Journal of Pharmaceutical and Biomedical Analysis 53 (2010) 350–358

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Application of near infrared spectroscopy for rapid analysis of intermediates of *Tanreqing* injection

Wenlong Li, Lihong Xing, Limin Fang, Jue Wang, Haibin Qu*

Pharmaceutical Informatics Institute, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 25 January 2010
Received in revised form 27 March 2010
Accepted 8 April 2010
Available online 8 May 2010

Keywords:
Tanreqing injection
Fourier transform near infrared spectroscopy
Partial least squares
Rapid analysis
Model updating

ABSTRACT

A method for rapid quantitative analysis of four kinds of *Tanreqing* injection intermediates was developed based on Fourier transform near infrared (FT-NIR) spectroscopy and partial least squares (PLS) algorithm. The NIR spectra of 120 samples were collected in transflective mode. The concentrations of chlorogenic acid, caffeic acid, luteoloside, baicalin, ursodesoxycholic acid (UDCA), and chenodeoxycholic acid (CDCA) were determined with the HPLC–DAD/ELSD as reference method. In the PLS calibration, the NIR spectra were pretreated with different methods and the number of PLS factors used in the model calibration was optimized by leave-one-out cross-validation. The performance of the final PLS models was evaluated according to the root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), BIAS, standard error of prediction (SEP), and correlation coefficients (R). The R values in the prediction sets were all higher than 0.93, and the SEPs for the 6 compounds are 1.18, 6.02, 2.71, 155, 126, 30.0 mg/l, respectively. The established models were used for the liquid preparation process analysis of *Tanreqing* injection in three batches, and a model updating method was proposed for the long-term usage of the established models. This work demonstrated that NIR spectroscopy is more rapid and convenient than the conventional methods to analyze the intermediates of *Tanreqing* injection, and the presented method is helpful to the implementation of process analytical technology (PAT) in pharmaceutical industry of Chinese Medicines Injections.

1. Introduction

CMIs (Chinese Medicine Injections) have good effects on the treatment of difficult and complicated diseases and have been widely used in China. However, in recent years, due to some reports on the adverse drug reactions of several kinds of CMIs, improvement of their qualities is a more urgent requirement for the CMI manufacturers. The U.S. FDA recommends the pharmaceutical industry to adopt process analytical technology (PAT) to obtain the quality information of their products as early as possible [1]. In view of this guide, process analysis methods and techniques are needed to provide real time information to ensure the quality and consistency of the products. However, the complex constituents of CMIs create a challenge in establishing quality control approaches for them, especially the method based on the separation of the compounds needs to be measured, which is time and labour consuming. In this situation, it is a good choice to develop novel measurement methods for rapid characterization of CMIs. Two articles [2,3] reported using UV spectroscopy for rapid analysis the intermediates of *Qingkailing* injection, which provided a good example for the PAT research of CMIs.

Tanreqing injection is a widely used patent drug in China. It is made from five kinds of TCM (Traditional Chinese Medicine) extracts, namely: *Radix Scutellariae*, *Forsythia Suspense*, *Flos Lonicerae*, Bear gall powder, and Cornu gorais, and was used chiefly in treating infection of the upper respiratory tract and serious influenza [4,5], the medicine also has satisfactory efficacy on the defense of SARS (Serious Acute Respiration Symptom) and A/H1N1 flu. In its manufacturing process, several kinds of intermediates need to be analyzed to ensure that the operation runs steadily. Several articles reported the analysis methods of the medicinal materials used in the *Tanreqing* injection [6–8] and the finished product [9–11]. However, no report has been seen for the analysis of the *Tanreqing* injection intermediates. In traditional way, this task could be accomplished with high performance liquid chromatography (HPLC), which is accurate but time consuming and unable to satisfy the rapid determination needs. To resolve this issue, the near infrared spectrometry (NIRS) technology is an appropriate alternative method.

NIR is the region of the electromagnetic spectrum that extends from about 780 to 2500 nm (or 12,800–4000 cm^{-1}), and it has become widely used in various analysis areas, such as foods,

* Corresponding author. Tel.: +86 571 88208428; fax: +86 571 88208428.
E-mail address: quhb@zju.edu.cn (H. Qu).

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.04.011

pharmaceuticals, and petroleum. [12–15]. The usefulness of this technique is mainly attributed to its speed, accuracy, economy and precision against other analytical techniques [16]. In recent years, NIR has been more and more widely applied with success in the research of TCMs, and the main subjects are content measurement [17–20] and geographical identification [21–25]. The analytical targets are mainly powdered solids and the modes of NIR measurement are mainly diffused reflection. However, as to the sample of solutions, transmission or transflective mode should be adopted. Several articles reported the NIR method to determine the quality parameters of liquid samples, such as wine [26,27], beverage [28,29] and edible oil [30,31] in the transflective mode, but no article about the NIR application in the CMI intermediates has been reported.

The aim of research was to establish an NIR method for the rapid determination of the concentrations of 6 active ingredients in *Tanreqing* injection intermediates with HPLC–DAD/ELSD as reference method.

2. Materials and methods

2.1. Sample preparation

The samples were obtained from a TCM pharmaceutical factory (Shanghai Kaibao Pharmaceutical Co., Ltd., Shanghai, China). In a liquid preparation process of *Tanreqing* injection, after all the five kinds of TCM extracts were put into the solvent feed tank, 10 ml of the mixture was taken as test sample 1; and then the pH of the mixture in the tank was adjusted with sodium hydroxide solution, after being decolorized with active carbon, 10 ml of the filtrate was taken as test sample 2; the rest filtrate was treated by again filtrating with triple filter, and 10 ml of the filtrate was taken as test sample 3; the rest filtrate was then hyperfiltrated and made into semi-finished products, 10 ml of which was collected as test sample 4. The detailed technical parameters are omitted here. The four kinds of intermediate test samples mentioned above were collected from the product line for 30 batches and 120 samples were collected in total.

