* according to equation (8)

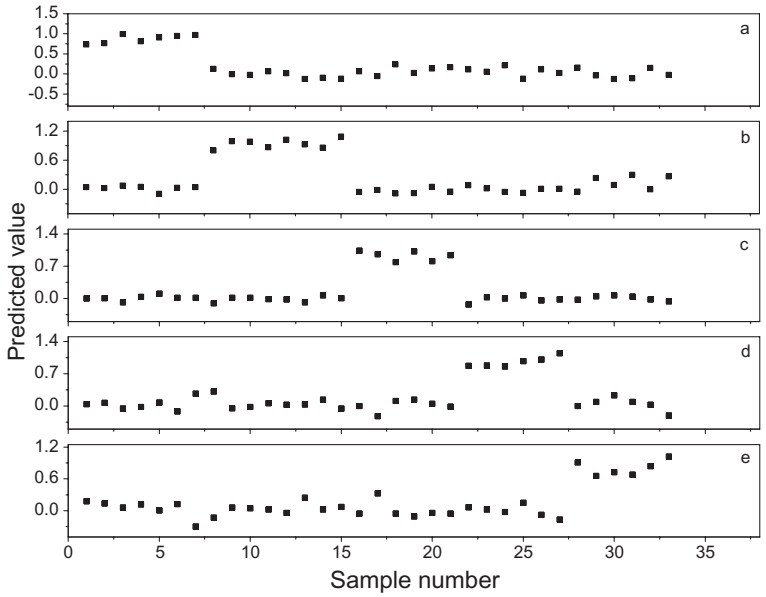

Figure 4. Predicted values of samples in the test set by using an optimal DPLS model. Letters a–e are the same ones as used for Figure 1.

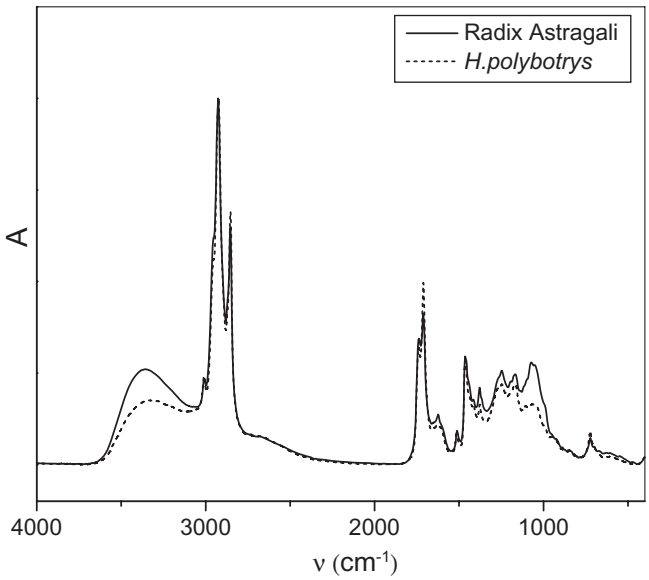

Figure 5. Representative IR spectra of Radix Astragali and *H. polybotrys*.

