Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 171 (2017) 351–360

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Rapid measurement of epimedin A, epimedin B, epimedin C, icariin, and moisture in Herba Epimedii using near infrared spectroscopy

Yue Yang [a], Xuesong Liu [a], Weili Li [b], Ye Jin [a], Yongjiang Wu [a], Jiyu Zheng [b], Wentao Zhang [b], Yong Chen [a,*]

[a] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China
[b] SPH Liaoning Herbapex Pharmaceutical (Group) Co., Ltd., Benxi 117200, China

ARTICLE INFO

Article history:
Received 1 March 2016
Received in revised form 8 August 2016
Accepted 17 August 2016
Available online 18 August 2016

Keywords:
Near infrared spectroscopy
Rapid measurement
Genetic algorithm (GA)
Support vector machine (SVM) regression
Herba Epimedii

ABSTRACT

In this work, near infrared (NIR) spectroscopy was used in combination with chemometrics to determine the epimedin A, epimedin B, epimedin C, icariin, and moisture contents of Herba Epimedii. The variable selection method genetic algorithm (GA) and regression tool support vector machine (SVM) were used to improve the model performance. Four different calibration models, namely Full-PLS, GA-PLS, Full-SVM, and GA-SVM, were established, and their performances in terms of prediction accuracy and model robustness were systemically studied and compared. In conclusion, the performances of the models based on the efficient variables selected through GA were better than those based on full spectra, and the nonlinear models were superior over the linear models. In addition, the GA-SVM model demonstrated the optimal performance in predicting five quality parameters (*viz.* epimedin A, epimedin B, epimedin C, icariin, and moisture). For GA-SVM, the determination coefficient (R_p^2), root-mean-square error (*RMSEP*), and residual predictive deviation (*RPD*) for the prediction set were 0.9015, 0.0268%, and 2.20 for epimedin A; 0.9089, 0.0656%, and 3.08 for epimedin B; 0.9056, 0.1787%, and 3.18 for epimedin C; 0.8192, 0.0657%, and 2.26 for icariin; and 0.9367, 0.2062%, and 4.12 for moisture, correspondingly. Results indicated that NIR spectroscopy coupled with GA-SVM calibration can be used as a reliable alternative strategy to measure the epimedin A, epimedin B, epimedin C, icariin, and moisture contents of Herba Epimedii because this technique is fast, economic, and nondestructive compared with traditional chemical methods.

1. Introduction

Herba Epimedii, which is obtained from the dry and aerial counterparts of many *Epimedium* species (Berberidaceae), is a popular traditional Chinese medicine (TCM) herb with a long history in China. This TCM is efficient in tonifying the kidney and dispelling wind chill; Herba Epimedii is also widely used to treat important symptoms, such as seminal emission, weakness of the limbs, and rheumatism [1]. Modern pharmacological experiments suggest that Herba Epimedii offers a promising therapeutic value for the treatment of coronary heart diseases, hypertension, diabetes, and osteoporosis [2–4].

Herba Epimedii contains high amounts of flavonoids which are fundamental pharmacological active constituents of plants. Several flavonoids, such as epimedin A, epimedin B, epimedin C, and icariin, were isolated from Herba Epimedii, and their therapeutic effects have been reported [5,6]. Therefore, as an important indicator of the quality of Herba Epimedii, the herb's flavonoid content attracted much attention in the past decade. Various instrumental analytical techniques have been applied to measure the flavonoid content of Herba Epimedii; these strategies include high performance liquid chromatography (HPLC) [7], capillary zone electrophoresis (CZE) [8], and micellar electrokinetic chromatography (MEKC) [9]. However, these traditional methods are all destructive, time consuming, laborious, and costly. Therefore, a rapid and nondestructive analytical method should be developed to analyze the flavonoid content of Herba Epimedii.

Near infrared (NIR) spectroscopy is becoming a powerful tool for analyzing a wide variety of samples, such as petroleum, pharmaceutical products, environmental components, and food products, because of this method's speed, economy, and precision compared with other analytical techniques [10]. NIR spectroscopy has also shown great potential and gained wide acceptance in TCM [11]. This method has been extensively embraced in the manufacturing process of TCM, including qualitative discrimination and quantitative determination of herb materials, intermediates and products [12–15], and separation monitoring [16–19].

Multivariate calibration models should be established for the qualitative and quantitative analyses of NIR spectroscopy, which can be achieved through the combination of rich-information spectroscopy and efficient regression tools provided by modern mathematics. However, most studies for quantitative determination of Herba Epimedii by using NIR mainly focused on the classical linear regression models

* Corresponding author.
E-mail address: chenyongzju@126.com (Y. Chen).

http://dx.doi.org/10.1016/j.saa.2016.08.033
1386-1425/

(PLS) based on a full spectrum [20,21]; little attempt has been made to develop nonlinear regression models based on selected spectral regions. For the full-range spectra, the existence of insignificant and irrelevant information can weaken the accuracy and robustness of the models. Therefore, spectral variable selection is essential to improve the performance of the final calibration models. In the present study, a promising procedure known as genetic algorithm (GA) was used to eliminate the uninformative variables and/or conduct wavelength selection to build a high-performance calibration model [22]. In addition, considering the complex correlation between the NIR spectra and the determined parameters, the linear regression tools might not be able to provide a complete solution to the modeling problem. Therefore, attempting to develop robust models by using nonlinear algorithm is very significant. A support vector machine (SVM), which performs excellently compared with other conventional learning algorithms, was applied in this study [23,24].

