



Construction of identification system for non-invasive analysis of macrolides tablets using near infrared diffuse reflectance spectroscopy[☆]

Yan-Chun Feng, Xue-Bo Zhang, Chang-Qin Hu^{*}

National Institute for the Control of Pharmaceutical and Biological Products, Beijing 100050, PR China

ARTICLE INFO

Article history:

Received 4 January 2009

Received in revised form 15 July 2009

Accepted 15 July 2009

Available online 23 July 2009

Keywords:

NIR diffuse reflectance spectroscopy

Identification system

Macrolides tablets

ABSTRACT

A NIR identification system consisting of two proposed models (Model I and Model II) has been developed for the analysis of 10 different products of macrolides tablets from different manufacturers. A total of 253 batches of the 10 products from 93 manufacturers in China were used for the system construction. First, a universal classification model (termed Model I) was constructed for 10 products using all the samples with the objective to distinguish the macrolides homologues which have the similar molecular structures. Secondly, 10 models (termed Model II) were developed separately for each product by using their samples. For each type of macrolides products, the two qualitative models are used in tandem as an identification system in mobile labs. Only when Model I and Model II both show acceptance results can an unknown sample be identified as "genuine". Internal and external validation showed almost 100% correct identity. Our study has shown that the analytical accuracy can be greatly improved when using this identification system and it will be efficient for quickly prescreening the drug quality in the open market and distribution channels.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Near infrared spectroscopy (NIRS) is a fast, nondestructive technique. Combined with chemometric algorithms, NIRS has become one of powerful tools in analytical area. In recent years, NIR spectroscopy has gained wide acceptance within the pharmaceutical industry [1–3] and several pharmacopoeia such as EP 5th edition [4], USP 32 [5], BP2009 [6] and Chinese Pharmacopoeia 2005 version [7] have adopted the NIR method either in an official chapter or in the appendix.

Counterfeit drugs are serious problems in countries all over the world, and are most severe in developing countries. The use of NIRS for identification of counterfeit drugs has been reported by several authors [8–13]. But so far, almost all of these reported NIR applications are used to identify if products were made by the authorized manufacturers. The identification usually was based on the comparison of the NIR spectrum of a sample with typical spectra of the authentic drug using multivariate modeling and classification algorithms. Meanwhile some parameters which could affect model construction were also considered.

[☆] Foundation item: National Key Technologies R&D Program Foundation of China (No. 2006BAK04A11).

^{*} Corresponding author. Tel.: +86 10 67095308; fax: +86 10 65115148.

E-mail address: hucq@nicpbp.org.cn (C.-Q. Hu).

Since 2002, a large scale study on rapid screening for counterfeit drugs using NIR has been carried out by the National Institute for the Control of Pharmaceutical and Biological Products (NICPBP, PR China). It is highly desirable that the proposed NIR prescreening system is based on universal models that can be used with multiple NIR instruments at different locations, and can accurately identify a given pharmaceutical product with the same INN (international nonproprietary name) made by different manufacturers all over China. The system based on such universal models will be much more efficient and operation will be simpler than the non-universal methods that need to be developed for different drug manufacturers, or adjusted for different instruments. In 2006, we reported on how to build a single universal quantification model for determination of API in different manufacturers' tablets [14]. At present, universal models for the identification of powder injections have been reported [15]. However, there are few reports mentioned on how to construct qualitative universal models for identification of oral preparations, which is more complicated than powder injections. In this study, the macrolides tablets including erythromycin, roxithromycin, clarithromycin, azithromycin, erythromycin ethylsuccinate, erythromycin estolate, acetylspiramycin, mydecamycin (meleumycin), kitasamycin, and acetylkitasamycin tablets which are structurally related compounds, were selected to develop such an identification system for oral preparation and investigate its practicability.

2. Experimental

2.1. Instrumentation

Three FT-NIR spectrometers, MATRIX-F2I008003, MATRIX-F2I001104 and EQUINOX55 from Bruker Optik GmbH (Ettlingen, Germany) were used in this experiment. All spectrometers are equipped with a 1.5 m fiberoptic diffuse reflectance probe and an extended TE-cooled indium gallium arsenide (InGaAs) detector. All the NIR spectra of different samples used to construct the system were recorded in these three FT-NIR spectrometers. Bruker OPUS software version 4.2 and 5.0 were used to collect and handle the data using chemometric analysis.