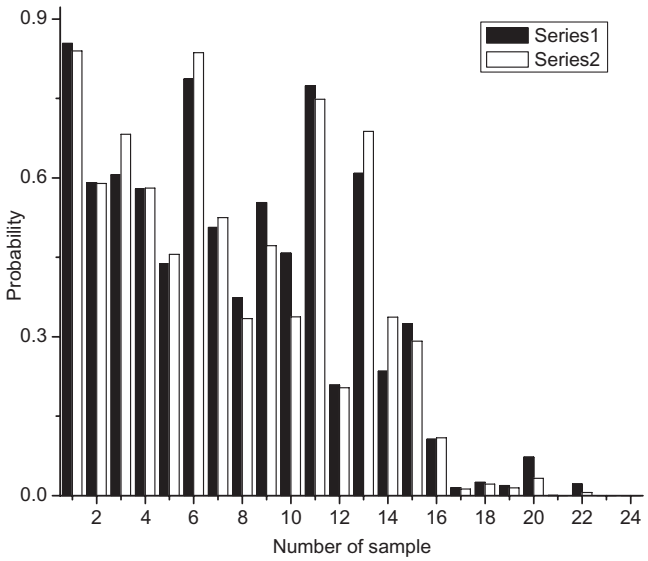

Figure 6. Probabilities based on Mahalanobis distances for unadulterated (nos 1–15) and adulterated samples (nos 16–24). Series 1 and 2 denote that probabilities are calculated by using the whole region (4000–400 cm^{-1}) and selected region (4000–1300 cm^{-1}) of IR spectra, respectively.

to determine whether the sample is acceptable or unacceptable for the known set.

$$D^2 \frac{(n-a)}{a(n-1)} > F_{a,n-a}(\alpha) \quad (8)$$

where n is the number of training samples, and a is the number of principal component scores used.

Figure 6 shows the results of detecting adulteration of Radix Astragali. The whole IR spectral region (4000–400 cm^{-1}) was used for calculating Mahalanobis distance and the corresponding probability level (Fig. 6, series 1). The number of principal components is set to 3 (84.4% cumulative variance) because this is sufficient to account for a specific percentage of total variance (80%) (Rencher, 2002).

As seen from probabilities based on the whole region in Fig. 6, the adulterated samples, except samples 16 and 20, have probability levels lower than 5% and are classified correctly. The rejection rate is 77.8%. The results suggest that Mahalanobis distance is sensitive to adulterated samples.

IR spectra consist of overlapping bands stemming from interaction of various components in the sample extracts, so different IR spectral regions may contain different information that gives rise to various capabilities for detecting adulteration. Commonly, an IR spectrum is divided into three regions: the functional group region (4000–1300 cm^{-1}), the fingerprint region (1300–910 cm^{-1}) and the aromatic region (910–650 cm^{-1}) (Settle, 1997). A criterion

Table 4. Selection of informative regions of IR spectra

	Three regions of IR spectra and their combination						
	4000–1300 cm^{-1}	1300–910 cm^{-1}	4000–910 cm^{-1}	910–650 cm^{-1}	910–650 and 4000–1300 cm^{-1}	1300–650 cm^{-1}	4000–650 cm^{-1}
RR[a] (%)	88.9	44.4	77.8	33.3	88.9	44.4	77.8
SP[b]	0.198	0.759	0.259	2.094	0.199	0.738	0.263

[a] RR refers to rejection rate.
[b] SP represents sum of probabilities for adulterated samples.

utilised for selecting the most informative regions to detect adulteration of Radix Astragali includes two requirements: high rejection rates and a minimum of sum of probabilities for adulterated samples (nos 16–24 shown in Fig. 6). Here, three spectral regions and their combination were considered for this purpose. Although different regions contain various numbers of spectral variables, it is possible to compare them owing to the corresponding probabilities employed for comparison. Because the number of principal components and the number of samples in the known set have been considered in the process of calculating probability, consistent performance will be achieved.

Table 4 gives the results of selected informative regions. According to the criterion mentioned above, the most informative region is the functional group region (4000–1300 cm^{-1}), and the corresponding rejection rate improves up to 88.9% as compared with the value of rejection rate based on the whole region. The number of principal components is still set to 3 (86.32% cumulative variance explained) and corresponding probabilities of samples are shown in Fig. 6 (series 2). Compared with the other two regions, the functional group region may consist of more useful information and have better capacity for detecting adulteration of Radix Astragali. Information from this region of IR spectra may be a better reflection of differences in the compositions of Radix Astragali and *H. polybotrys*.

Acknowledgements

Support from the Shandong Province Natural Science Foundation, China (no. Q2006C05) for this study is gratefully acknowledged.