This study aimed to (1) investigate the feasibility of determining the epimedin A, epimedin B, epimedin C, icariin, and moisture contents of Herba Epimedii via NIR spectroscopy; (2) compare the performances of the models based on the full-range spectra or the efficient spectral variables; and (3) compare the performances of linear and nonlinear regression models.

2. Materials and methods

2.1. Plant materials and reagents

Herba Epimedii materials were supplied by a pharmaceutical factory (SPH Liaoning Herbapex Pharmaceutical Group, Co., Ltd., Liaoning, China). Medicinal grade ethanol (70%, v/v) was used as extraction solvent. HPLC-grade acetonitrile was purchased from Merck (Darmstadt, Germany). Epimedin A, epimedin B, epimedin C, and icariin were purchased from Chengdu Must Bio-Technology Co., Ltd.

2.2. Sample preparation

The dried aerial parts of the samples were pulverized and passed through an 80-mesh sieve. The powdered samples were kept in ziplock bags and stored in a dark place prior to analysis. Each sample was weighed accurately (0.2 g) and soaked with 25 ml of 70% ethanol solution. The mixture was extracted by ultrasonication for 30 min at 40 °C. After cooling, the extracts were centrifuged at 3800 rpm for 10 min, and the supernatant was used for HPLC analysis. A total of 117 Herba Epimedii samples were obtained for analysis.

2.3. Reference assays

Epimedin A, epimedin B, epimedin C, and icariin were quantitatively analyzed on an Agilent 1200 HPLC system (Agilent Technologies, USA), which consisted of a vacuum degasser, a quaternary pump, an autosampler, a thermostatic column compartment, and a UV detector. A Waters Xbridge C18 HPLC column (particle size: 5 μm, diameter: 4.6 mm, length: 250 mm) was utilized, and the mobile phase consisted of (A) 0.1% phosphoric acid and (B) acetonitrile; the isocratic elution was 26% B for 20 min. The flow rate was set to1.0 ml min^{-1}, and the injection volume was 10 μl. The column temperature was retained at 25 °C, and UV detection was accomplished at 270 nm.

The moisture contents were determined using the loss-on-drying method. Samples were heated to constant weight at 105 °C, and the results were expressed in wt%.

2.4. NIR spectra acquisition

NIR spectra acquisition was performed in reflectance mode by using a Bruker Matrix-F Fourier transform NIR spectrometer (Bruker Optics Inc., Germany) with an integrating sphere. Each spectrum was collected in the region of 12,000–4000 cm^{-1} with are solution of 8.0 cm^{-1} and average of 32 scans. The spectral data were measured at intervals of 7.714 cm^{-1}, which resulted in 1037 variables. The NIR spectra were recorded as absorbance by using air as the reference standard.

2.5. Multivariate data analysis

Calibration models were developed using PLS and SVM regression algorithms by relating NIR spectra to the reference measurements.

The SVM method, which was proposed by Vapnik [25], was successfully applied to NIR spectroscopy nonlinear prediction model because this method demonstrates superior generalization and accurate prediction capabilities, as well as can avoid over-fitting problems [26–29].

The basic concept underlying SVM is the mapping of original data points to a higher dimensional feature space and construct an optimal separating plane from which the distance to all the data points is minimum. Mapping data into a higher dimensional space is often implemented through a kernel function [30]. Generally, four possible choices exist for the kernel function of SVM regression: linear, polynomial, sigmoid, and radial basis function (RBF). In the present study, RBF was selected as the kernel function of SVM regression because the prediction effect of the SVM model based on the RBF function is better, and the theoretical systems of RBF are more developed than other kernel functions [31]. Optimization of the regularization parameter C and RBF kernel parameter g is the key step in SVM as their combined values determine the regression performance. The regularization parameter C controls the trade-off between minimizing the training error and minimizing the model complexity. The RBF kernel parameter g defines the width of the kernel and significantly affects the predictive ability of SVM. To perform this optimization, an exhaustive grid search technique [32] was used, and both parameters were optimized within the region of 2^{-20} to 2^{20}. For each combination of C and g parameters, the mean square error of cross validation (*MSECV*) was calculated, and the combination with the lowest *MSECV* value was selected as the optimal one.

2.6. Variable selection

As a well-known variable selection method [33], genetic algorithm (GA) was used to select the most appropriate NIR spectral wavelengths to reduce the spectral complexity and improve the calibration model.

This algorithm is based on the principles of genetics and natural selection, of which the calculations were performed using the following steps: (i) coding each “chromosome” and creating an initial population, (ii) evaluating the performance of each “chromosome” by using the fitness value as the criterion for guiding the GA to the global optimum, (iii) selecting the better performing “chromosomes” from the population, (iv) processing crossover and the random mutation of “chromosomes,” and (*v*) replacing the parent “chromosomes” by reinserting the offspring in the next population in terms of their performance. Steps 2 to 4 were repeated until the termination criterion was achieved.

2.7. Evaluation of model performance

The performances of the established models were evaluated using the following standards: root-mean-square error of calibration (*RMSEC*), root-mean-square error of prediction (*RMSEP*), standard error of calibration (*SEC*), standard error of prediction (*SEP*), residual predictive deviation (*RPD*), determination coefficients of calibration (R_c^2), and determination coefficients of prediction (R_p^2). *RMSEC* and *RMSEP* were recorded to determine the models' efficiency. Moreover, *RPD*, which is defined as the ratio of standard deviation (*SD*) of the reference data to the *SEP*, was also calculated to assess how well the calibration model can predict compositional data. Other statistics, such as slope and bias, are considered to discern systematic errors and investigate the correlation between the reference and the NIR models [34].