

Near-Infrared Spectroscopy Combined with Absorbance Upper Optimization Partial Least Squares Applied to Rapid Analysis of Polysaccharide for Proprietary Chinese Medicine Oral Solution

Jiexiong Su¹, Xinkai Gao¹, Lirong Tan¹, Xianzhao Liu¹, Yueqing Ye¹, Yifang Chen², Kaisheng Ma², Tao Pan^{2*}

¹Infinitus (China) Company Ltd., Jiangmen, China

²Department of Optoelectronic Engineering, Jinan University, Guangzhou, China

Email: *tpan@jnu.edu.cn

Received 29 January 2016; accepted 13 March 2016; published 16 March 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Near-infrared (NIR) spectroscopy was applied to reagent-free quantitative analysis of polysaccharide of a brand product of proprietary Chinese medicine (PCM) oral solution samples. A novel method, called absorbance upper optimization partial least squares (AUO-PLS), was proposed and successfully applied to the wavelength selection. Based on varied partitioning of the calibration and prediction sample sets, the parameter optimization was performed to achieve stability. On the basis of the AUO-PLS method, the selected upper bound of appropriate absorbance was 1.53 and the corresponding wavebands combination was 400 - 1880 & 2088 - 2346 nm. With the use of random validation samples excluded from the modeling process, the root-mean-square error and correlation coefficient of prediction for polysaccharide were 27.09 mg·L⁻¹ and 0.888, respectively. The results indicate that the NIR prediction values are close to those of the measured values. NIR spectroscopy combined with AUO-PLS method provided a promising tool for quantification of the polysaccharide for PCM oral solution and this technique is rapid and simple when compared with conventional methods.

Keywords

Near-Infrared Spectroscopic Analysis, Proprietary Chinese Medicine Oral Solution, Polysaccharide, Absorbance Upper Optimization Partial Least Squares

*Corresponding author.

How to cite this paper: Su, J.X., Gao, X.K., Tan, L.R., Liu, X.Z., Ye, Y.Q., Chen, Y.F., Ma, K.S. and Pan, T. (2016) Near-Infrared Spectroscopy Combined with Absorbance Upper Optimization Partial Least Squares Applied to Rapid Analysis of Polysaccharide for Proprietary Chinese Medicine Oral Solution. *American Journal of Analytical Chemistry*, 7, 275-281.
<http://dx.doi.org/10.4236/ajac.2016.73025>

1. Introduction

Proprietary Chinese medicine (PCM) oral solution is a kind of health-care nourishing product, which is convenient to eat. According to the theory of traditional Chinese medicine, modern research results and practical experience, it is crafted by extracting some active components from a variety of Chinese herbal medicine. The compound polysaccharide, as the main active ingredients of PCM oral solution, can effectively regulate and enhance human immunity, prevent diseases and improve physical fitness. In the process of producing PCM oral solution, the real-time determination of the polysaccharide content is the necessary guarantee of monitoring the quality of the products. The conventional method [1] needs sample pretreatment and chemical reagent, which is difficult for real-time monitoring of production quality. Therefore, a rapid, simple, and reagent-free method has the significant value in practice.

Near-infrared (NIR) spectroscopy primarily reflects absorption of overtones and combination of vibrations of X-H functional groups (such as C-H, O-H, and N-H). Because of weak absorption strength, most of samples can be measured directly without preprocessing. This rapid, simple and non-destructive technique has obvious advantages and is commonly used in many areas, including agriculture [2]-[6], food [7] [8], environment [9], biomedicine [10]-[13] and pharmaceuticals [14] [15]. However, to the best of our knowledge, a quantification method for the determination of polysaccharide in the PCM oral solution using NIR spectroscopy has not been developed yet. Since the NIR spectra have serious overlapping and no significant absorption band, especially for the PCM oral solution with multiple components, appropriate chemometric methods must be employed to obtain wavelength optimization and quantitative analysis models with high signal-to-noise ratio (SNR). It can achieve extracting information variables and remove the noise interference. Partial least squares (PLS) regression has been recognized as an effective multivariate analysis method, and has been widely applied in the spectral analysis field [2]-[13].

Zengjian oral solution is a well-known brand product of PCM healthy oral solution, which is produced via refining polysaccharide from natural plant such as tremella, enoki and Chinese wolfberry etc. In this study, absorbance upper optimization PLS (AUO-PLS) was proposed, and NIR spectroscopy combined with AUO-PLS method was successfully applied to the rapid and reagent-free quantification of polysaccharide for Zengjian oral solution.

