

# Research and Reviews: Journal of Pharmacy and Pharmaceutical Sciences

## Qualitative and Quantitative Analysis of Heparin during Precipitation by Near-Infrared Spectroscopy

Lian Li<sup>#</sup>, Jinfeng Wang<sup>#</sup>, Hengchang Zang<sup>\*</sup>, Hui Zhang, Wei Jiang, Shang Chen and Fengshan Wang<sup>\*</sup>

School of Pharmaceutical Sciences, Shandong University and National Glycoengineering Research Center, No. 44 Wenhuaxi Road, Jinan 250012, P.R. China

<sup>#</sup>These authors have contributed equally to this work.

### Research Article

Received date: 28/03/2016

Accepted date: 07/05/2016

Published date: 12/05/2016

#### \*For Correspondence

Hengchang Zang and Fengshan Wang, School of Pharmaceutical Sciences, Shandong University and National Glycoengineering Research Center, No. 44 Wenhuaxi Road, Jinan 250012, P.R. China, Tel: +86-0531-88380268.

**E-mail:** zanghcw@126.com, wangfengshansdu@hotmail.com

**Keywords:** Near infrared spectroscopy, Heparin, Ethanol precipitation, Principal component analysis, Partial least square.

#### ABSTRACT

Heparin is a glycosaminoglycan (GAG) that plays an important role in the blood coagulation system. Its quality is of great importance, so it is necessary to develop a fast analytical method during the manufacture process to analyse the quality of heparin produced. In this study, the heparin contents of 80 samples collected from five batches during the precipitation process were analysed using near-infrared (NIR) spectroscopy and a chemometrics approach. This was done in order to improve the efficiency, to understand the process directly and accurately, and to reduce the variation in product quality during manufacturing. First, the principal component analysis (PCA) method was applied to study the stability and the characteristic trajectory of all of the batches of heparin from ethanol precipitation qualitatively. Then, partial least square (PLS) regression, combined with several spectral pretreatment methods and variable selection methods, was performed to quantitatively predict the heparin contents during the ethanol precipitation process. The results showed that the values of the coefficient of determination ( $R^2$ ), the root mean square error of prediction (RMSEP) and the residual predictive deviation (RPD) were 0.974, 1.105 g/l and 6.37, respectively. This approach has a considerable potential for on-line monitoring of the heparin contents of each ethanol precipitation process. Additionally, it will cause a large transformation in the pattern of production of the pharmaceutical industry by application of NIR spectroscopy in the future.

### INTRODUCTION

Pharmaceutical manufacturing processes usually consist of a series of unit operations, where each operation usually has a great influence on the quality attributes of the products. Hence, it is important to monitor the critical quality attributes (CQA) and the critical process parameters (CPP), which in turn affect the CQA<sup>[1]</sup>. Conventional pharmaceutical manufacturing is generally accomplished by off-line testing to evaluate quality. Consequently, the quality parameters acquired by laboratory testing would usually lag behind production. It is not conducive to the timely adjustment of process parameters, and it may even lead to rework or rejection of intermediate or final products, resulting in an increased cost of production<sup>[2]</sup>. Considering the quality and cost of production, it is important to use a fast and effective process analytical technology (PAT) in the manufacturing process for a solution of the aforementioned problems.

PAT is used to design, analyse, and control manufacturing through timely measurements (i.e. during processing) of

critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring the final product quality [3]. Since its initiation by the American Food and Drug Administration (FDA), PAT has been widely applied in the pharmaceutical industry [4] to provide a systematic understanding of the different variables that affect the processes, such as raw materials, equipment, temperature, and pH. Interactions among the variables affecting the formulation in real time and near-real time provide a scope for the development of robust processing conditions [5]. Recently, progress and interest in PAT have become a major trend in the pharmaceutical industry. Near-infrared (NIR) spectroscopy, one of the widely applied PAT tools, uses electromagnetic waves in the region of 800-2500 nm (12500-4000  $\text{cm}^{-1}$ ). The NIR region is dominated by bands that can be attributed to functional groups containing a hydrogen atom, such as C-H, O-H, and N-H [6]. Due to its fast, non-destructive, and low cost nature, NIR spectroscopy has proven to be an effective analytical tool in the pharmaceutical industry [7]. Additionally, there will be a large transformation in the pattern of production of the pharmaceutical industry by application of NIR spectroscopy.