2.2. HPLC analysis

To determine the concentrations of the 6 active compounds, a gradient elution HPLC method was established. An Agilent 1100 HPLC system (Agilent Technologies, USA) with a vacuum degasser, a quaternary pump, an autosampler, a thermostatic column compartment, a diode array detector (DAD) and an evaporative light scattering detector (ELSD) were used. Separation was performed on Zorbax SB-C_{18} column (4.6 mm × 250 mm with 5 μm particle size) at 30 °C. The mobile phase consisted of (A) 0.15% CH_3COOH–H_2O and (B) 0.15% CH_3COOH–acetonitrile (v/v). The gradient program was as follows: initial 90% (A), at 0–10 min, linear change from 90% to 85% (A), at 10–20 min, linear change from 85% to 75% (A), at 20–30 min, linear change from 75% to 65% (A), at 30–35 min, linear change from 65% to 10% (A), keep 10% (A) at 35–45 min. Re-equilibration duration was 10 min between individual runs. The flow rate of the mobile phase was 0.8 ml min^{-1}. The detection wavelength for chlorogenic acid and caffeic acid was 326 nm, and for luteoloside and baicalin 276 nm. For the detection of UDCA and CDCA, the ELSD was used. The drift tube temperature and the nitrogen flow rate of the ELSD were set at 100 °C and 1.5 l min^{-1}, respectively. The samples in this research were 1:25 (v/v) diluted before centrifugation at the rotating speed of 10,000 rpm for 10 min, and 10 μl supernatant fluid was injected into the HPLC system for analysis.

2.3. NIR apparatus and software

The NIR spectra of the samples were collected at 4 cm^{-1} interval over the spectral region 4000–10,000 cm^{-1} with an Antaris MX FT-NIR System (Thermoscientific, Madison, USA) equipped with an optical fiber transflectance adapter. Samples were scanned with a 2 mm path length and equilibrated at 25 °C for 10 min before scanning to ensure that the samples were analyzed at the same temperature. Each spectrum was obtained by averaging 64 scans, and all the spectra were recorded as the logarithm of the reciprocal, $\log(1/R)$. The NIR instrument was controlled by a compatible PC, and a RESULT Operation workstation was used for data acquisition.

All the computations, including selection of the spectra wavebands, mathematical pretreatment, principal component analysis and partial least squares regression, were performed using TQ analyst software package (Version 8.0, Thermoscientific, Madison, USA) and The Unscrambler (Version 7.5, Camo Inc., Trondheim, Norway).

2.4. Spectral data pretreatment

The data acquired from NIR spectrometer contain background information and noises besides sample information. In order to obtain reliable, accurate and stable calibration models, it is necessary to preprocess the spectra before modeling. In this work, several data preprocessing methods, including derivation, multiplicative scatter correction (MSC), standard normal variate transformation (SNV), Savitsky–Golay (S.G.) filter, and Norris derivative filt were applied in order to minimize the interference effect, including correction of scatter effect, elimination of baseline shift, enhancement of spectral differences, and smoothing the spectra. The effects of different pretreatment methods were compared and the optimized combination was selected.

2.5. Calibration of models

The algorithm used for building the calibration models was PLS regression. The principle and the application of this algorithm were well documented in Refs. [32,33]. The performance of the final PLS model was evaluated in terms of RMSEC, RMSECV, RMSEP, BIAS, SEP and *R*. Detailed computing formula of the above mentioned parameters can be found in Ref. [19].

The optimum latent variable (LV) numbers of factors is determined using leave-one-out (LOO) cross-validation method. In the LOO cross-validation, the spectrum of one sample of the calibration set was removed from the set, and a PLS regression model was built with the remaining spectra of the calibration set, the left-out sample was predicted with this model and the procedure was repeated leaving out each of the samples of the calibration set. Predicted residual error sum square (PRESS) is the sum of square of deviation between predicted and reference values of all the samples in the cross-validation and the equation for the calculation of its value was expressed as below:

$$\mathrm{PRESS} = \sum_{i=1}^{n} (\hat{C}_i - C_i)^2$$

where C_i is the reference measurement result for sample i, $\hat{C}_i$ is the estimated value of the sample i. When the model is constructed with sample i removed, n is the number of calibration samples. The PRESS value decreases with the increase of the LV numbers, and when the PRESS value tends to be constant, the optimum LV numbers are obtained.

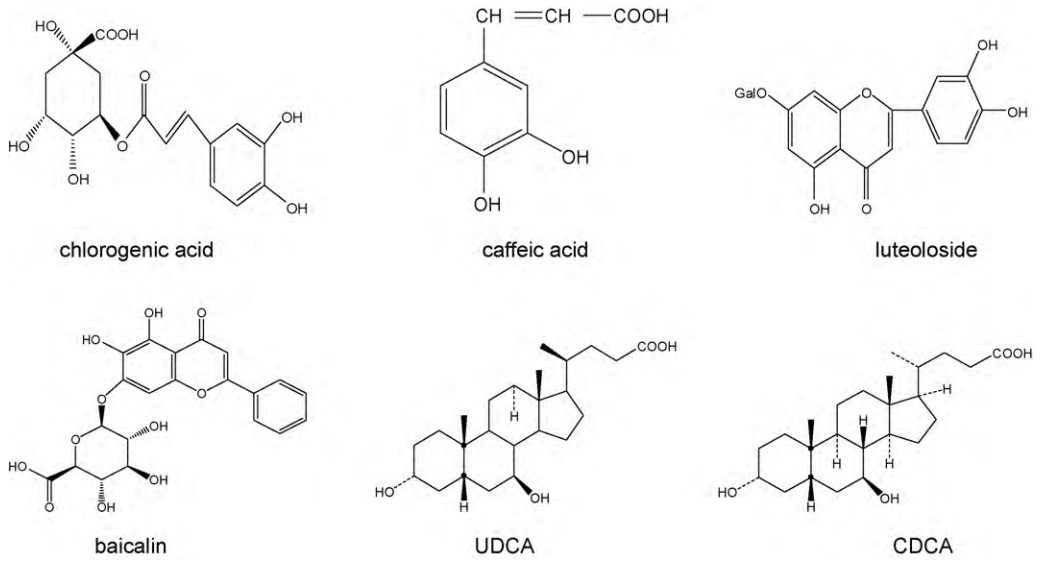

Fig. 1. The chemical structures of the six investigated analytes.

3. Results and discussion

3.1. Determination of active ingredients with HPLC–DAD/ELSD

Prior to NIRS quantitative analysis, a robust HPLC reference method has to be established as reference method. Chlorogenic acid, caffeic acid, luteoloside, and baicalin have good ultraviolet absorption, so they were detected with DAD; however, the UDCA and CDCA have no characteristic ultraviolet absorption and were detected with ELSD. The chemical structures of the six investigated analytes are shown in Fig. 1. Use the methods described in Section 2.2, the 120 samples were analyzed. All the 6 active compounds are baseline separated (see the chromatogram in Fig. 2) and can be determined accurately. The standard curves were established, and the methodology parameters were investigated before the realistic sample analyses (see in Table 1). Because the main purpose of this article is to establish an NIR method, more detailed information about the HPLC analysis method is omitted here.