2.2. Materials

All the samples used in the experiment were collected from Chinese market by NICBPB between 2000 and 2004, and these samples were complied with their specifications.

2.3. Recoding of NIR spectra

All the samples of each product were randomly divided into three groups which were measured by three instruments, respectively. Six tablets were randomly selected from each batch for spectra collection. Diffuse reflectance spectra were recorded using fiber optic probes from one surface randomly of each tablet at 8 cm^{-1} resolution with 64 co-added scans over the spectral range $4000\text{--}12,000\text{ cm}^{-1}$. Therefore, six original spectra of each batch were used for the identification system development and validation.

2.4. Theory

Either the standard or factorization method available with OPUS software can be used to model construction. In this study, the factorization method [16] was used because it can compress data and suppress noise. A strategy for developing a classification model using factorization method consists of reference spectra collection, preprocessing, calculation of mean spectra and setup of distance threshold. After obtaining the NIR spectra, spectral preprocessing techniques and range were selected based on the best performance of the classification model, and then the average spectra for each group were calculated. In general, the larger the selectivity (described below), the better the performance of the classification model is. When factorizing an identification library, s average spectra (i.e. s groups) are transformed into s factor spectra, and then the original spectra was represented as linear combinations of these factor spectra (loadings):

$$a = T_{1a} \cdot f_1 + T_{2a} \cdot f_2 + T_{3a} \cdot f_3 + \dots$$

Vector a shows the original spectrum and the factor spectra are denoted f_1, f_2, f_3 , etc. T indicates the coefficients (scores) required to reconstruct the original spectrum. These factor spectra are orthogonal to each other. The effect a certain factor has on the reproduction of reference spectra is indicated by the respective Eigenvalue. The factor spectra are sorted according to these Eigenvalues. The first factor spectrum is the most important one and thus has the highest Eigenvalue. To calculate the spectral distance D between the two spectra a and b , the T coefficients are used in the factorization method:

$$D = \sqrt{\sum_i (T_{ia} - T_{ib})^2}$$

Another parameter used in the classification process was the threshold distance (D_T), which defines the tolerance of the classification model. In this case, a fixed algorithm was applied in calculating the threshold value. First, the distance (D_i) between the reconstructed spectrum and their corresponding average spectra was calculated. And then the threshold for each group is calculated from the $\max D_i$ and the standard deviation S_0 of D_i :

$$D_{\text{threshold}} = D_{\max} + X \times S_0$$

Whereas the default x value is 0.25.

The ability of the model to uniquely identify each group (cluster) was judged by the selectivity (S). Selectivity was calculated as the ratio of the distance (D) between average spectra to the sum of threshold values T_1 and T_2 (cluster radii): $S = D/(T_1 + T_2)$. When $S < 1$, the two groups overlap, and when $S > 1$, the two clusters are separated and can be uniquely identified.

When using the classification model to identify an unknown sample, the average spectrum of the unknown sample is compared with every mean spectrum in the library. There are two possible outcomes: (i) the hit quality (the distance between the test spectrum to the reference average spectrum) for the test sample was lower than the threshold value of the reference group, and no other hits were found to meet this criterion. The result is reported as "IDENTICAL," i.e., the test sample is identical to the reference. (ii) The hit quality is greater than the threshold, and no match is found of the test sample to any reference group in the spectral library. In this case the sample was reported as "NOT IDENTICAL." This situation could arise if the sample was not scanned properly, there was a lack of enough representativeness in the reference library, or the test sample was not the drug reported on its label (that was, this sample was counterfeit).

In this paper Model I and Model II were all constructed used factorization method. The scores and factors in this method depends on data set, changing the data set will change factors and therefore score of all spectra. There were 10 groups in Model I and only 2 groups in Model II, so the two models entries and threshold for the same group were different.