Heparin is a glycosaminoglycan (GAGs) composed of polymers of alternating derivatives of  $\alpha$ -D-glucosamine and an O-sulfated uronic acid ( $\alpha$ -L-iduronic acid or  $\beta$ -D-glucuronic acid) [8]. It affects and regulates metabolism and physiological functions [9]. The production process of heparin ranges from the evaluation of the raw material to storage of the product. Ethanol precipitation is a critical unit operation in heparin production. It is a widely used technique for purifying and concentrating heparin from aqueous solutions. Ethanol is added as an Antisolvent owing to its simplicity and large capacity, which can efficiently precipitate heparin from an aqueous phase in the presence of a salt. There are numerous CPPs, including initial compositions of the extract, ethanol concentrations and precipitation temperature, whose fluctuations can lead to an unstable product and can also cause batch-to-batch variation [10]. Traditionally, the methods for monitoring the ethanol precipitation of heparin were primarily based on empirical or literature data. Research on the internal evolution was so scarce that it was difficult to accurately control the process of ethanol precipitation, leading to variations in the product quality. This is a common problem in GAGs production. This work discusses the utility of NIR spectroscopy for qualitative and quantitative monitoring of the process of ethanol precipitation of heparin. For process analysis by NIR spectroscopy, the qualitative methods generally involved principal component analysis (PCA) [11], multivariate statistical process control (MSPC) [12], moving block of standard deviation (MBSD) [13] and two-dimensional correlation spectroscopy [14], while the partial least squares (PLS) regression was mostly applied to quantitative analysis. Currently, NIR spectroscopic analysis of the ethanol precipitation process is mostly used in the production of traditional Chinese medicines (TCM) such as *cinobufacini* [15], *Lonicera japonica* [16,17], Danshen injection [10] and Danhong injection [18]. However, there has been no report about the monitoring of ethanol precipitation of heparin by NIR spectroscopy.

The aim of this study is to assess whether NIR spectroscopy has the potential to provide in-process information and monitoring of heparin precipitation from ethanol. PCA and PLS regression methods were applied in this work for qualitative and quantitative analysis of NIR spectroscopy. First, PCA was applied to identify the characteristic trajectory of all of the batches of heparin precipitated from ethanol. Then, PLS regression, combined with several spectral pre-treatment methods and variable selection methods, was performed to predict the heparin contents during the ethanol precipitation.

## MATERIALS AND METHODS

### Materials and chemicals

Heparin sodium and its reference standard were provided by Zaozhuang Sainuo Kang Biochemical Co. Ltd. (Zaozhuang, China) and the National Institute for the Control of Pharmaceutical and Biological Products (Beijing, China), respectively. Concentrated sulfuric acid was purchased from Beijing Chemical Works (Beijing, China). Anhydrous ethanol was obtained from Tianjin Fuyu Chemical Co., Ltd (Tianjing, China). All chemicals used were of analytical grade or higher purity. The deionized water was obtained from a Millipore Milli-Elix/RiOs ultra-pure water system (Bedford, MA, USA).

### Methods

#### Ethanol precipitation

Five hundred millilitres of an aqueous solution of heparin sodium (solid-liquid ratio 7:100, w/v) was adjusted to pH 6.5 with 0.1 M hydrochloric acid. Then, 600 ml of a well-stirred, anhydrous ethanol was continuously pumped into the aqueous heparin sodium solution at a flow rate of 9 ml/min. After addition of the anhydrous ethanol was complete, stirring was stopped and the suspension was allowed to stand. The entire process lasted 80 min, and then 1 ml of the suspension was collected in tubes. Samples were collected every 5 min, and 16 samples were collected for every batch. Therefore, five batches yielded 80 samples. All of the collected samples were centrifuged, and the supernatants were stored for further research.