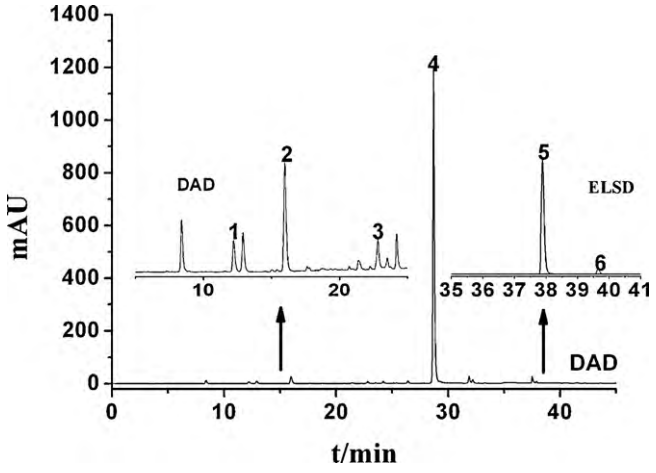

Fig. 2. HPLC chromatogram of *Tanreqing* injection intermediate at optimized conditions. (The peaks marked with 1–6 are chlorogenic acid, caffeic acid, luteoloside, baicalin, UDCA, CDCA, respectively.)

3.2. NIR spectral features

The raw and preprocessed spectra of *Tanreqing* injection intermediates are shown in Fig. 3. As can be seen in the raw spectra, 3 sharp fluctuations appear in the region of 7103–7320, 5310–5440, and 4540–4810 cm^{-1}, which register as 3 sharp peaks in the derivative absorption spectra. Those wavebands have correlation with the 2nd overtone stretch and deformation vibration of C–H in $–CH_2$, the vibration of the 2nd overtone of the carbonyl group, and the stretch and deformation vibration of O–H bands [34,35], respectively. The three bonds are widespread in the target molecules, so their concentration information can be reflected in the NIR spectra, and this can be viewed as the foundation of NIR quantitative analysis of the *Tanreqing* injection intermediates.

3.3. Principal component analysis

Whether the four kinds of intermediates can be calibrated in the same model needs to be inspected, and the method adopted is principal component analysis. From the score plots of the first 3 principal components (see in Fig. 4), which accounted for 98% of the variability in the spectral data set, there are no apparent

Table 1
The calibration curves and the methodology parameters of the reference method.

Compounds	t_R (min)	Calibration curves	R	Linearity ranges (μg/ml)	Limit of detection (ng)	Repeatability (RSD %, $n=6$)	Recovery (%, $n=3$)
Chlorogenic acid	12.15	$Y=36.38X-1.950$	0.9999	0.210–4.52	0.350	0.82%	100%
Caffeic acid	16.10	$Y=64.61X-1.926$	0.9999	0.165–5.40	0.285	0.90%	101%
luteoloside	22.82	$Y=16.22X+1.541$	0.9999	0.105–10.9	0.352	1.2%	99.9%
Baicalin	28.70	$Y=34.00X+329.6$	0.9998	20.3–406	0.530	0.85%	98.4%
UDCA	37.91	$\lg Y=0.4571\lg X+6.194$	0.9973	19.8–396	95.8	1.5%	99.0%
CDCA	39.63	$\lg Y=0.5416\lg X-1.881$	0.9999	17.9–179	60.1	1.5%	102%

Y: the peak area of chromatogram; X: the concentration of each active compound.

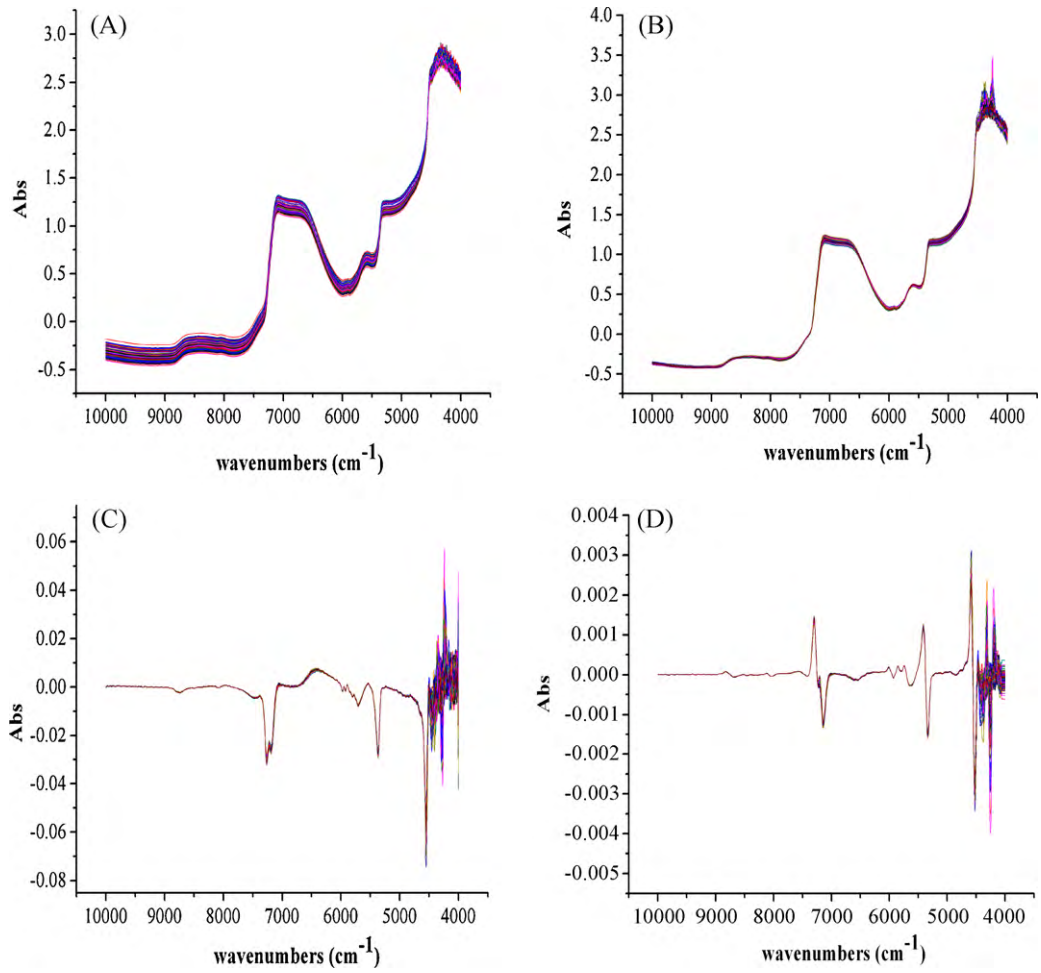

Fig. 3. Raw NIR spectra (A) and spectra preprocessed by MSC (B), Savitzky–Golay filter + 1-Der (C) and Norris derivative filt + 2-Der (D) of researched samples.

groupings in the sample set. The results indicate that there is no significant difference among the four kinds of *Tanreqing* injection intermediates, and they can be viewed as homogeneous samples. In fact, the major constituents of these four kinds of samples are very similar. The chief differences exist in the contents of active ingredients and impurities, and there is no qualitative difference. The research also indicates that the models established with single kind of intermediates are no better than the models established with the whole samples, and, moreover, the latter possesses better prediction performance due to the wider representativeness of the whole samples.