3. Results and discussion

3.1. Experimental design

The counterfeit drugs in the Chinese market can be divided into four types: (1) containing no active ingredients; (2) containing the incorrect amount of active ingredients; (3) containing different but structurally related active ingredients which have lower prices than the right one; (4) containing different active ingredients totally unrelated to the correct one. The first and fourth types of counterfeit drugs can be differentiated by comparing their NIR spectra to the typical spectra of authentic drugs, and the second type of counterfeit drugs can also be recognized by a universal NIR quantitative model [14,17]. Therefore, the identification of the third type of counterfeit drugs became the biggest challenge for NIR model construction. In this study, we hypothesized that 10 kinds of macrolides tablets as 10 different groups could be put into one classification model, and we attempted to differentiate these pharmaceutical products containing chemical similar compounds from each other by selecting appropriate parameter combinations (the characteristic spectral ranges, spectral pretreatment methods and chemometric technique for spectra comparison and their thresholds).

However it is difficult to select one kind of parameter combination to differentiate all these 10 products. A multi-level classification model (Model I) was proposed to solve this problem. The method of how to build up this model was as follows. The first

Table 1
Products used for Model I construction.

Group no.	Product (tablets)	Number of manufacturers used for calibration	Number of batches used for calibration	Number of NIR spectra used for calibration
1	Azithromycin	19	50	300
2	Erythromycin	16	50	300
3	Acetylspiramycin	2	5	30
4	Acetylkitasamycin	1	4	24
5	Erythromycin ethylsuccinate	26	50	300
6	Mydecamycin (meleumycin)	4	8	48
7	Roxithromycin	12	50	300
8	Clarithromycin	8	26	156
9	Kitasamycin	3	7	42
10	Erythromycin estolate	2	3	18

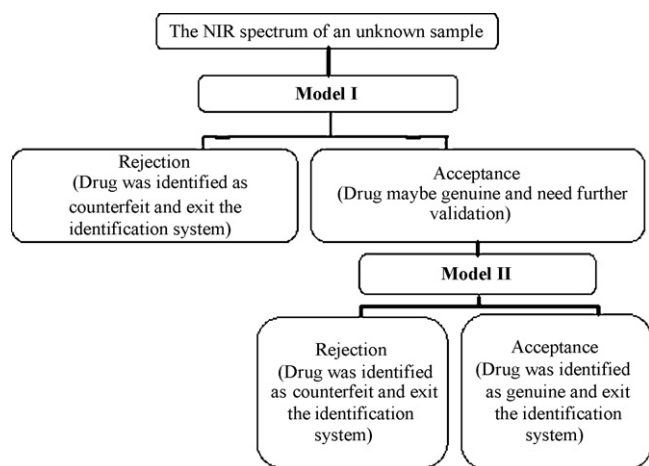


Fig. 1. Flowchart using NIR identification system to analyze an unknown sample.

step was to construct the primary classification model (1st level) by optimizing these parameters (e.g. the characteristic spectral ranges or the spectral pretreatment methods). There may be some groups confused with each other in the 1st level. So the second step was to put these groups into the next sub-level of the model (2nd level), often referred to as a sub-library. Then the choice of parameters was repeated for the sub-library. If there were still some groups could not be separated in the 1st and 2nd sub-libraries, a 3rd sub-library would be constructed in the same way until all groups can be uniquely identified.

Table 2
The parameters used in the Model II for Roxithromycin tablets.

Parameters	Values
Spectral pretreatment	Vector normalization; first derivative; 5 points smoothing
Spectral range	4200–6000 cm^{-1}
Chemometrics algorithm	Factorization method and only factor 1 was used for model construction

However, another problem showed up when the Model I construction was finished. Since different manufacturers might adopt different formulations of excipients to produce one kind of pharmaceutical products, the corresponding NIR spectra could have many variations. Therefore, the risk is that other preparations that were not included in Model I might be falsely recognized as one of these groups of Model I. For example, spectra of roxithromycin tablets from 18 manufacturers are quite different from each other, leading a larger threshold (0.72) than other groups in the selected spectral range. As a result, 281 of 300 spectra of vitamin C tablets were incorrect identified as roxithromycin tablet during the external validation, and the distance between the average spectra of roxithromycin tablets and vitamin C tablets was 0.68. Therefore, a proposed Model II was constructed for each product in order to solve this problem. Different from the objective of Model I to separate the pharmaceutical products with structurally related APIs, Model II was developed to prevent other products from being falsely recognized. For Model II, only samples of one product were included in and the spectral range utilized was different from Model I. When analyzing one sample, Model I and Model II were tandem used as an identification system (Fig. 1). Only when the two models showed acceptance simultaneously, the result of this unknown sample is