The stability of the spectral analysis model is very important in practice. Numerous experiments show that differences in partitioning of calibration and prediction sample sets can result in fluctuations in predictions and parameters (e.g. the number of PLS factors), thus leading to unstable results [3] [5] [8] [9] [12]. In the current study, a rigorous process of calibration, prediction, and validation based on randomness and stability was performed to achieve the goal of spectroscopic analysis.

2. Materials and Methods

2.1. Experimental Materials, Instruments, and Measurement Methods

A total of 1533 Zengjian oral solution samples were collected from infinitus (China) Company Ltd. The polysaccharide concentrations of these samples were measured with a UV-2300 UV-Vis spectrophotometer (Shanghai Tianmei, China) using mineral chameleon titration method. Mineral chameleon titration is capacity analysis method with potassium permanganate solution as titrant. It requires the use of chemical reagents, and by color reaction to achieve accurate quantification of the polysaccharide concentration of a sample. The measured values ranged from $330.26 \text{ mg}\cdot\text{L}^{-1}$ to $679.99 \text{ mg}\cdot\text{L}^{-1}$, and the mean value and standard deviation were 484.67 and $52.53 \text{ mg}\cdot\text{L}^{-1}$, respectively, which were used as the reference values for the calibration modeling of NIR spectroscopic analysis. Based on the obtained calibration model, a new method without chemical reagent for rapid determination of polysaccharide concentration of the PCM oral solution samples can be established with NIR spectroscopy.

An XDS Rapid ContentTM Solution Grating Spectrometer (FOSS, Denmark) equipped with a transmission accessory and a 2-mm cuvette was used for spectroscopy. The scanning spectrum spanned 400 nm to 2498 nm with a 2-nm wavelength gap, including the overall NIR region and a part of the visible region. Wavebands of 400 - 1100 nm and 1100 - 2498 nm were used for silicon and plumbous sulfide detection, respectively. Each sample was scanned thrice, and the mean value of the three measurements was used for modeling. The spectra were obtained at $25^\circ\text{C} \pm 1^\circ\text{C}$ and a relative humidity of $45\% \pm 1\%$.

2.2. Calibration, Prediction, and Validation Process with Stability

First, the 693 samples were randomly selected from a total of 1533 samples as the validation sample set, which were not subjected to the modeling optimization process. Then, the remaining 840 samples were used as modeling sample set and were further randomly divided into calibration (420 samples) and prediction (420 samples) sample sets for 100 times. The calibration and prediction models were established for all 100 divisions, and the model parameters were optimized depending on the mean prediction effects for all divisions to obtain objective and stable models.

The root-mean-square errors (SEC, SEP) and correlation coefficients (R_C , R_P) for calibration and prediction in modeling set were calculated, respectively. For each division (i) of calibration and prediction sets, they were denoted as SEC_i , SEP_i , $R_{C,i}$ and $R_{P,i}$, respectively, $i = 1, 2, \dots, 100$. The mean values (SEP_{Ave} , $R_{P,Ave}$) and standard deviations (SEP_{SD} , $R_{P,SD}$) of SEP_i and $R_{P,i}$ for all the divisions were further calculated, respectively. These values were used to analyze model prediction accuracy and stability. The equation $SEP^+ = SEP_{Ave} + SEP_{SD}$ was used as a comprehensive indicator of prediction accuracy and stability of a model. A smaller value of SEP^+ indicated higher accuracy and stability. The model parameters were selected to achieve minimum SEP^+ . The selected model was then revalidated against the validation sample set. The root-mean-square error and correlation coefficient of prediction in validation sample set were then calculated and denoted as SEP and R_P , respectively. The calculation formulas are as follows:

$$SEP = \sqrt{\frac{\sum_{k=1}^m (\tilde{C}_k - C_k)^2}{m-1}} \quad (1)$$

$$R_P = \frac{\sum_{k=1}^m (C_k - C_{Ave})(\tilde{C}_k - \tilde{C}_{Ave})}{\sqrt{\sum_{k=1}^m (C_k - C_{Ave})^2 \sum_{k=1}^m (\tilde{C}_k - \tilde{C}_{Ave})^2}} \quad (2)$$

where m is the number of validation samples; C_k and \tilde{C}_k are the measured and predicted polysaccharide concentrations of the k th validation sample, respectively; C_{Ave} and \tilde{C}_{Ave} are the mean measured polysaccharide value and the mean predicted polysaccharide value of all the validation samples, respectively.