3.4. Calibration of models

3.4.1. Division of calibration and validation sets

A total of 120 samples were separated into calibration set and validation set according to the concentration values. For each model, the calibration set consisted of 70 samples, and the remaining 50 samples were used as validation set, which was used to validate the calibration model, and made sure that a stable model was achieved. The statistical values of the active ingredient contents in calibration and validation sets are listed in Table 2. The ranges of calibration set covered the larger scale and the concentrations are evenly distributed in both data sets, which was helpful for developing a stable and robust calibration model.

3.4.2. Selection of regression method

Different regression methods, such as multiple linear regression (MLR), principal component regression (PCR) and PLS were compared according to the quality of corresponding models. Taking the UDCA as example, the performance parameters of the optimized calibration models established with different regression methods are listed in Table 3. The results indicated that the PLS model has the best predictive ability. Similar results can be founded in the calibration models of other compounds; therefore, all the 6 models were developed by PLS method. It is well known that, in the PLS

Table 2
The statistics of active ingredient contents in calibration sets and validation sets.

Active ingredients	Average value ($mg\,l^{-1}$)	Calibration sets ($mg\,l^{-1}$)	Validation sets ($mg\,l^{-1}$)
Chlorogenic acid	66.75	56.66–83.47	57.55–82.46
Caffeic acid	59.83	33.60–79.46	34.35–76.80
Luteoloside	137.92	108.44–172.17	114.92–170.45
Baicalin	7070.00	5710.37–8642.94	5799.91–8593.09
UDCA	6868.34	4274.30–8537.75	4411.94–8184.80
CDCA	556.27	207.65–883.59	237.36–874.64

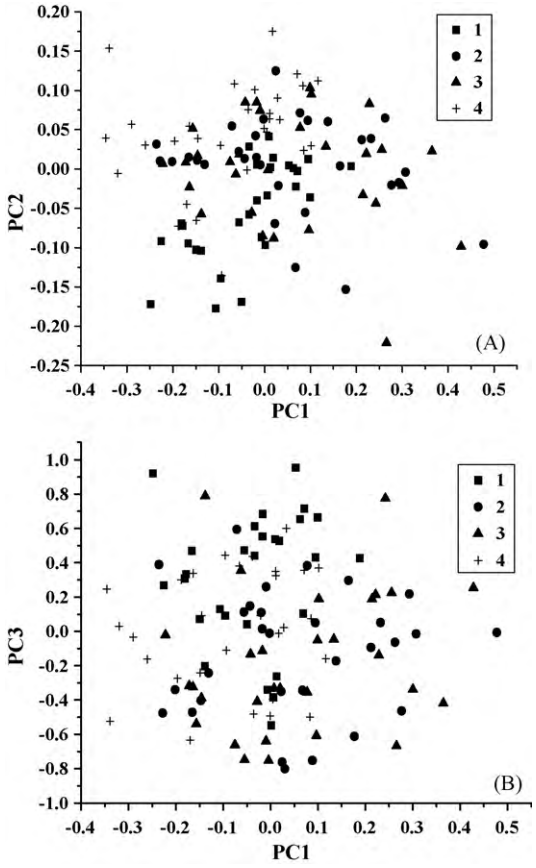

Fig. 4. Scattergrams of principal component analysis of samples from 4 different steps ((A) PC1–PC2; (B) PC1–PC3).

calibration, the most important influence factors include the waveband selection, the number of LVs, and the spectrum pretreatment method. So, the three issues are discussed hereinafter.

3.4.3. NIR waveband selection

The NIR region is divided into short-wave NIR (SW-NIR) and common NIR at 1300 nm (7692 cm^{-1}). The SW-NIR region is considered as the absorption band of high overtones, while the latter belongs to 1st or 2nd overtone. The absorption intensity will decrease when the overtone increases. Thus, SW-NIR is usually applied in the transmission analysis with long path length, and common NIR is used in diffuse reflection analysis [36]. The transflective mode combines the illumination characteristics of both transmission and reflective technologies. For this reason more information is obtainable; therefore, the useable region can cover the whole NIR bands [37], which is wider than either SW-NIR region or common NIR region. It can be seen from Fig. 3 that most spectral regions are smooth, but at the end part, there was some noise brought in by spectral differentiation process due to the optical fiber absorption. Therefore, the 4000–4500 cm^{-1} wave band should be discarded during the calibration process. In this research, the selected waveband was 4500–8000 cm^{-1}, which covered both

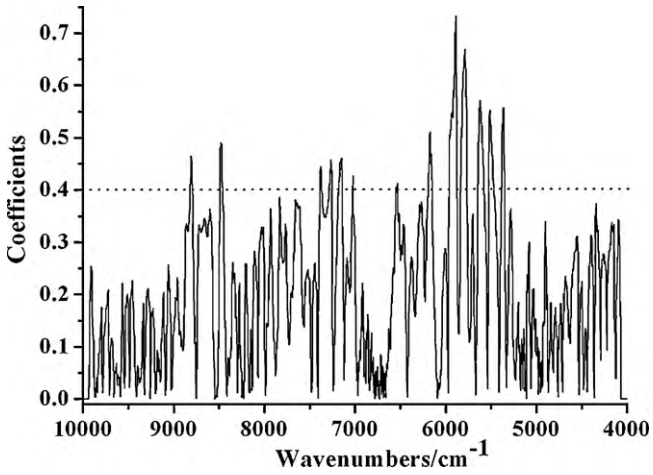

Fig. 5. Correlogram of the NIR spectra and the UDCA concentrations.

SW-NIR and common NIR regions. To confirm the correctness of the band selection, the correlation coefficients of the spectra were investigated. Taking the UDCA as example, the correlogram is shown in Fig. 5, in which it can be seen that the variables with higher coefficients (>0.4) are mostly distributed in the selected region, and also the similar cases can be observed with other compounds.