#### Reference method

The heparin content of the supernatants was determined by UV-Vis spectroscopy, as reported in this work, with minor modifications. A solution containing a known concentration of heparin sodium was prepared and used as the reference standard. Each supernatant was diluted with water to attain a heparin concentration that was approximately equal to that in the standard solutions. Then, 0.4 ml of the sample solutions and five standard solutions (400  $\mu\text{g}/\text{ml}$ ) containing 0 to 160  $\mu\text{g}$  of heparin were collected in test tubes. Next, 3 ml of sulfuric acid (90% (v/v)) containing 0.025 M sodium tetra borate was carefully added to each

test tube. The mixture was then heated in a water bath at 90 °C for 10 min and then immediately cooled to room temperature. The absorbance was measured at 298 nm using a Cary 100 Bio UV/Vis spectrophotometer (Varian Inc., Walnut Creek, CA). A control was prepared with water using the same method.

### NIR analyser

The transmittance spectra of the supernatants were acquired by an Antaris II Fourier-transform near infrared spectrophotometer (Thermo Fisher Scientific, USA) with an InGaAs detector and RUSULT 3 software. All of the samples were packed into 6 × 50 mm glass tubes (Kimble Chase, Germany) and then scanned over a range of 10000 to 4000 cm<sup>-1</sup> using transmission mode in the laboratory environment (room temperature, 30-50% relative humidity). The number of co-added scans and resolution were 32 and 4 cm<sup>-1</sup>, respectively.

### Data processing

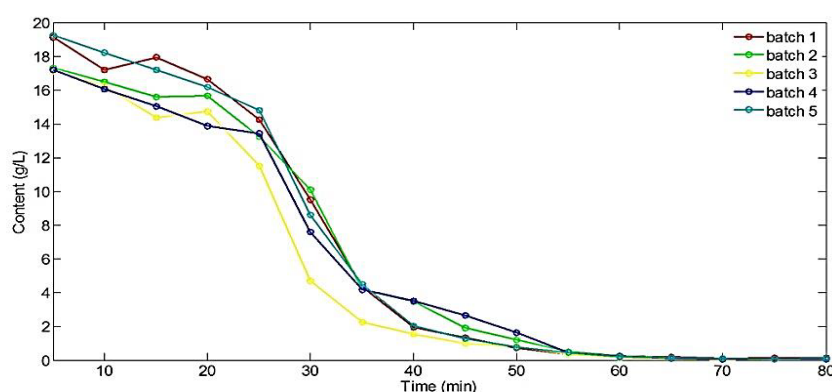
Principal component analysis (PCA) and partial least squares (PLS) regression methods were used for data processing. All of the calculations were performed using Matlab 2010a (Mathwork Inc., USA). A score scatter plot obtained using PCA was used to identify the characteristic trajectory of all batches. PLS regression was performed to establish a quantitative analysis model to predict the heparin content during the ethanol precipitation process. Four batches of samples were randomly divided into a calibration set, and the remaining batch was used for the validation set. The transmission spectra were pre-treated with several pre-processing and variable selection methods to produce simpler models and better prediction. The optimal number of latent variables was determined using venetian blinds cross-validation. The coefficient of determination ( $R^2$ ), root mean square error of prediction (RMSEP) and root mean square error of cross-validation (RMSECV) were used to evaluate the quality of the models. In addition, the residual predictive deviation (RPD), which is defined as the ratio of the standard deviation in the validation set and the standard error of validation, was used to further evaluate how well the calibration model could predict compositional data.

## RESULTS AND DISCUSSION

### Determination of the heparin content

#### Results of reference methods

The standard curve for determining heparin concentrations was prepared by plotting heparin concentration along the horizontal x-axis and absorbance along the vertical y-axis ( $y=0.0077x-0.0398$ ,  $r^2=0.9999$ ). The content distribution of the heparin samples during the ethanol precipitation process, divided into the calibration set and the validation set, is shown in **Table 1**, while the heparin concentration in each sample collected over time is shown in **Figure 1**. The heparin content of the samples was found to vary between 19.25 and 0.038 g/l (**Table 1**). Additionally, all of the batches showed a similar trajectory for the change in heparin content over time (**Figure 1**).