2.3. Selection of Number of PLS Factors with Stability

The number of PLS factors (F) is an important parameter of PLS method that corresponds to the number of spectral latent variables corresponding to sample information. The selection of a reasonable F is both necessary and difficult. If F was set too small, the sample information in the spectra was unable to be fully reflected. If F was set too big, extra noises would be led into the model, the prediction ability would descend in both cases. In the present study, F was selected according to minimum SEP^+ based on all divisions for the calibration and prediction sample sets. Thus, the optimal number of PLS factors exhibited stability and practicality.

2.4. AUO-PLS Method

Lambert Beer's law is described by the following equation:

$$A(\lambda) = -\lg \left(\frac{I_1(\lambda)}{I_0(\lambda)} \right) = \lg \left(\frac{1}{T(\lambda)} \right) \quad (3)$$

where λ is the wavelength; $A(\lambda)$ is the absorbance; $I_0(\lambda)$ and $I_1(\lambda)$ are the intensity of incident light and the intensity of transmitted light through the sample, respectively; and $T(\lambda)$ is the transmittance, *i.e.*, the ratio of transmitted light intensity and incident light intensity. Conversely, Equation (3) can then be expressed as follows:

$$T(\lambda) = \frac{I_1(\lambda)}{I_0(\lambda)} = 10^{-A(\lambda)} \times 100\% . \quad (4)$$

According to the above equation, e.g. when $A(\lambda) = 4$, the transmitted light intensity was merely one ten thousandth of the incident light intensity, *i.e.*, the 99.99% of the incident light was absorbed by the sample. In this case, the transmitted light was very weak and was difficult to detect; it would thus likely cause noise in the spectrum. Therefore, wavelength selection with appropriate absorbance values, which correspond to a high quality of sample information and low levels of noise, is necessary. In this study, a novel PLS-based wavelength selection method, named absorbance upper optimization PLS (AUO-PLS) was proposed on the basis of the selection of the upper bound of absorbance, which can appropriately minimize noise bands. The specific steps are as follows:

Step 1: A region of wavelength screening (Δ) was set in advance for the entire scanning region according to the physical and chemical characteristics of the measured objects and the instrument properties. Meanwhile, in the average spectrum for all samples within the region Δ , the minimum and maximum values of absorbance were denoted as A_{\min} and A_{\max} , respectively. An appropriate step of absorbance (ε) was set.

Step 2: Set some value A^* , $A^* \geq A_{\min}$, the upper bound of absorbance A_{upper} was changed from A^* to A_{\max} with the step ε . According to relationship between wavelength and absorbance within the region Δ , for each A_{upper} , the absorbance interval ($A_{\min}, A_{\text{upper}}$) corresponded to a wavebands combination.

Step 3: Every obtained wavebands combination was employed for establishing the PLS calibration and prediction models. The corresponding SEP_{Ave} , $R_{\text{p,Ave}}$, SEP_{SD} , $R_{\text{p,SD}}$ and SEP^+ values were then calculated.

Step 4: According to minimum SEP^+ , the optimal A_{upper} was determined, and the wavebands combination corresponded ($A_{\min}, A_{\text{upper}}$) was also selected.

In this study, the region Δ was set to be the entire scanning region (400 - 2498 nm) with 1050 wavelengths. The A_{\min} was greater than or close to zero, and the A_{\max} value was less than or close to five, therefore, A_{\min} and A_{\max} were set to 0 and 5, respectively. Noticed that around 1450 nm is another obvious absorption peak with absorbance value 1.40. In order to retain the relevant information of the region, the A^* value was set as 1.40 (namely set $A_{\text{upper}} > 1.40$), because the main purpose in here is to remove the noise bands with saturate absorption. The absorbance step ε was set to 0.01 and the number of PLS factors (F) was set to $F \in \{1, 2, \dots, 30\}$. **Figure 1** shows a sketch map of the relationship between wavelength and absorbance for the case in which the absorbance value $A_{\text{upper}} = 1.53$ and the corresponding wavebands combination is 400 - 1880 & 2088 - 2346 nm.