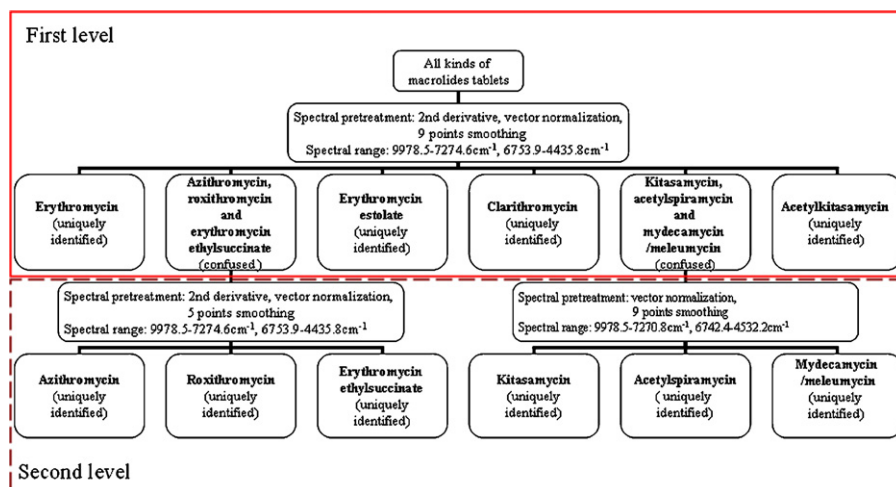


Fig. 2. The two-level NIR classification model for identification of macrolides tablets.

Table 3
Identification results of independent macrolides samples.

Product (tablets)	Number of batches used for validation	Number of NIR spectra used for validation	Accuracy (%)	
			Model I	Combination of Model I and Model II
Erythromycin	50	298	100	100
Azithromycin	35	210	100	100
Erythromycin ethylsuccinate	81	485	99.6	99.6
Erythromycin estolate	1	6	100	100
Roxithromycin	35	209	100	100
Total	202	1208	99.9	99.9

Table 4
External validation results of other products.

	Drug named on its label (tablets)	Number of batches used for validation	Number of NIR spectra used for validation	Identification results	
				Model I	Combination of Model I and Model II
Counterfeit drugs	Erythromycin ethylsuccinate	7	14	8 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Roxithromycin	1	2	2 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Astemizole	2	4	1 spectra was incorrectly identified as Kitasamycin	All samples were identified correctly
	Domperidone	2	4	2 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Gliclazide	1	2	All samples were identified correctly	All samples were identified correctly
	Compound Paracetamol and Amantadine Hydrochloride	1	2	2 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
Genuine drugs	Aspirin	50	299	All samples were identified correctly	All samples were identified correctly
	Estazolam	13	78	53 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Anminophylline	36	216	41 spectra were incorrectly identified as Erythromycin	All samples were identified correctly
	Pyrazinamide	20	119	All samples were identified correctly	All samples were identified correctly
	Ibuprofen	50	299	All samples were identified correctly	All samples were identified correctly
	Prednisolone Acetate	9	54	24 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Dexamethasone Acetate	31	183	150 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Diazepam	13	78	17 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Paracetamol	49	293	All samples were identified correctly	All samples were identified correctly
	Metronidazole	50	300	6 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Ribavirin	12	72	All samples were identified correctly	All samples were identified correctly
	Nimodipine	35	207	47 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Vitamin B1	50	300	161 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Vitamin B2	48	288	All samples were identified correctly	All samples were identified correctly
	Vitamin B4	9	54	18 spectra were incorrectly identified as Kitasamycin; 6 spectra were incorrectly identified as Roxithromycin	All samples were identified correctly
	Vitamin B6	50	300	31 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Vitamin C	50	300	281 spectra were incorrectly identified as Roxithromycin	All samples were identified correctly
	Cimetidine	50	300	All samples were identified correctly	All samples were identified correctly
	Nifedipine	11	65	8 spectra were incorrectly identified as Kitasamycin	All samples were identified correctly
	Nitroglycerin	8	46	All samples were identified correctly	All samples were identified correctly
Metformin Hydrochloride	50	300	2 spectra were incorrectly identified as Erythromycin	All samples were identified correctly	
Indometacin	35	209	7 spectra were incorrectly identified as Erythromycin	All samples were identified correctly	
Total	743	4388	867 spectra were incorrectly identified	All samples were identified correctly	
Accuracy (%)			80.24%	100%	

Table 5
The identification results randomly collected from 200 mobile labs.