**Figure 1.** Time course of heparin content during the ethanol precipitation process.

**Table 1.** Information table for the calibration and validation sets.

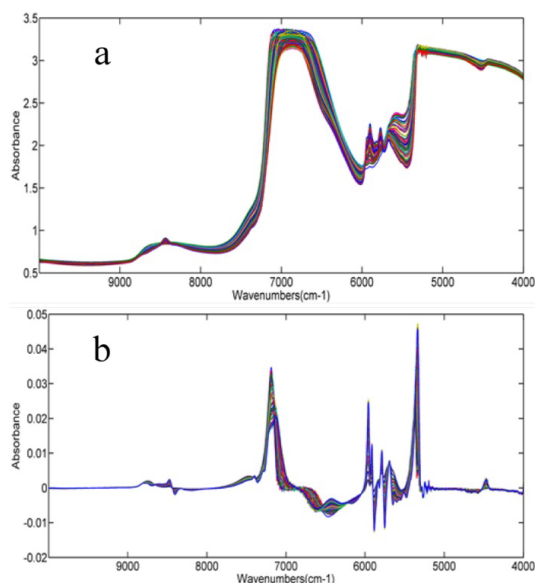
Sample sets	n <sup>a</sup>	Range(g/l)	Mean(g/l)	SD <sup>b</sup> (g/l)
Calibration set	64	19.25-0.038	6.08	7.12
Validation set	16	17.32-0.052	6.26	7.04

<sup>a</sup> n=number of samples, <sup>b</sup> SD =standard deviation.

### Spectral interpretation

Raw NIR spectra of the supernatants from ethanol precipitation are shown in **Figure 2a**. The peak at approximately 6900 cm<sup>-1</sup> corresponds to a strong absorbance originating from a combination of HOH symmetric and asymmetric stretching vibrations, while the peak at approximately 5150 cm<sup>-1</sup> corresponds to a strong absorbance originating from a combination of HOH

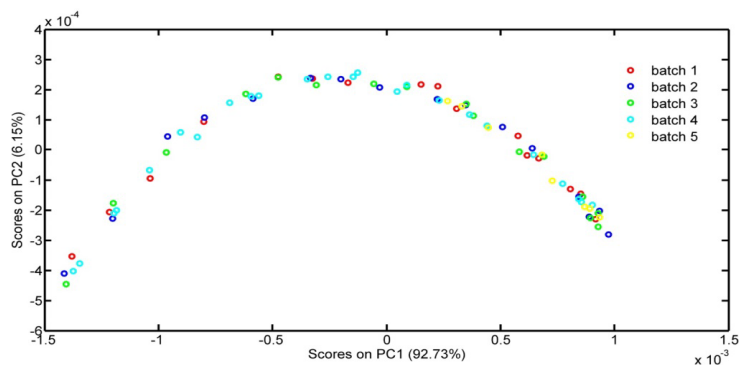
asymmetric stretching and bending vibrations [19]. However, unlike the spectrum of the aqueous heparin solution that does not contain ethanol [20], there were conspicuous peaks at approximately 6000-5400  $\text{cm}^{-1}$  in the spectra of the supernatants that can be attributed to the absorbance of ethanol. These peaks appeared after the addition of ethanol, and the absorbance increased with the increase in the amount of ethanol added. Hence, the spectra were further subjected to pre-processing to linearize the response of the variables and to remove extraneous sources of variation that are of no interest to the analysis [21]. As shown in **Figure 2b**, after pre-processing of the spectra using the first-order Savitzky-Golay derivative method with a filter width of 15 data points, the sensitivity of the spectra improved and the overlapping peaks in the original spectra could be distinguished. This was conducive to the extraction of specific information from the spectra and has been used for subsequent multivariate analysis.