3. Results and Discussion

3.1. Wavebands Combination Selection with AUO-PLS

The NIR spectra of the 1533 samples of Zengjian oral solution in the entire scanning region (400 - 2498 nm) are shown in **Figure 2**. As indicated in the figure, a saturate absorption region appears at about 1900 - 2000 nm. The saturate region was caused by strong absorption of water molecules and scattering of some tangible components

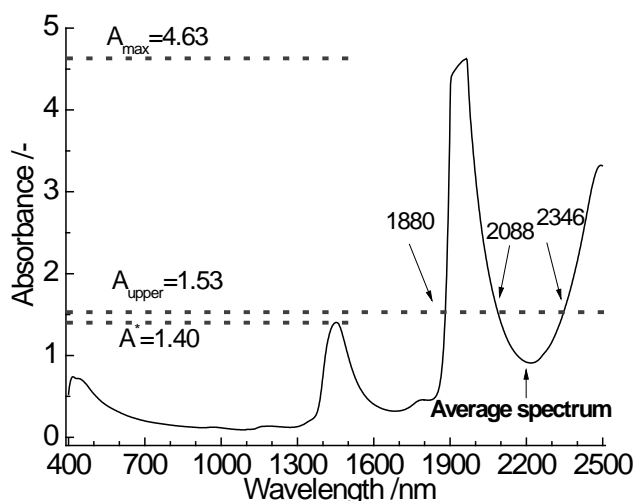


Figure 1. Sketch map for relationship between wavelength and absorbance.

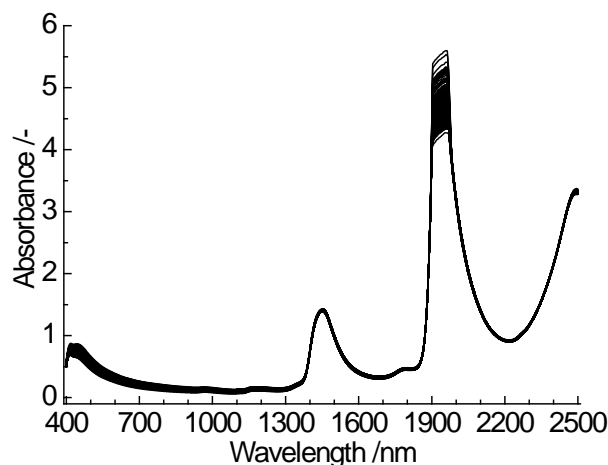


Figure 2. NIR spectra of 1533 samples of Zengjian oral solution in the entire scanning region (400 - 2498 nm).

in oral solution samples. AUO-PLS method mentioned in Section 2.4 was performed to avoid the noise wavebands with high absorption.

The SEP^+ values for each upper bound of absorbance A_{upper} are shown in Figure 3. The results showed that, the prediction polysaccharide value achieved the minimum SEP^+ when about $A_{upper} = 1.53$. The corresponding wavebands combination was 400 - 1880 & 2088 - 2346 nm with 871 wavelengths, and the prediction accuracy and stability results (SEP_{Ave} , $R_{P,Ave}$, SEP_{SD} , $R_{P,SD}$, and SEP^+) are summarized in Table 1. As a comparison, the full PLS model based on the entire scanning region was also established, and the prediction effects were also summarized in Table 1. The SEP^+ value for optimal AUO-PLS model was $27.81 \text{ mg} \cdot \text{L}^{-1}$, which was obviously better than that of the full PLS model. The relative SEP value (RSEP) for the optimal AUO-PLS model was 5.6%. The results show that, by avoiding the noise wavebands with high absorption, the prediction ability was improved and model complexity was reduced.

3.2. Model Validation

The randomly selected validation samples, which were excluded in the modeling optimization process, were used to validate the adopted AUO-PLS model. The PLS regression coefficients were calculated using the spectral data and measured polysaccharide concentrations of all modeling samples depending on the selected parameter F . The predicted polysaccharide concentrations of the validation samples were then calculated using the obtained regression coefficients and spectra of the validation samples.

Figure 4 shows the relationship between the NIR predicted and measured values of the 693 validation samples. The evaluation values (SEP and R_p) for validation effect were $27.09 \text{ mg} \cdot \text{L}^{-1}$ and 0.888, respectively. The results indicate that the NIR prediction values of the validation samples are close to those of the measured values. Satisfactory validation effects were achieved for the random samples because stability was considered in the modeling optimization process.