References

Aldridge PK, Evans CL, Ward HW, Colgan ST, Boyer N, Gemperline PJ. 1996. Near-IR detection of polymorphism and process-related substances. *Anal Chem* **68**: 997–1002.
Alsberg BK, Kell DB, Goodacre R. 1998. Variable selection in discriminant partial least-squares analysis. *Anal Chem* **70**: 4126–4133.
Cai WS, Li YK, Shao XG. 2008. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemom Intell Lab Syst* **90**: 188–194.
Centner V, Massart DL, Noord OE, Jong S, Vandeginste BM, Sterna C. 1996. Elimination of uninformative variables for multivariate calibration. *Anal Chem* **68**: 3851–3858.
Chen G, Wang XL, Wong WS, Liu XD, Xia B, Li N. 2005. Application of 3 ′ untranslated region (UTR) sequence-based amplified polymorphism analysis in the rapid authentication of Radix Astragali. *J Agric Food Chem* **53**: 8551–8556.
Chong IG, Jun CH. 2005. Performance of some variable selection methods when multicollinearity is present. *Chemom Intell Lab Syst* **78**: 103–112.
Fowlkes EB, Mallows CL. 1983. A method for comparing two hierarchical clusterings. *J Am Stat Assoc* **78**: 553–569.
Gemperline PJ, Boyer NR. 1995. Classification of near-infrared spectra using wavelength distances: comparison to the Mahalanobis distance and residual variance methods. *Anal Chem* **67**: 160–166.
Hastie T, Tibshirani R, Friedman J. 2001. *The Elements of Statistical Learning*. Springer: New York; 214–217.
Hua R, Sun SQ, Zhou Q, Noda I, Wang BQ. 2003. Discrimination of fritillary according to geographical origin with Fourier transform infrared spectroscopy and two-dimensional correlation IR spectroscopy. *J Pharm Biomed Anal* **21**: 407–413.
Jiang SY, Ye CG. 2006. Experienced identification of adulterated slices of some common traditional Chinese medicines. *Li Shi Zhen Med Mater Med Res* **17**: 1509.
Lin LZ, He XG, Lindenmaier M, Nolan G, Yang J, Cleary M, Qiu SX, Cordell GA. 2000. Liquid chromatography–electrospray ionization mass spectrometry study of the flavonoids of the roots of *Astragalus mongholicus* and *A. membranaceus*. *J Chromatogr A* **876**: 87–95.
Ma XQ, Shi Q, Duan JA, Dong TTX, Tsim KWK. 2002. Chemical analysis of Radix Astragali (Huangqi) in China: a comparison with its adulterants and seasonal variations. *J Agric Food Chem* **50**: 4861–4866.
Mahalanobis PC. 1936. On the generalised distance in statistics. *Proc Natl Inst Sci India* **2**: 49–55.
Morey LC, Blashfield RK, Skinner HA. 1983. A comparison of cluster analysis techniques within a sequential validation framework. *Multivariate Behav Res* **18**: 309–329.
Normile D. 2003. The new face of traditional Chinese medicine. *Science* **299**: 188–190.
Rencher AC. 2002. *Methods of Multivariate Analysis* (2nd Edn.). John Wiley & Sons: New York; 397.
Settle FA. 1997. *Handbook of Instrumental Techniques for Analytical Chemistry*. Prentice-Hall: Englewood Cliffs, NJ.
Song JZ, Yiu HHW, Qiao CF, Han QB, Xu HX. 2008. Chemical comparison and classification of Radix Astragali by determination of isoflavonoids and astragalosides. *J Pharm Biom Anal* **47**: 399–406.
Wang YH, Yuan JR, 2004. Application of infrared spectroscopy in materia medica identification. *Acta Univ Trad Med Sin Pharmacol Shanghai* **18**: 62–65.
Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* **58**: 109–130.
Xue TH, Roy R. 2003. Studying traditional Chinese medicine. *Science* **300**: 740–741.
Yip PY, Kwan HS. 2006. Molecular identification of *Astragalus membranaceus* at the species and locality levels. *J Ethnopharmacol* **106**: 222–229.