3.4.4. Comparing of different pretreatment methods for the spectra

Data pre-processing is critical in obtaining the quantitative models. In the *Tanreqing* intermediates of different stages, the insoluble solid particle may be different and generate different scattering effect. The MSC method is used to correct the variations in light scattering due to different particle size distribution [38] and make the spectra of the four kinds of intermediates much more similar. The spectra always have two kinds of deformation, one is parallel move, the other is rotate. Differentiations, including 1st and 2nd, help to eliminate the influence of the two kinds of spectra deformation, respectively. However, both differentiation methods may lead to amplification of noises. Therefore, before the derivation, the spectrum must be smoothed to filter noises. Different pretreatment methods for the spectrum were investigated; the criterion of selection was the performance of the calibration models, which was evaluated with RMSEC, RMSECV, and most of all, the value of RMSEP and the *R* of validation set. Taking the calibration model of baicalin as example, the calibrate results via different pretreatment methods were listed in Table 4, and the "MSC + S-G filt + 1st derivative" combination was selected. The spectra pretreatment methods for the calibration models of different compounds were selected in the same way, but the methods used for different compounds are different. To observe the spectra more intuitively, the spectra pretreated with different methods are also shown in Fig. 3.

3.4.5. Determination of the optimum LV numbers

Every calibration model has an optimum LV number. Less LVs used in the model will cause the 'under-fitting' phenomenon, which

Table 3
The performance parameters of the UDCA calibration models established with different regression methods.

Regression methods	Calibration		Cross-validation		External validation	
	R	RMSEC	*R*	RMSECV	*R*	RMSEP
MLR	0.8622	206	0.8537	211	0.8071	254
PCR	0.9209	159	0.9118	176	0.8994	202
PLS	0.9883	101	0.9822	113	0.9735	133

Table 4
The comparison of different calibration models of baicalin concentration developed with differently pretreated spectra.

Pretreated methods	LVs	Calibration		Cross-validation		External validation	
		R	RMSEC	*R*	RMSECV	*R*	RMSEP
Raw spectra	9	0.8092	276	0.3899	437	0.4247	488
MSC	8	0.9016	299	0.8972	274	0.8883	246
SNV	10	0.9191	259	0.8503	310	0.7592	303
S-G + 1d	7	0.7355	382	0.7648	188	0.8431	192
ND + 2d	6	0.8442	9273	0.9177	211	0.9015	241
SNV + ND + 2d	7	0.9109	178	0.8542	230	0.8548	244
MSC + S-G + 1d	7	0.9772	139	0.9630	155	0.9579	164

means insufficiency fitting. Including more LVs in the model will fit the calibration set better, but rupture the predictions of other samples. This phenomenon is called 'over-fitting' of a model, The cause is that specific information related to the calibration samples was included in the model, which may lead to poorer prediction results for the samples not in the calibration set. In this research, the optimum number of factors is determined using LOO cross-validation, and the relational graph of PRESS value against number of LV numbers was plotted to observe when the curves flatten out. Taking the CDCA calibration model as example, the determination method for the optimum LV numbers and the correlogram of the LOO cross-validation with the optimum LVs are shown in Fig. 6.

3.4.6. Establishment and validation of the calibration models

All spectra were pretreated with corresponding method, and 6 calibration models were established. After the calibration models were established, the external validation was performed to verify their accuracy. Taking chlorogenic acid calibration model as example, the correlation diagram and the residuals plot are shown in Fig. 7. Intuitively, the model has satisfactory fitting result and predictive power, the correlation is high and the residuals are low.

A good model should have lower RMSEC, RMSECV, RMSEP and higher *R*, but smaller differences between RMSECV and RMSEP, because large ones indicate that too many LVs are introduced in the model and the noises are also modeled. The performance param-

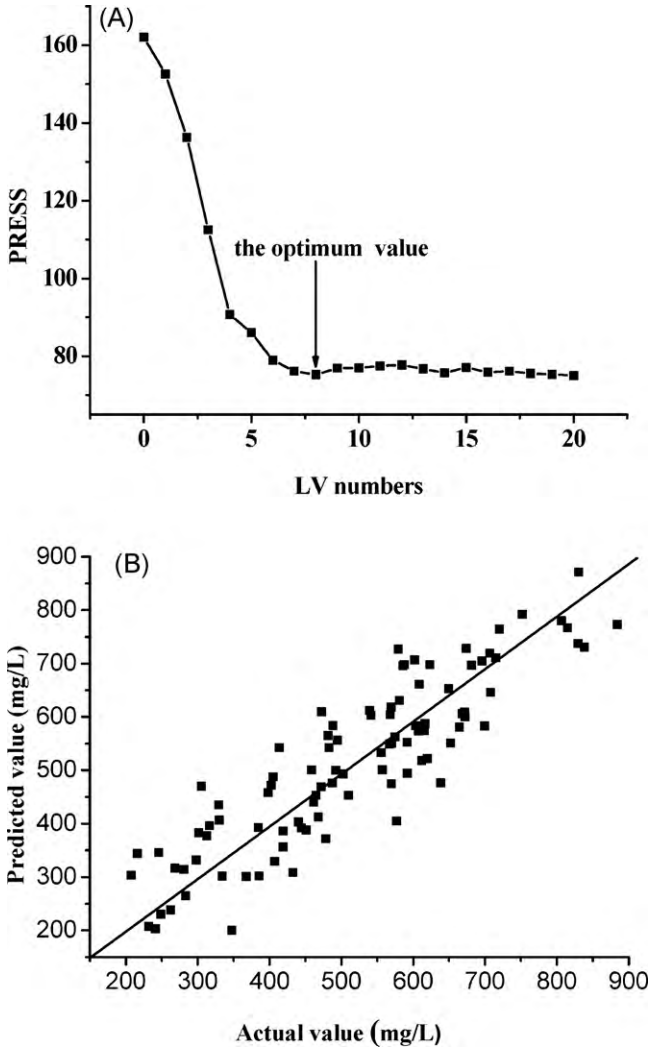

Fig. 6. Relational graph of PRESS and LV numbers (A) and the correlation diagram of cross-validation with the optimum LV numbers (B) of CDCA.