No.	Products	Number of manufacturers were evaluated	Number of batches were evaluated	Number of batches were incorrectly identified	Accuracy (%)
1	Roxithromycin	23	90	0	100.00
2	Erythromycin ethylsuccinate	48	173	3	98.26
3	Azithromycin	36	88	13	85.23
4	Acetylspiramycin	35	167	52	68.86
5	Kitasamycin	12	43	15	65.12
6	Clarithromycin	14	35	17	51.43
7	Midecamycin	9	46	30	34.78

identified as “genuine”. The identification accuracy was significantly improved by this tandem strategy.

3.2. Construction of Model I

A multi-level classification model was constructed using the factorization method available in the IDENT package of Bruker OPUS with NIR spectra of the 10 macrolides products. The number of calibration spectra, manufacturers and batches of each product was shown in Table 1. The constructed classification model consisted of a main identification library with two sub-libraries, shown schematically in Fig. 2. The spectral preprocessing techniques and spectral ranges used in model construction were also shown in Fig. 2. The main library contained all the 10 products, while the sub-libraries only contained the products that could not be distinguished by the parameters used in the main library. More sensitive parameters were generally used in the sub-libraries for the differentiation of these products. When performing identification, if the spectrum of an unknown sample cannot be uniquely identified in the main library it will be automatically passed to the next level to be analyzed in the appropriate sub-library.

3.3. Construction of Model II

Model II for each of the products was constructed after Model I was finished. Here roxithromycin tablet was used as an example to explain the Model II construction. All spectra of roxithromycin tablets used in Model I were also participated in for Model II development. Different from Model I, there were only two groups in Model II: one group was the spectra of all the roxithromycin samples; another group was the spectra of the roxithromycin standard substance. The parameters for Model II of roxithromycin were shown in Table 2. In fact, there are no fixed rules on how to construct Model II. The groups in Model II should be fewer in number or different from those in Model I, so one has a better chance for different spectral range selection in Model II for one product. Additionally the chemometric technique (pretreatment and algorithm) may be different from Model I.

3.4. External validation

The proposed Model I and the combination using of Model I and Model II were challenged with independent samples of the 10 products. These batches were not previously used for the system construction. The detail results were shown in Table 3. In addition, the tablets of other products which were not included in the system were also considered for external validation (see Table 4).

3.5. Using and updating of the identification system

The developed identification system was equipped in more than 300 mobile vehicles in China and these mobile vehicles are used as mobile labs to quickly evaluate pharmaceutical products in the open market as well as distribution channels at the scene [18]. The

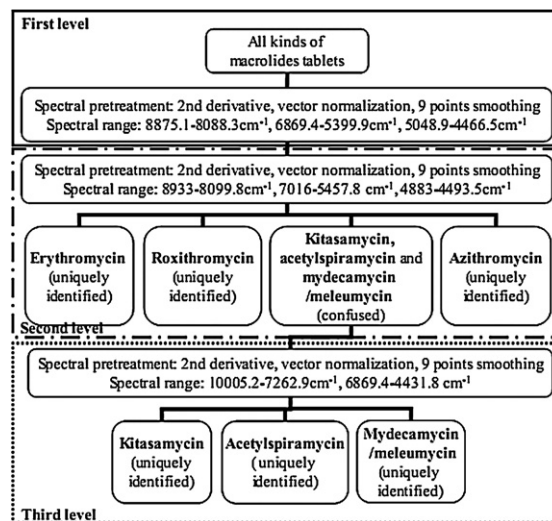


Fig. 3. The three-level Model I for macrolides tablets after updating.

results randomly collected from 200 mobile labs are summarized in Table 5. The accuracy of this identification system can reach up to 95% if the representative of the samples in the system was strong enough, which means the identification system can be used as a fast tool for drug prescreening. However, for those products such as acetylspiramycin, kitasamycin, clarithromycin and midecamycin, the identification system did not function well possibly because few samples were available for model construction.