**Figure 2.** Near-infrared spectra of supernatants from ethanol precipitation. (a) Raw spectra. (b) Preprocessed spectra by first-order Savitzky-Golay derivative with a filter width of 15 data point.

### Qualitative analysis by PCA

PCA is a widely used tool in chemometrics for data compression and information extraction. PCA finds combinations of variables that describe major trends in the data. Here, pre-processed NIR spectra of five batches were analysed using PCA, where the characteristic trajectory of the all batches could be examined by scores and loadings. The heparin ethanol precipitation system mainly consisted of water, ethanol and heparin. Heparin precipitated continuously with the addition of ethanol; thus, the ratio of the three components changed continuously in the supernatants. **Figure 3** shows the score scatter plot for the first two principal components, which explained 98.88% of the total variance. The first principal component (PC1) score with 92.73% of the explained variance increased steadily as the ethanol precipitation process proceeded, while the second principal component (PC2), explaining 6.15% of the total variance, increased first but then started to decrease after 40 min. The figure demonstrated that all five batches had a similar score trajectory, and the scores of all of the batches started at the bottom left and ended at the bottom right. As shown in **Figure 1**, heparin initially precipitated rapidly with the addition of ethanol, but after 40 min, a small amount of heparin remained in the supernatant and precipitated slowly. After 80 min, very little heparin could be detected in the collected samples. Accordingly, **Figure 3** showed that the principal component scores reached the inflection point at 40 min. This result suggested that the score scatter plot has the potential to reflect the stability of the internal ethanol precipitation system. First, a given mass of heparin is precipitated with the addition of ethanol, thus destroying the equilibrium of the system. Then, the heparin content curve and the principal component scores both reached the inflection point at 40 min. Later, the small amount of the heparin that remained in the supernatants gradually precipitated, and the system stabilized.



**Figure 3.** Score scatter plot of the NIR spectra for the five batches of heparin precipitated.

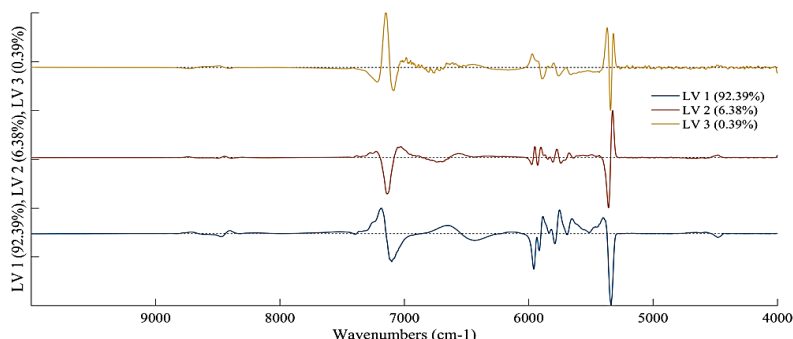
## Quantitative analysis by PLS

In this section, PLS regression was performed to establish the quantitative analysis models that would predict the heparin concentrations during the ethanol precipitation process. First, a PLS quantitative analysis model was developed with fully pre-processed spectra, and the values of RMSECV and RMSEP were 1.499 g/l and 1.114 g/l, respectively. However, the variables are not of equal importance to the model, and some of them are noisy enough to even disrupt the analysis. Variable selection is a critical step in modelling because it has been shown that the predictive ability can be enhanced and the complexity of the model can be reduced by choosing a judicious variable selection method [22]. For this purpose, four different variable selection methods, e.g., correlation coefficient, genetic algorithm (GA), forward interval partial least squares (FiPLS) and backward interval partial least squares (BiPLS) models, were employed to establish calibration models of good predictive ability based on the preprocessed spectra. The number of latent variables was selected based on the RMSECV calculated using the venetian blinds cross-validation method. Random reordering of the data set was performed five times in order to reach an optimized solution [23]. The results obtained with PLS regression after variable selections are shown in **Table 2**.