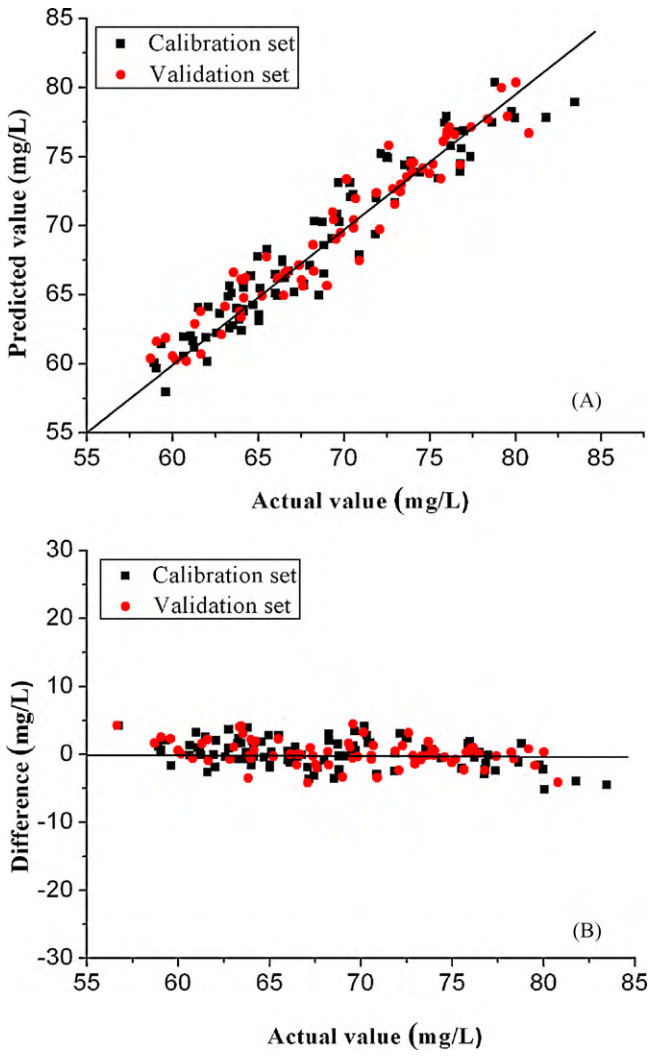

Fig. 7. Correlation diagram (A) and the residuals plot (B) of chlorogenic acid calibration models.

Table 5
The characteristic parameters of the established PLS models.

Active ingredients	Chlorogenic acid	Caffeic acid	Luteoloside	Baicalin	UDCA	CDCA
LVs	10	4	4	7	9	6
RMSEC	1.16	4.53	2.65	139	101	32.4
R (calibration)	0.9713	0.9640	0.9695	0.9772	0.9883	0.9592
RMSECV	1.19	6.33	2.88	155	113	37.2
R (cross-validation)	0.9633	0.9288	0.9432	0.9630	0.9822	0.9578
RMSEP	1.24	5.78	2.73	164	133	35.4
R (validation)	0.9532	0.9307	0.9309	0.9579	0.9735	0.9522
BIAS	0.00206	0.00310	−0.00212	−0.00114	0.00522	0.00118
SEP	1.18	6.02	2.71	155	126	30.0
RMSECV/RMSEP	0.9596	1.0951	1.0549	0.9451	0.9115	1.0508

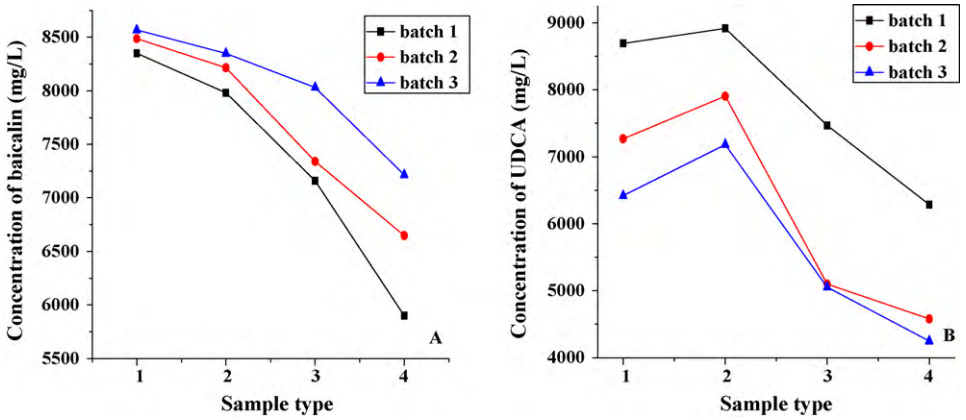

Fig. 8. Concentrations of baicalin (A) and UDCA (B) of 3 batches predicted with the established models.

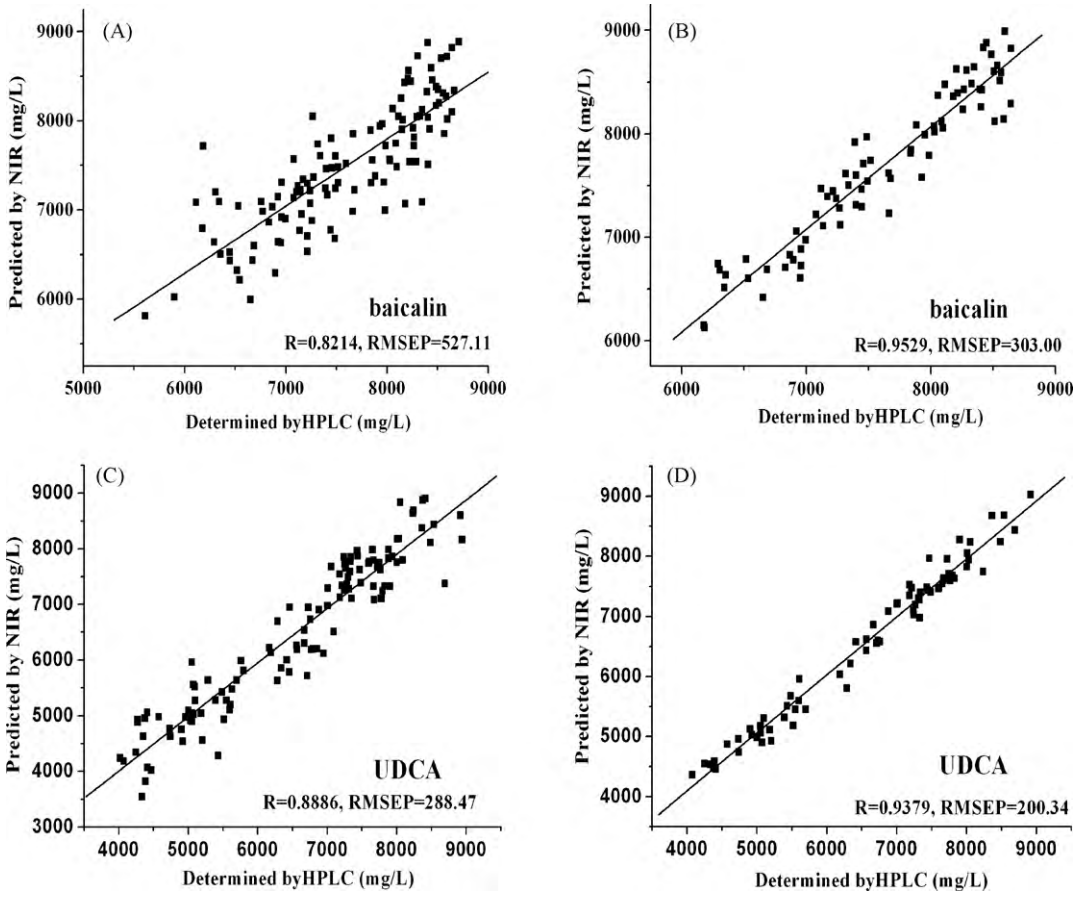

Fig. 9. Correlation diagrams of HPLC–NIR measured values for baicalin and UDCA before (A, C) and after (B, D) the calibration models updating.

eters of all the 6 models are listed in Table 5. The *R* values of the calibration, cross-validation and external validation were all higher than 0.93, and the values of RMSECV/RMSEP are all close to 1 (between 0.9 and 1.1), which indicates good predictive and extrapolation ability.