In order to solve this problem, we put the new spectra of acetylspiramycin, kitasamycin, clarithromycin and midecamycin collected from the mobile labs into the identification system and updated it. These new NIR spectra came from the samples which were rejected by the identification system in the mobile labs but demonstrated as genuine using the legal laboratory method. The parameters of Model I were changed (Fig. 3.) by updating. We found that the accuracy of the identification system for those four products markedly increased in the latter using of this system in the mobile labs. It is obvious that model construction is not a process accomplished in one effort and timely model updating is a robust way by adding the information of new products to original model to construct a new universal identification model. By this way we can make sure all the identification systems in the mobile labs work well.

4. Conclusion

Our study has demonstrated that it is feasible to build an identification system for analysis of macrolides tablets from different manufacturers by near infrared diffuse reflectance spectroscopy. Meanwhile, we confirmed the necessity of the combination of Model I (a classification model containing several different groups of chemical similar compounds) and Model II (a model only focus

on single group of Model I, but other groups in this model should be fewer in number or different from Model I) for the complicated oral preparations. Model I was used to separate structurally related compounds and Model II was used to differentiate the products which will be likely to be ambiguous with the analyzed product but outside Model I. The identification accuracy can be dramatically improved when Model I and Model II was tandem used.

References

- [1] M. Blanco, J. Coello, H. Iturriaga, S. MasPOCH, C. de la Pezuela, *Analyst* 123 (1998) 135R–150R.
- [2] G. Reich, *Adv. Drug Deliv. Rev.* 57 (2005) 1109–1143.
- [3] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, *J. Pharm. Biomed. Anal.* 44 (2007) 683–700.
- [4] Council of Europe, *European Pharmacopoeia*, fifth edition, Aubin, Ligugé, 2004, pp. 59–63.
- [5] The United States Pharmacopoeial Convention, *The United States Pharmacopoeia 32 and The National Formulary 27*, United Book Press, Baltimore, 2008, pp. 622–626.
- [6] British Pharmacopoeia Commission, *British Pharmacopoeia 2009*, The Stationery Office, London, 2008, pp. A148–A152.
- [7] The State Pharmacopoeia Commission of The People's Republic of China, *Pharmacopoeia of the People's Republic of China*, Chemical Industry Press, Beijing, 2005, pp. A185–A187.
- [8] World congress of pharmacy and pharmaceutical sciences, *J. Pharm.* 271 (2003) 465–472.
- [9] S.H.F. Scafi, C. Pasquini, *Analyst* 126 (2001) 2218–2224.
- [10] B.A. Olsen, M.W. Borer, F.M. Perry, R.A. Forbes, *Pharma. Technol.* 26 (2002) 62–71.
- [11] M.J. Vredendregt, L. Blok-Tip, R. Hoogerbrugge, D.M. Barends, D. de Kaste, *J. Pharm. Biomed. Anal.* 40 (2006) 840–849.
- [12] O.Ye. Rodionova, L.P. Houmøller, A.L. Pomerantsev, P. Geladi, J. Burger, V.L. Dorofeyev, A.P. Arzamastsev, *Anal. Chim. Acta* 549 (2005) 151–158.
- [13] A.K. Deisingh, *Analyst* 130 (2005) 271–279.
- [14] Y.C. Feng, C.Q. Hu, *J. Pharm. Biomed. Anal.* 41 (2006) 373–384.
- [15] X.M. Chong, C.Q. Hu, Y.C. Feng, H.H. Pang, *Vib. Spectrosc.* 49 (2009) 196–203.
- [16] Bruker Optics GmbH, *Spectroscopic Software-Ident Manual Version 5*, 2004, pp. 51–57.
- [17] Y. Lei, Z.Y. Luo, C.Q. Hu, *J. Near Infrared Spectrosc.* 16 (2008) 349–355.
- [18] R. Mukhopadhyay, *Anal. Chem.* 79 (2007) 2622–2627.