ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
21 (1999) 407–413

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Discrimination of herbal medicines according to geographical origin with near infrared reflectance spectroscopy and pattern recognition techniques

Young-Ah Woo [a], Hyo-Jin Kim [a,*], JungHwan Cho [b], Hoeil Chung [c]

[a] *College of Pharmacy, Dongduk Women's University, Seoul 136-714, South Korea*
[b] *College of Pharmacy, Sookmyung Women's University, Seoul 140-742, South Korea*
[c] *SK Corp. Production Technology Center, Ulsan 680-130, South Korea*

Received 30 October 1998; received in revised form 20 April 1999; accepted 7 June 1999

Abstract

Herbal medicines have an important role in clinical therapy in Asian countries such as Korea, Japan, and China. The objective of this study is to develop a nondestructive and accurate analytical method to discriminate herbal medicines according to geographical origin. Even though they are the same species, their qualities are different by growing conditions such as climate and soil. Near infrared (NIR) reflectance spectroscopy and a pattern recognition technique were applied for discrimination of herbal medicines according to geographical origin (Korea and China). Astragali Radix (AR), Ganoderma, and Smilacis Rhizoma (SR) were examined. It is shown that the representative NIR reflectance spectra in each group are different according to geographical origin after second derivatization to enhance spectral features. Also, the NIR reflectance spectra of Chinese and Korean samples were differentiated using principal component (PC) score plots. To establish the discrimination rule, Mahalanobis distance and discriminant analysis with PLS2 were utilized.

Keywords: Near infrared reflectance spectroscopy; Pattern recognition; Discrimination; Herbal medicines; Geographical origin

1. Introduction

Historically, herbal medicines have played an important role in clinical therapy in Asian areas. Recently, the pharmaceutical value of Korean medicines like ginseng has been increased owing to the ability to cure disease without considerable side effects. Herbal medicines are widely used and cultivated in Asian areas such as Korea, China, and Japan. Even though herbal medicines come from the same species, the quality and efficacy are somewhat different according to growing conditions based on geographical origin. Therefore, a rapid and accurate analytical method to determine the origin is essentially required for the correct value estimation and for the prevention of illegal distribution. However, it is not easy to determine the geographical origin with existing

* Corresponding author.

0731-7085/99/$ - see front matter
PII: S0731-7085(99)00145-4

analytical tools as well as through visible inspection. Since there are tens of major components, which are slightly different according to growing conditions such as geographical origin, we can not select only several specific components as essential criteria. Near infrared (NIR) reflectance spectroscopy can be an excellent candidate for the determination of geographical origin of herbal medicines because it is fast, accurate, and nondestructive. NIR pattern recognition method was applied for the discrimination of roasted coffees [1] and vegetable oils [2]. Those researches were based on the classification of chemically different samples. In present study, we applied pattern recognition techniques to classify the same species of different geographical origin of herbal medicines.

Astragali Radix (AR), Ganoderma, and Smilacis Rhizoma (SR) were chosen since those are very popular herbal medicines. Astragali Radix, a root of *Astragalus membranaceus* is clinically used in Korea and China to improve a reduced immune response, often for the elderly [3]. Ganoderma (*Ganoderma lucidum*), an oriental fungus, has been widely used as a remedy for promotion of health and longevity in Korea, China and other Asian countries [4]. Smilacis Rhizoma, a rhizome of *Smilax glabra* are used in the Orient as a traditional medicine for chronic skin disease and syphilis [5].

We first applied principal component analysis (PCA) to ascertain the possibility of discrimination with NIR reflectance spectroscopy. Also, two pattern recognition techniques, Mahalanobis distance method coupled with Hotelling's T^2 statistics [6] and discriminant PLS2 [7] were applied to develop discrimination techniques.