3.5. Application of the models for the analysis of unknowns

The established PLS calibration models are uploaded to the NIR instrument, and the instrument applies the calibrations to the NIR spectrum of unknown samples and outputs the concentrations of 6 active compounds simultaneously. The models were used to analyze samples from 3 different batches. The concentrations of baicalin and UDCA, which are the most important quality indicators of the *Tanreqing* injection, are shown in Fig. 8. It can be seen that the concentrations of baicalin in type 1 samples have minor differences, while the ultimate values vary greatly after 3-step operations. This phenomenon suggested that to obtain products with stable baicalin contents, the operating conditions should be rigidly controlled in these 3 steps. As to the concentrations of UDCA, the values rise after pH adjusted, and there is a sharp decline after filtration and hyperfiltration. From this feature, it can be concluded that to minimize the loss of UDCA, which is extracted from a costly medicinal material, bear gall powder, the pH should be raised appropriately within the bounds.

The established methods were used in manufacturing for routine analysis, and the results illustrated the feasibility and superiority of NIR technology. Using the presented method, main compounds' content is monitored and controlled during the manufacturing process, which allows us to check whether the product meets the specification criteria and also to ensure its final quality. This could be viewed as a PAT application and can be seen as the first step to establish a parametric release system.

3.6. Update of the established models

The calibration models are generally built according to the standard or referred method and the concentration variable at certain regions. Hence, this indirect analysis of NIRS results in an accumulated error occurrence. It takes long time and needs knowledge of chemometrics to build calibration models. However, each model is usually used in the limitation of time and spatial variation although some approaches for model transfer have been proposed. Thus, without further extended researches, these problems would always hinder the potential application. Model updating is a commonly used method to solve the problem of calibration invalidation, which can be achieved by adding spectra of new products or measuring on new instruments to the original model and calibrating it again. Thus, new model will cover more variability and can be utilized in new conditions [39].

The established models were used in the manufacture of *Tanreqing* injection for 3 months, and the accuracy of the prediction was decreased, so the models needed to be updated. To update the models, a new set of 120 samples were taken from the same product line and in the same way as mentioned above. When the old models were used directly to predict the concentration of 6 active compounds in the 120 samples, the correlation coefficient of the predicted value and the actual value is much lower than before. Then, one-thirds of 120 samples were added to the old calibration sets to reconstruct new models. Using the new models to predict the remaining two-thirds of the samples, the correlation coefficients are increased considerably and the inaccuracy became much smaller. This change can be observed in Fig. 9.

4. Conclusions

In this study, an FT-NIR method was developed for the simultaneous determination of 6 active ingredients, including chlorogenic acid, caffeic acid, luteoloside, baicalin, UDCA and CDCA in *Tanreqing* injection intermediates with an HPLC–DAD/ELSD method as reference method. The calibration models were validated with satisfactory *R* values and can be updated to ensure robustness for the long-term usage in industry manufacturing. Due to its rapid, accurate and robust properties, the presented model method was successfully used in monitoring four kinds of *Tanreqing* injection intermediates to improve the efficiency of quality control and assurance of the CMI as an alternative for the reference method. To our best knowledge, this research firstly reported the NIRS application in CMI intermediates analysis. The presented method provides a promising tool for *Tanreqing* injection intermediates analysis and can be consulted in solving similar problems. It is expected that the presented method could also be a promising tool for the implementation of PAT in other CMIs manufacturing.

Acknowledgments

Financial supports from the Program for New Century Excellent Talents in University (NCET-06-0515) and the Science and Technology Program of Zhejiang Province (2008C13004-1). The authors would also like to acknowledge the cooperators from Kaibao Pharmacy Co., Ltd., for supporting this work and for permission to publish the above results.

References

[1] Food and Drug Administration, Guidance for Industry: PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance, Food Drug Administration, Washington, DC, USA, 2004.
[2] X. Zhu, N. Li, X. Shi, Y. Qiao, Z. Zhang, Study on the Chinese Medicinal *Qingkailing* injections intermediate by support vector machines and ultraviolet spectrometry, Spectrosc. Spectrom. Anal. 28 (2008) 1626–1629.
[3] X. Zhu, N. Li, X. Shi, Y. Qiao, Z. Zhang, Study on Chinese Medicinal *Qingkailing* injection intermediate by least squares support vector machines and ultraviolet spectrometry, Chin. J. Anal. Chem. 36 (2008) 770–774.
[4] T. Wu, X. Yang, X. Zeng, P. Poole, Traditional Chinese medicine in the treatment of acute respiratory tract infections, Respir. Med. 102 (2008) 1093–1098.
[5] Y. Lv, D. Li, X. Liu, Q. Sai, Clinical observation on tanreqing injection curing 158 suffers of wind-heat cough, Acta Chin. Med. Pharm. 36 (2008) 20–22.
[6] C.R. Li, L. Zhou, G. Lin, Z. Zuo, Contents of major bioactive flavones in proprietary Traditional Chinese Medicine products and reference herb of *Radix Scutellariae*, J. Pharm. Biomed. Anal. 50 (2009) 298–306.
[7] M.T. Ren, J. Chen, Y. Song, L.S. Sheng, P. Li, L.W. Qi, Identification and quantification of 32 bioactive compounds in *Lonicera* species by high performance liquid chromatography coupled with time-of-flight mass spectrometry, J. Pharm. Biomed. Anal. 48 (2008) 1351–1360.
[8] A. Roda, F. Piazza, M. Baraldini, Separation techniques for bile salts analysis, J. Chromatogr. B 717 (1998) 263–278.
[9] J. Wang, H.B. Qu, Q. Shao, HPLC-UV-ELSD simultaneous determination of major component contents in Tanreqing injection, Chin. J. Pharm. Anal. 29 (2009) 1804–1807.
[10] W.M. Wu, L.J. Chen, L.K. Zhu, Determination of baicalin and chlorogenic acid in Tanreqing injection with high performance liquid chromatography, J. Pediatric Pharm. 12 (2006) 38–40.
[11] Y.X. Lou, G.H. Nie, N.Q. Liu, Content determination of baicalin, ursodeoxycholic acid and chenodeoxycholic acid in *Tanreqing* injection solution by HPLC-ELSD, Chin. J. Chin. Mater. Med. 34 (2009) 1862–1864.
[12] F. Liu, F. Zhang, Z. Jin, Y. He, H. Fang, Q. Ye, W. Zhou, Determination of acetolactate synthase activity and protein content of oilseed rape (*Brassica napus* L.) leaves using visible/near-infrared spectroscopy, Anal. Chim. Acta 629 (2008) 56–65.
[13] T.T. Zou, Y. Dou, H. Mi, J.Y. Zou, Y.L. Ren, Support vector regression for determination of component of compound oxytetracycline powder on near-infrared spectroscopy, Anal. Biochem. 355 (2006) 1–7.
[14] K. Hidajat, S.M. Chong, Quality characterization of crude oils by partial least squares calibration of NIR spectral profiles, J. Near Infrared Spectrosc. 8 (2000) 53–59.
[15] B. Liebmann, A. Friedl, K. Varmuza, Determination of glucose and ethanol in bioethanol production by near infrared spectroscopy and chemometrics, Anal. Chim. Acta 642 (2009) 171–178.

