Chemometrics and Intelligent Laboratory Systems 156 (2016) 181–187

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Application of fluorescence spectroscopy combined with interval partial least squares to the determination of enantiomeric composition of tryptophan

Long Jiao [a,*], Shan Bing [a], Xiaofeng Zhang [a], Yunxia Wang [b], Hua Li [a,b]

[a] College of Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an 710065, PR China
[b] College of Chemistry and Materials Science, Northwest University, Xi'an 710069, PR China

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form 6 June 2016
Accepted 9 June 2016
Available online 11 June 2016

Keywords:
Fluorescence spectroscopy
Interval partial least squares
Partial least squares
Enantiomeric composition
Tryptophan

ABSTRACT

The application of interval partial least squares (*i*PLS) method to the fluorescence spectroscopy analysis was investigated. The practicability of spectral region selection in the fluorescence spectroscopy analysis was studied. A method which combines fluorescence spectroscopy and *i*PLS was proposed for determining the enantiomeric composition of tryptophan (Trp). Fluorescence spectroscopy was used to measure the diastereomeric interaction between the enantiomers of Trp and bovine serum albumin, which plays the role of chiral selector. *i*PLS was used to select the spectral region and build the calibration model between the fluorescence spectral data and enantiomeric composition of Trp. The spectral region from 300 nm to 356 nm was selected and used to build the calibration model. Leave-one-out cross validation and external test validation were used to validate the obtained models. By selecting the spectral region, *i*PLS provided better prediction performance, in comparison to the full-spectrum PLS model. The root mean square relative error (*RMSRE*) of leave-one-out validation has decreased from 10.24 to 7.17, and the *RMSRE* of external test validation has decreased from 8.59 to 7.06. In addition, four local-spectrum PLS models were developed in order to make a comparison to the *i*PLS model. The result demonstrates the *i*PLS model is superior to the four local-spectrum PLS models. The result of this work demonstrates the proposed method is practicable for determining the enantiomeric composition of Trp at trace level. When there is 2.50 μmol·L^{-1} Trp in the samples, the enantiomeric composition can be accurately determined. Moreover, the result demonstrates the selection of spectral region has significant influence on the fluorescence spectroscopy analysis and *i*PLS is a practicable method for spectral region selection.

1. Introduction

Chirality is a crucial factor in numbers of chemical and biological reactions. With the rapid development of chiral chemistry, enantiomeric pure products are needed in chemical and pharmaceutical industries. As a result, the determination of enantiomeric composition for chiral compounds has gained much attention [1,2]. The determination of enantiomeric composition for chiral compounds is always a hard work because the enantiomers of a chiral compound have very similar, often identical, physical and chemical properties. Chromatography is commonly used to determine the enantiomeric purity of chiral compounds. However, time-consuming chromatographic methods are not attractive, and fast determining techniques seem much more promising for high throughput analytical strategies. It is still of great importance to develop fast and low-cost methods for determining the enantiomeric composition of chiral compounds [3–10]. As a kind of rapid, simple, low-cost and reliable methods, spectroscopic methods including UV–vis [3–8], fluorescence [9–11] and near-infrared (NIR) [12,13] have already gained much attention and been applied to the determination of enantiomeric composition of chiral compounds. Among these methods, fluorescence spectroscopy is regarded as a very important and promising method for determining enantiomeric composition of chiral compounds at trace level.

In these researches, it has been demonstrated there is a quantitative relationship between the enantiomeric composition of chiral compounds and their spectra. Thus, PLS was used to model the quantitative relationship between the spectral data and enantiomeric composition. However, PLS is a global calibration method, which builds calibration models by using the data from full spectral range [14]. In spectroscopic analysis, some spectral regions may contain chemical information due to other analytes, non-modeled interferences, background variations and interactions. These spectral regions will degrade the calibration model and should not be used to build the model [15,16]. Hence, application of the multivariable calibration methods to multi-component spectroscopic analysis usually requires spectral region selection for building well-

* Corresponding author.
E-mail address: mop@xsyu.edu.cn (L. Jiao).

http://dx.doi.org/10.1016/j.chemolab.2016.06.005
0169-7439/

fitted model and avoiding non-modeled interferences [17–19]. Several researches [14,20–23] have demonstrated that spectral region selection can lead to improvements of prediction ability over standard full-spectrum PLS model in IR (includes mid-infrared and near-infrared) and Raman analysis. Similar to IR and Raman spectra, fluorescence spectra may contain some regions which contain non-modeled information and should be excluded from the calibration model. Spectral region selection is likely to benefit the fluorescence spectroscopy analysis. However, the influence of spectral region selection on the fluorescence spectroscopy analysis has not been studied yet. Interval partial least squares (*i*PLS) is a local multivariable calibration method which can detect the spectral region that gives the most important information to build the calibration model. It has been successfully applied to IR and Raman analysis [14,16, 20–30]. Therefore, the application of *i*PLS to the fluorescence spectroscopy analysis was investigated in this work. Fluorescence spectroscopy combined with *i*PLS was applied to determining the enantiomeric composition of Trp. Fluorescence spectroscopy was used to measure the diastereomeric interaction between the enantiomers of Trp and bovine serum albumin (BSA), which plays the role of chiral selector. *i*PLS was used to select the spectral region and build the calibration model.