2. Experimental

2.1. Samples preparation

All herbal medicines were acquired from The Experimental Station of the Natural Agriculture Products Inspection Office (NAPIO), Seoul, Korea. Korean samples used in this study were collected from various cultivation areas. Korean Samples for Bokryung Radix and Austragali Radix that were mainly cultivated in the area of Kangwon and Chungbuk in Korea were used. Korean samples for Ganoderma that were cultivated in the area of Chungbuk in Korea were used. Chinese samples for all three herbal medicine that were cultivated in the north-east of China were used. Fifty-one Korean and 46 Chinese AR, 27 Korean and 16 Chinese Ganoderma, 67 Korean and 23 Chinese SR samples were studied. Dried samples were acquired. All samples were powdered using a cyclone mill (Udymill, USA) fitted with 1 mm screen. The particle size of powder was below 20 meshes.

2.2. Near infrared reflectance spectra

NIR reflectance spectra were collected over the 1100–2500 nm spectral region with a NIRSystems model 6500 spectrometer (Foss NIRSystems, MD) equipped with a quartz halogen lamp and PbS detector. The spectra were collected with 2 nm data intervals. The spectra were acquired with a circular sample cup with a quartz window (38 mm in diameter and 10 mm in thickness). Each sample spectrum was obtained by averaging 32 scans. All of the spectra were recorded as log $(1/R)$ with respect to a ceramic reference standard. Sample spectra were divided into training and test set using random selection as shown in Table 1.

2.3. Data analysis

PCA and discriminant analysis with PLS2 were performed using WINISI software (Foss NIRSystems, MD). Mahalanobis distance was performed using MATLAB® (The MathWorks Inc.). Second derivatives of all the raw spectra were calculated before the data processing.

3. Results and discussion

3.1. Near infrared reflectance spectra

By examination of raw reflectance spectra of AR, Ganoderma, and SR, no significant spectral

Table 1
Data set preparation

Sample	Number of spectra in training set		Number of spectra in test set	
	Korea	China	Korea	China
AR	35	32	16	14
Ganoderma	19	11	8	5
SR	46	16	21	7

differences were observed based on the origin. However, when spectra were averaged and second derivatized, reasonable spectral differences were observed between samples from Korea and China. Fig. 1 shows the averaged second derivative spectra of AR from Korea and China. The most significant differences are observed around the 1700 nm band which corresponds to the first overtone of CH stretch [8]. It is hard to find the clear rationale for spectral differences; however, these are clearly resulted from the compositional or differences of Korean and Chinese AR. Similarly for Ganoderma and SR, considerable spectral differences are observed in the 1300–1400 nm range. The corresponding spectra are shown Figs. 2 and 3, respectively. Overall, AR, Ganoderma, and SR are very complex and resulting NIR reflectance spectra are highly overlapped; however, unique spectral differences between Korean and Chinese samples can provide qualitative information.

3.2. Principal component analysis

Principal component analysis (PCA) has been performed to examine the qualitative difference between Korean and Chinese herbal medicines used in this study. Before performing PCA, all of the sample spectra were preprocessed using a second derivatization algorithm to reduce baseline variations and enhance the spectral features. The full NIR reflectance spectral range (1100–2500 nm) was used instead of using specific spectral regions. The spectral variations based on the origin of herbal medicines are expected to present throughout whole NIR range because herbal medicines are composed of huge numbers of chemically different individual components.

All of the spectra of 51 Korean and 46 Chinese AR, 27 Korean and 16 Chinese Ganoderma, and 67 Korean and 23 Chinese SR samples were used for PCA. Figs. 4–6 show the three dimensional (3D) score plots using first, second, and third factors. The initial three factors, which describe the most spectral variations related to origin, are

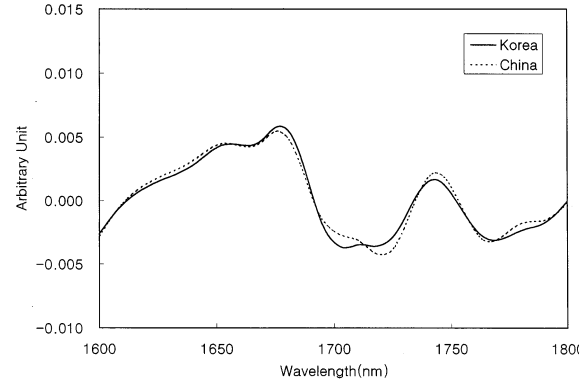

Fig. 1. The average second derivative spectra of Astragali Radix from Korea and China.