**Table 2.** Results of the variables selection algorithms for the ethanol precipitation process.

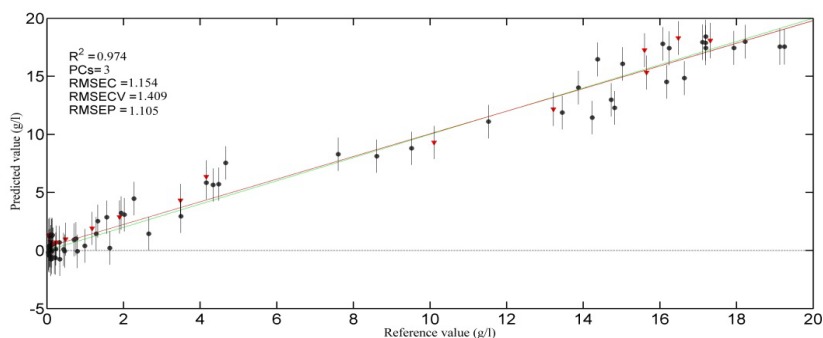
Methods	RMSECV (g/l)	R <sup>2</sup>	RMSEP (g/l)	PCs	RPD
-	1.4998	0.972	1.1139	3	6.32
Correlation Coefficient	1.4105	0.973	1.3688	5	5.14
GA	1.2920	0.974	1.7592	4	4.00
FiPLS	1.4089	0.972	1.1054	3	6.37
BiPLS	1.3096	0.978	1.3242	4	5.32

Incidentally, compared to the previous PLS model established with full spectra, the RMSECV values obtained from the calibration models were smaller for different variable selection methods. However, only the RMSEP value generated from the FiPLS method was less than that of the full spectra, which indicated that that the predictive ability of the model was enhanced by the FiPLS method. Moreover, a relatively high RPD indicated that the models were robust and very effective in predicting the chemical composition. Additionally, the RPD value calculated using the FiPLS method was greater than five, demonstrating the robustness and power of the calibration models [24]. When the exact positions of the NIR variables selected by the FiPLS algorithm were displayed, they corresponded very closely with the regions of 9034.87–8940.37 cm<sup>-1</sup>, 8649.17–8554.68 cm<sup>-1</sup>, 7974.21–7686.87 cm<sup>-1</sup>, 7395.67–7301.18 cm<sup>-1</sup> and 5949.32–5758.4 cm<sup>-1</sup>. The loadings of the first three latent variables are shown in **Figure 4**.



**Figure 4.** Loadings from the PCA of the NIR spectra obtained during the ethanol precipitation process.

The peaks in loading 1 between 5600 and 6000 cm<sup>-1</sup> can be attributed to the OH bond. However, because of the complexity of the NIR spectra it could not be conclusively determined whether the absorption was a result of the presence of ethanol in the system. Therefore, as shown in **Figure 5**, the calibration model developed based on the preprocessed spectra and the FiPLS variable selection method provided better results when compared with the other models. The values of R<sup>2</sup>, RMSECV and RMSEP were 0.974, 1.409 g/l and 1.105 g/l, respectively, and the number of latent values was three, which avoided over-fitting in multivariate calibration. However, the RMSECV was higher than the RMSEP calculated from the FiPLS model. This could be due to the presence of some samples that had a significant influence on the calibration. When those samples were eliminated, the predictability suffered. That also led to an increase in the RMSECV. The detailed reason for this observation is still under investigation.



**Figure 5.** Comparison of reference values measured by UV analysis vs. predicted values obtained from PLS model established based on pre-treated spectra and FiPLS variable selection method.

## CONCLUSIONS

Based on the results obtained in this study, the heparin ethanol precipitation process could be qualitatively and quantitatively monitored using NIR spectroscopy, combined with chemometric methods. First, the score scatter plot for the first two principal components obtained by the PCA method could be used to represent the stability and the characteristic trajectory of heparin ethanol precipitation process qualitatively. Next, the PLS calibration models were established based on the pre-treated spectra and the variable selection methods for quantitative analysis of the heparin content in the supernatants. This study shows a great potential toward providing a solution for rapid, real-time monitoring of the heparin ethanol precipitation process. Indeed, this study has the potential to provide a feasibility reference for in-line monitoring of ethanol precipitation of the GAGs.