4. Conclusion

Wavelength selection is crucial for spectroscopic analysis, as it improves the effectiveness of prediction, reduces model complexity, and aids in the design of a specialized spectrometer with a high signal-to-noise ratio. The proposed AUO-PLS method focused on the optimization of upper bounds of absorbance to avoid noise interference caused by high absorbance. Based on the relationship between wavelength and absorbance, the appropriate wavebands combination was selected. NIR spectroscopy combined with the proposed AUO-PLS method was successfully employed for the reagent-free and rapid quantitative analysis of polysaccharide for Zengjian oral solution. A rigorous process of calibration, prediction, and validation based on randomness and stability was performed to produce objective and stable models. We believe that AUO-PLS has such applicability and can be also applied to other brand product of PCM healthy oral solution.

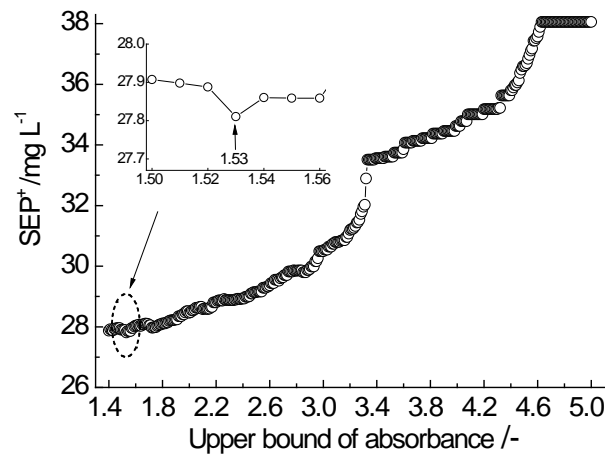


Figure 3. SEP⁺ values for each upper bound of absorbance with AUO-PLS method.

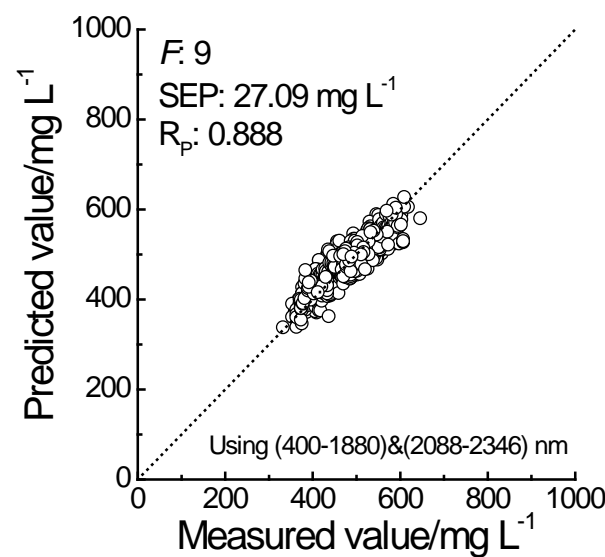


Figure 4. Relationship between the predicted and measured values of the validation samples with AUO-PLS method.

Table 1. Prediction effects of full PLS and AUO-PLS models for polysaccharide.									
Methods	Adopted wavebands	A_{upper}	F	SEP _{Ave} (mg·L ⁻¹)	SEP _{SD} (mg·L ⁻¹)	R _{p,Ave}	R _{p,SD}	SEP ⁺ (mg·L ⁻¹)	RSEP (%)
Full PLS	400 - 2498	-	12	36.75	1.30	0.723	0.019	38.05	7.6
AUO-PLS	400 - 1880 & 2088 - 2346	1.53	9	26.97	0.84	0.858	0.010	27.81	5.6

Acknowledgements

This work was supported by Foundation of Infinitus (China) Company Ltd.

References

[1] Yang, L.F. and Liu, Z. (2009) The Research of Determination Method of Polysaccharide Content in Compound Preparations. *Modem Traditional Chinese Medicine*, **28**, 244-246.