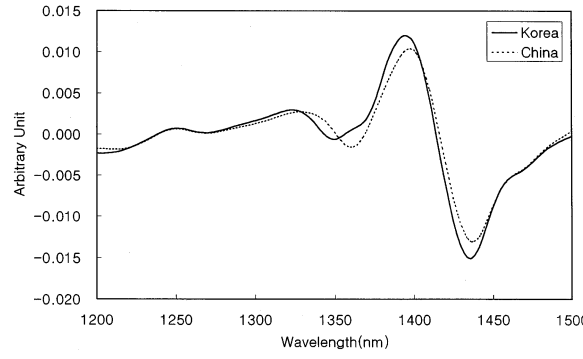

Fig. 2. The average second derivative spectra of Ganoderma from Korea and China.

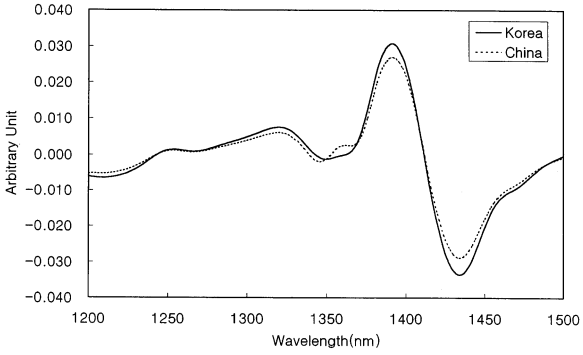

Fig. 3. The average second derivative spectra of Smilacis Rhizoma from Korea and China.

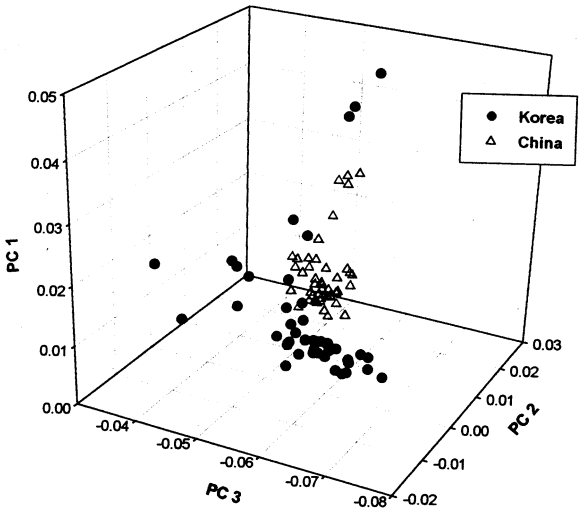

Fig. 4. The three-factor plot in PCA for Astragali Radix.

used to make differentiation clearer. For AR, the scores are divided moderately into two groups (Korea and China), however some points overlap each other. On the other hand, in the case of Ganoderma, the samples from both countries are clearly differentiated without overlapping. This result shows that PCA can adequately discriminate Ganoderma samples from different geographical origins. For SR, all samples are generally classified into two groups; however, there are considerable overlaps between the two origins. It can be assumed that some Chinese SR samples are similar to Korean ones in terms of composition.

By examining PCA 3D plots, it is expected that discriminating AR and SR samples from two different geographical origins is feasible; however, it is not perfect. The result of Ganoderma shows the clear classification based on two different origins. Overall, these results sufficiently show that discrimination of herbal medicines according to geographical origin using a principal component is possible. Therefore, for actual discrimination, two different methods of Mahalanobis distance and discriminant PLS2 were utilized.