[16] L. Bokobza, H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise (Eds.), Near-Infrared Spectroscopy: Principles, Instruments, Applications, Wiley-VCH, Weinheim, Germany, 2002, p11.
[17] C.C. Lau, C.O. Chan, F.T. Chau, D.K.W. Mok, Rapid analysis of *Radix puerariae* by near-infrared spectroscopy, J. Chromatogr. A 1216 (2009) 2130–2135.
[18] I. Rager, G. Roos, P.C. Schmidt, K.A. Kovar, Rapid quantification of constituents in St. John's wort extracts by NIR spectroscopy, J. Pharm. Biomed. Anal. 28 (2002) 439–446.
[19] C.O. Chan, C.C. Chu, D.K.W. Mok, F.T. Chau, Analysis of berberine and total alkaloid content in *Cortex Phellodendri* by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultra-visible spectrometric detection, Anal. Chim. Acta 592 (2007) 121–131.
[20] Q. Chen, J. Zhao, S. Chaitep, Z. Guo, Simultaneous analysis of main catechins contents in green tea (*Camellia sinensis* (L.)) by Fourier transform near infrared reflectance (FT-NIR) spectroscopy, Food Chem. 113 (2009) 1272–1277.
[21] Y.A. Woo, H.J. Kim, K.R. Ze, H. Chung, Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of *Angelicae gigantis Radix*, J. Pharm. Biomed. Anal. 36 (2005) 955–959.
[22] J. Lu, B. Xiang, H. Liu, S. Xiang, S. Xie, H. Deng, Application of two-dimensional near-infrared correlation spectroscopy to the discrimination of Chinese herbal medicine of different geographic regions, Spectrochim. Acta. Part A 69 (2008) 580–586.
[23] X.F. Luo, X. Yu, X.M. Wu, Y.Y. Cheng, H.B. Qu, Rapid determination of *Paeoniae Radix* using near infrared spectroscopy, Microchem. J. 90 (2008) 8–12.
[24] L. Wang, F.S.C. Lee, X.R. Wang, Near-infrared spectroscopy for classification of licorice (*Glycyrrhizia uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content, LWT 40 (2007) 83–88.
[25] Y. Chen, M.Y. Xie, Y. Yan, S.B. Zhu, S.P. Nie, C. Li, Y.X. Wang, X.F. Gong, Discrimination of *Ganoderma lucidum* according to geographical origin with near infrared diffuse reflectance spectroscopy and pattern recognition techniques, Anal. Chim. Acta 618 (2008) 121–130.
[26] D. Cozzolino, M.J. Kwiatkowski, R.G. Dambergs, W.U. Cynkar, L.J. Janik, G. Skouroumounis, M. Gishen, Analysis of elements in wine using near infrared spectroscopy and partial least squares regression, Talanta 74 (2008) 711–716.
[27] H.Y. Yu, X.Y. Niu, H.J. Lin, Y.B. Ying, B.B. Li, X.X. Pan, A feasibility study on on-line determination of rice wine composition by Vis-nir spectroscopy and least-squares support vector machines, Food Chem. 113 (2009) 291–296.
[28] L.M. Reid, T. Woodcock, C.P. O'Donnell, J.D. Kelly, G. Downey, Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data, Food Res. Int. 38 (2005) 1109–1115.
[29] Y.N. Shao, Y. He, Nondestructive measurement of the internal quality of bayberry juice using Vis/NIR spectroscopy, J. Food Eng. 79 (2007) 1015–1019.
[30] L. Wang, F.S.C. Lee, X.R. Wang, Y. He, Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance, Food Chem. 95 (2006) 529–536.
[31] O. Galtier, N. Dupuy, Y. Le Dréau, D. Ollivier, C. Pinatel, J. Kister, J. Artaud, Geographic origins and compositions of virgin olive oils determinated by chemometric analysis of NIR spectra, Anal. Chim. Acta 595 (2007) 136–144.
[32] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[33] D.C. Slaughter, D. Barrett, M. Boersig, Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy, J. Food Sci. 61 (1996) 695–697.
[34] Q.S. Chen, J.W. Zhao, H. Lin, Study on discrimination of roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition, Spectrochim. Acta. Part A 72 (2009) 845–850.
[35] E. Stark, K. Luchter, Near infrared analysis (NIRA): a technology for quantitative and qualitative analysis, Appl. Spectrosc. Rev. 22 (1986) 335–399.
[36] H.Y. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, Trends Food Sci. Technol. 18 (2007) 72–83.
[37] W.L. Li, H.B. Qu, Rapid Quantification of phenolic acids in *Radix Salvia Miltrorrhiza* extract solutions by FT-NIR spectroscopy in transflective mode, J. Pharm. Biomed. Anal. 52 (2010) 425–431.
[38] H.H. Zhao, Y.L. Yan, The effects of noise on NIR analysis and related mathematic pretreatments and models, Spectrosc. Spectrom. Anal. 26 (2006) 842–845.
[39] X.B. Zhang, Y.C. Feng, C.Q. Hu, Feasibility and extension of universal quantitative models for moisture content determination in beta-lactam powder injections by near-infrared spectroscopy, Anal. Chim. Acta 630 (2008) 131–140.