[2] Moron, A. and Cozzolino, D. (2002) Application of Near Infrared Reflectance Spectroscopy for the Analysis of Or-

- ganic C, Total N and pH in Soils of Uruguay. *Journal of Near Infrared Spectroscopy*, **10**, 215-221. <http://dx.doi.org/10.1255/jnirs.338>
- [3] Chen, H.Z., Pan, T., Chen J.M. and Lu, Q.P. (2011) Waveband Selection for NIR Spectroscopy Analysis of Soil Organic Matter Based on SG Smoothing and MWPLS Methods. *Chemometrics and Intelligent Laboratory Systems*, **107**, 139-146. <http://dx.doi.org/10.1016/j.chemolab.2011.02.008>
- [4] Pan, T., Wu, Z.T. and Chen, J.M. (2012) Waveband Optimization for Near-Infrared Spectroscopic Analysis of Total Nitrogen in Soil. *Chinese Journal of Analytical Chemistry*, **40**, 920-924.
- [5] Pan, T., Li, M.M. and Chen, J.M. (2014) Selection Method of Quasi-Continuous Wavelength Combination with Applications to the Near-Infrared Spectroscopic Analysis of Soil Organic Matter. *Applied Spectroscopy*, **68**, 263-271. <http://dx.doi.org/10.1366/13-07088>
- [6] Chen, J.M., Pan, T., Liu, G.S. and Han, Y. (2014) Selection of Stable Equivalent Wavebands for Near-Infrared Spectroscopic Analysis of Total Nitrogen in Soil. *Journal of Innovative Optical Health Sciences*, **7**, 1-9. <http://dx.doi.org/10.1142/S1793545813500715>
- [7] Chen, J.Y., Iyo, C. and Kawano, S. (2002) Effect of Multiplicative Scatter Correction on Wavelength Selection for Near Infrared Calibration to Determine Fat Content in Raw Milk. *Journal of Near Infrared Spectroscopy*, **10**, 301-307. <http://dx.doi.org/10.1255/jnirs.346>
- [8] Liu, Z.Y., Liu, B., Pan, T. and Yang, J.D. (2013) Determination of Amino Acid Nitrogen in Tuber Mustard Using Near-Infrared Spectroscopy with Waveband Selection Stability. *Spectrochimica Acta. Part A: Molecular and Biomolecular Spectroscopy*, **102**, 269-274. <http://dx.doi.org/10.1016/j.saa.2012.10.006>
- [9] Pan, T., Chen, Z.H., Chen, J.M. and Liu, Z.Y. (2012) Near-Infrared Spectroscopy with Waveband Selection Stability for the Determination of COD in Sugar Refinery Wastewater. *Analytical Methods*, **4**, 1046-1052. <http://dx.doi.org/10.1039/c2ay05856a>
- [10] Pan, T., Liu, J.M. and Chen, J.M. (2013) Rapid Determination of Preliminary Thalassaemia Screening Indicators Based on Near-Infrared Spectroscopy with Wavelength Selection Stability. *Analytical Methods*, **5**, 4355-4362. <http://dx.doi.org/10.1039/c3ay40732b>
- [11] Pan, T., Huang, W.J., Liu, Z.Y. and Yao, L.J. (2012) Near-Infrared Spectroscopic Analysis of Hemoglobin with Stability Based on Human Hemolysates Samples. *American Journal of Analytical Chemistry*, **3**, 19-23. <http://dx.doi.org/10.4236/ajac.2012.31004>
- [12] Xie, J., Pan, T., Chen, J.M., Chen H.Z. and Ren, X.H. (2010) Joint Optimization of Savitzky-Golay Smoothing Models and Partial Least Squares Factors for Near-Infrared Spectroscopic Analysis of Serum Glucose. *Chinese Journal of Analytical Chemistry*, **38**, 342-346. <http://dx.doi.org/10.3724/sp.j.1096.2010.00342>
- [13] Han, Y., Chen, J.M., Pan, T. and Liu, G.S. (2015) Determination of Glycated Hemoglobin Using Near-Infrared Spectroscopy Combined with Equidistant Combination Partial Least Squares. *Chemometrics and Intelligent Laboratory Systems*, **145**, 84-92. <http://dx.doi.org/10.1016/j.chemolab.2015.04.015>
- [14] Luybaert, J., Massart, D.L. and Vander Heyden, Y. (2007) Near-Infrared Spectroscopy Applications in Pharmaceutical Analysis. *Talanta*, **72**, 865-883. <http://dx.doi.org/10.1016/j.talanta.2006.12.023>
- [15] Reich, G. (2005) Near-Infrared Spectroscopy and Imaging: Basic Principles and Pharmaceutical Applications. *Advanced Drug Delivery Reviews*, **57**, 1109-1143. <http://dx.doi.org/10.1016/j.addr.2005.01.020>