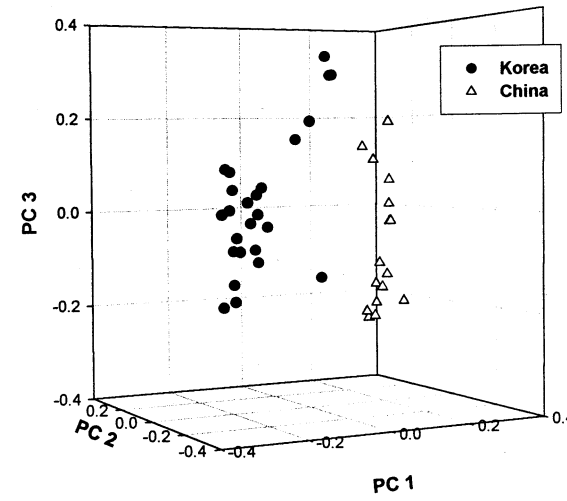

Fig. 5. The three-factor plot in PCA for Ganoderma.

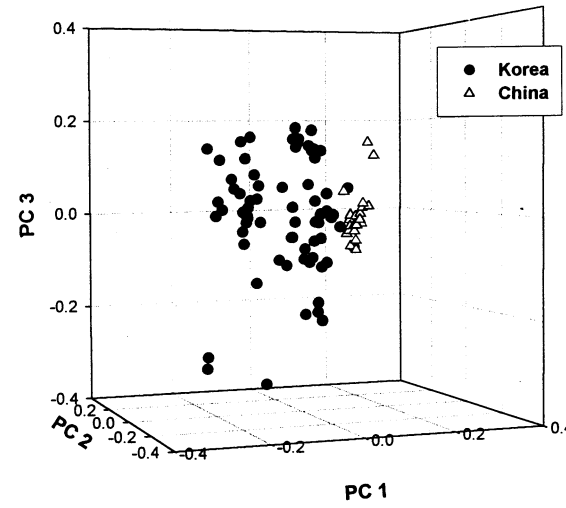

Fig. 6. The three-factor plot in PCA for Smilacis Rhizoma.

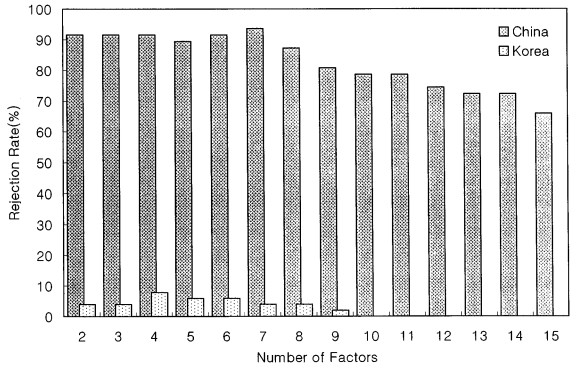

Fig. 7. Rejection rate of Chinese samples of Astragali Radix with Mahalanobis distance.

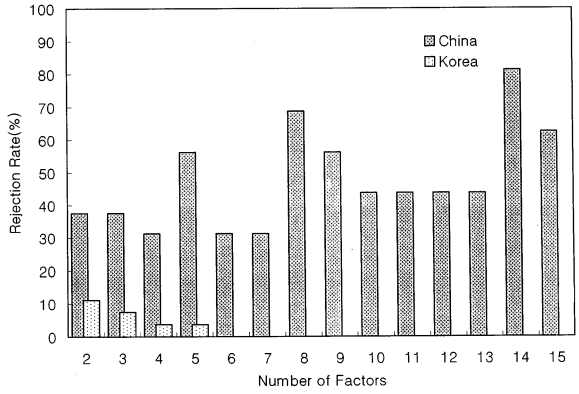

Fig. 8. Rejection rate of Chinese samples of Ganoderma with Mahalanobis distance.

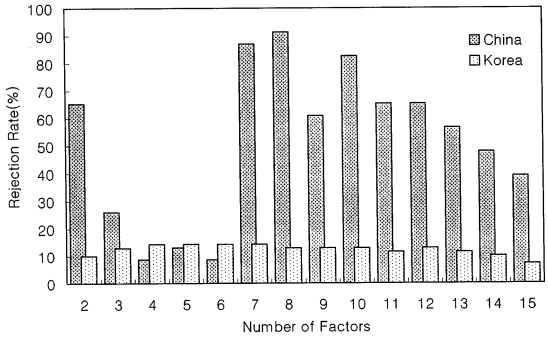

Fig. 9. Rejection rate of Chinese samples of Smilacis Rhizoma with Mahalanobis distance.