2. Experimental and methods

2.1. Reagents

Enantiomeric pure L-Trp and D-Trp were purchased from Sigma–Aldrich (U. S.) and used as received. BSA (fraction V) was purchased from Sino-American Biotechnology Co. (Shanghai, China). Milli-pore water was produced by a Milli-Q synthesis A10 system (Merck Group, Germany) and used for preparing sample solutions. All sample solutions contain a fixed concentration of BSA (5.00 μmol·L^{-1}) and a fixed concentration of Trp (2.50 μmol·L^{-1}) at pH 8.00, while the enantiomeric compositions of Trp are varied.

2.2. Spectra

The fluorescence spectra were collected and recorded with a Cary Eclipse fluorescence spectrophotometer (Varian, U. S.). A fluorometer cell with a 1.0 cm path length was used. Sample excitation was done at 280 nm, and the emission was scanned between 300 nm and 850 nm. The spectra were dark off and blank corrected.

2.3. Data and software

RMSRE was used to assess the prediction accuracy of the obtained model. It is defined as Eq. (1):

$$RMSRE = \sqrt{\frac{\sum (RE_i)^2}{n}} \tag{1}$$

where RE_i is the relative error of the *i*th sample, and *n* is the total number of the used samples.

The calculation of PLS was carried out with PLS_toolbox 5.5 (Eigenvector research incorporated, U.S.). The calculation of *i*PLS was carried out with iToolbox (programmed by Prof. L. Nørgaard, KVL, Denmark, published on http://www.models.kvl.dk/iToolbox). The independent variable matrix (**X** matrix) of PLS and *i*PLS was constructed by assembling the spectral data as row vectors. The mol fraction of L-Trp was used as the dependent variable (**Y** matrix) to develop the calibration model. The **X** and **Y** matrix were mean-centered prior to PLS and *i*PLS.

2.4. Variable importance in projection

Variable importance in projection (VIP) scores is the calculated variable importance in projection from a regression model [31–35]. It is commonly used to select variable in PLS regression. The idea behind the VIP score is to accumulate the importance of each variable *j* being reflected by *w* from each component. The VIP score is namely a weighted sum of squares of the PLS weights, which takes into account the explained variance of each PLS dimension. The VIP score V_j for each variable *j* in a regression model is defined as Eq. (2):

$$V_j = \sqrt{p\sum_{a=1}^{A}\left[SS_a\left(w_{aj}/\|w_a\|^2\right)\right]\Big/\sum_{a=1}^{A}(SS_a)} \tag{2}$$

where SS_a denotes the sum of squares explained by the *a*th component. Hence, the V_j weights is a measure of the contribution of each variable according to the variance explained by each PLS component where $(w_{aj}/\|w_a\|)^2$ represents the importance of the *j*th variable.

The variables are selected according to the VIP scores. Variable *j* can be excluded if V_j is lower than a user-defined threshold. It is generally accepted that a variable with V_j greater than 1 can be considered important in regression model. Variables with V_j less than 1 are less important and might be good candidates for exclusion from the model. The "greater than one" rule is generally used as a criterion for variable selection because the average of squared VIP scores is equal to 1.

2.5. Interval partial least squares

Interval partial least squares is a graphically oriented local modeling method. It has been commonly used in IR and Raman spectroscopy analysis. Because its algorithm has been elaborated in the article [14], only a brief outline of *i*PLS is presented here for the sake of brevity. When *i*PLS is used to analyze spectral data, its main force is to provide an overall view of the relevant information in different spectral regions, thereby finding important spectral regions and removing interferences from other regions. This method is able to illustrate local calibration models in a graphical display, focusing on a choice of better intervals. In *i*PLS method, the input data are divided into a number of equidistant intervals (regions), and the local PLS model is developed for each interval. The prediction performance of the developed local PLS models and the global model (developed from full spectrum) is calculated and compared. The comparison is usually based on the root mean square error of cross validation (RMSECV). The interval with the smallest RMSECV is chosen and regard as the optimal one. The local PLS model of this interval is recorded and regarded as the final *i*PLS model.

2.6. Leave one out cross validation

Leave-one-out cross-validation [36–39] is a commonly used algorithm for estimating the predictive performance of a multivariable calibration model. Usually, practical calibration experiments have to be based on a limited set of available samples. The idea behind the leave-one-out cross validation algorithm is to predict the property value of each sample in turn with the calibration model which is developed from the other samples. When applying the algorithm to a dataset including *n* samples, the calibration modeling is performed *n* times, each time using (*n*-1) samples for modeling and one sample for testing. Hence, the procedure of leave-one-out cross validation can be divided into *n* segments. In each segment *i* ($i = 1,\ldots, n$), there are three steps: (1) taking sample *i* out as temporary 'test set', which is not used to build the calibration model, (2) developing a calibration model with the remaining (*n*-1) samples, (3) testing the established model with sample *i*, computing and storing the prediction error of the sample. The advantage of leave-one-out cross validation over random sub-sampling is that each sample is used for validation exactly once.