3.3. Mahalanobis distance

The probability based on Hotelling's T^2 statistics was calculated from Mahalanobis distance and the threshold value is set to 0.05% to determine rejects. The spectral range from 1300 to 2300 nm is used. To develop an identification method for geographical origin of samples with a Mahalanobis distance value, Korean samples were selected as reference and Mahalanobis distance as well as probability were calculated. Standard sample set was previously developed as Korean samples. After constructing a qualitative model, we can classify samples as acceptable and reject samples with a significance level of 0.05%. Thus, it is expected that Chinese samples are identified as rejects. When being plotted as functions of the number of factors from 2 to 15, as Figs. 7–9 show, all Chinese samples were not classified as reject. The highest rejection rates for AR, Ganoderma, and SR are 93.6, 81.3 and 91.3%, respectively. It is noteworthy that the results of Ganoderma are not good, though the classification Ganoderma using PCA is good in Fig. 5. This feature proves that it is critical to select proper pattern recognition method for each sample. However, this result suggested the possibility of the application of NIR reflectance spectroscopy and Mahalanobis distance.

3.4. Discriminant PLS2

Alternatively, the discriminant analysis based on PLS2 has been utilized. In this study, there are only two different classes from different geographical origins. Therefore, it is possible to build calibration models by spectral data as well as providing a priori knowledge of the class membership of each sample in a training set. By this way, it is expected to build the better calibration models using spectral data and information of origin. For this purpose, the discriminant PLS2 method was used. PLS2 utilizes not only the spectral data itself but also the external knowledge of the origin of samples based on different geographical origin. Discriminant PLS2 places one group as 1 and the other group as 2; 2 meaning the sample is in the desired group, and 1 meaning the sample is not in the group. The general concept is that the pre-

dicted value of 2.0 is the perfect match, 1.0 is not in the group, and 1.5 can go either way.

To perform discriminant PLS2, the training and a test set described in Table 1 were used. The

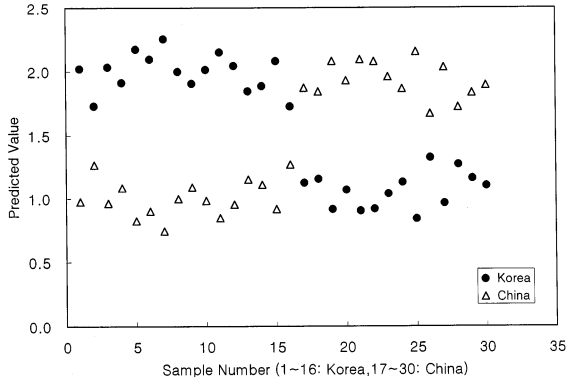

Fig. 10. The PLS2 prediction result of Astragali Radix.

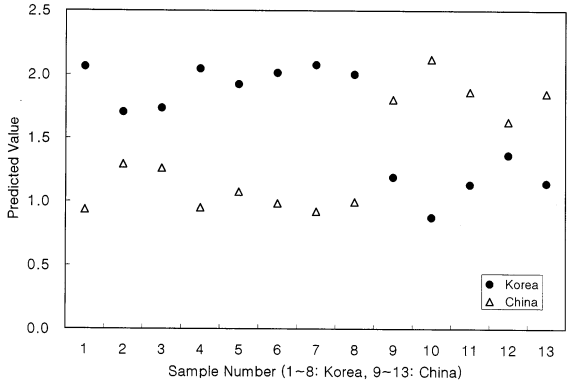

Fig. 11. The PLS2 prediction result of Ganoderma.

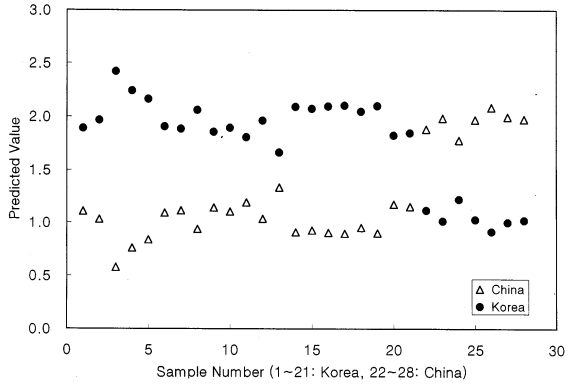

Fig. 12. The PLS2 prediction result of Smilacis Rhizoma.

training set is used to build calibration models, and test set to validate the performance of calibration models. Additionally, it is very important to collect the representative samples from many different geographical origins in one country for the data set. For this purpose, Korean samples were collected from various growing places; therefore, the training set contains wide variations, such as growing condition and composition, related to the origin of herbal medicine. Before performing PLS, all the spectra were also preprocessed by taking second derivatives. Figs. 10–12 show the prediction results for all three herbal medicines. All samples of test set are accurately discriminated according to geographical origin. As shown in Fig. 5, Ganoderma samples are clearly differentiated in 3D score plot using PCA; therefore, predicted result with PLS2 algorithm shows the very clear classification. The prediction results of AR and SR show relatively worse classification results, however each sample was clearly identified based on its geographical origin and acceptable.

These results definitely demonstrate the applicability of NIR reflectance spectroscopy combined with the pattern recognition technique to determine the geographical origin of herbal medicine such as AR, Ganoderma and SR, for the first time. By using discriminant PLS2, which simultaneously utilizing the spectral data and knowledge of origin, the discrimination capability has been significantly improved compared to the Mahalanobis distance. The herbal medicines studied here are only a few examples out of huge numbers of its categories. It can be expected that this method be applied further for the same purpose to other herbal medicines in the future.

4. Conclusions

The results described in this research open the possibility of discriminating herbal medicines according to their geographic origin using NIR reflectance spectroscopy combined with pattern recognition methods, such as Mahalanobis distance and discriminant PLS2. This study shows that, if the origins are confined to two countries,

discriminant PLS2 will be appropriate for future applications. Without NIR reflectance spectroscopy, the rapid identification of origins of herbal medicine, such as AR, Ganoderma, and SR, was impossible or only limited to human eye examination that is lack of objectiveness. NIR reflectance spectroscopy requires minimum steps of pretreatment of grinding and sieving to control the particles size in this case. Practically, direct NIR scanning of herbal medicines without grinding, which makes this measurement much more convenient, is easy and possible. Future work will be directed to examine: (1) the performance of calibration models for more AR, Ganoderma, and SR samples; (2) feasibility of NIR reflectance spectroscopy for the direct measurement without sampling grinding; (3) applicability of scatter correction algorithms; and (4) feasibility of NIR reflectance spectroscopy for other herbal medicines, i.e. ginseng.

Acknowledgements

The authors would like to thank Foss NIRSystems and DooRee Tech for technical support.

References

[1] M.J. Martin, F. Pablos, A.G. Gonzalez, Anal. Chim. Acta 320 (1996) 191.
[2] K.M. Bewig, A.D. Clarke, C. Roberts, N. Unklesbay, JAOCS 71 (1994) 195.
[3] K. Kajimura, Y. Takagi, K. Miyano, Y. Sawabe, M. Mimura, Y. Sakagami, H. Yokoyama, K. Yeneda, Biol. Pharm. Bull. 20 (1993) 1178.
[4] S.Y. Wang, M.L. Hsu, H.C. Hsu, C.H. Tzeng, S.S. Lee, M.S. Shiao, C.K. Ho, Intl. J. Cancer 70 (1997) 699.
[5] T. Namba, Coloured Illustrations of Wakan-Yaku, Hoikushya Publishing Co., 1980.
[6] R. Vong, A. Geraldi, S. Wold, K. Esbensen, J. Chemom. 2 (1988) 281.
[7] H.L. Mark, D. Tunnell, Anal Chem. 57 (1985) 1449.
[8] O.H. Wheeler, Chem. Rev. 59 (1959) 629.