## ACKNOWLEDGEMENTS

We are grateful to the financial support of the 863 program (Hi-tech research and development program of China) under contract number 2012AA021505 and the Science and Technology Development Program of Shandong Province (2009GG10002081).

## REFERENCES

- Rathore A, et al. Process analytical technology (PAT) for biopharmaceutical products”, *Anal Bioanal Chem.* 2010;398:137.
- Xiaoli C, et al. The On-line near Infrared Spectroscopy Process Analytical Technique and Its Application's, *Modern Scientific Instruments*, 2004;2:3.
- Center for Drug Evaluation and Research, Food and Drug Administration, Guidance for Industry, PAT: A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. 2004.
- Hinz DC. Process analytical technologies in the pharmaceutical industry: the FDA's PAT initiative. *Anal Bioanal Chem.* 2006;384:1036-1042.
- Challa S and Potumarthi R. Chemometrics-Based Process Analytical Technology (PAT) Tools: Applications and Adaptation in Pharmaceutical and Biopharmaceutical Industries, *Appl Biochem Biotechnol.* 2013;169:66.
- Ozaki Y. Near-infrared spectroscopy its versatility in analytical chemistry. *Anal Sci.* 2012;28:545-563.
- Roggo Y, et al. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies.
- Casu B. Structure of heparin and heparin fragments, *Ann N Y Acad Sci*, 1989;556.
- Weitz DS and Weitz JI. Update on heparin: what do we need to know? *J Thromb Thrombolysis.* 2010;29: 199-207.
- Huang H and Qu H. In-line monitoring of alcohol precipitation by near-infrared spectroscopy in conjunction with multivariate batch modeling, *Anal Chim Acta.* 2011;707:47.
- Parris J, et al. Monitoring API drying operations with NIR, *Spectroscopy*, 2005;20-34.
- Pollanen k, et al. Dynamic PCA-based MSPC charts for nucleation prediction in batch cooling crystallization processes, *Chemometr Intell Lab Syst.* 2006;84(1-2):126.
- Sekulic SS, et al. On-line monitoring of powder blend homogeneity by near-infrared spectroscopy. *Anal Chem.* 1996;68:509-513.
- Noda I. Recent advancement in the field of two-dimensional correlation spectroscopy, *J Mol Struct.* 2008;2:883.
- Sun D. Fast analytical method on content and transfer rate of marker component in ethanol precipitation of cinobufacini by near infrared spectroscopy”, *Zhongguo Zhong yao za zhi (Chinese journal)*, 2011;36:2479.
- Wu Z, et al. Development and validation of NIR model using low-concentration calibration range: rapid analysis of *Lonicera japonica* solution in ethanol precipitation process, *Anal Methods.* 2012;4 1084.
- Wu Z, et al. validation of a NIR quantification method for the determination of chlorogenic acid in *Lonicera japonica* solution in ethanol precipitation process. *J Pharm Biomed Anal.* 2012;62:1-6.
- Jin Y, et al. Near infrared spectroscopy in combination with chemometrics as a process analytical technology (PAT) tool for on-line quantitative monitoring of alcohol precipitation. *J Pharm Biomed Anal.* 2013;77:32-39.
- Workman J and Weyer L, *Practical guide to interpretive near-infrared spectroscopy*, CRC press, 2007.
- Zang H, et al. Application of near-infrared spectroscopy combined with multivariate analysis in monitoring of crude heparin purification process”, *Spectrochim Acta A Mol Biomol Spectrosc*, 2013;8:109.
- Rozenstein O, et al. Comparing the Effect of Pre-processing Transformations on Methods of Land-Use Classification Derived From Spectral Soil Measurements, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015;8:6

22. Xiaobo Z, et al. Variables selection methods in near-infrared spectroscopy. *Anal Chim Acta*. 2010;667:14-32.
23. Catana C, Linear and nonlinear methods in modeling the aqueous solubility of organic compounds. *J Chem Inf Model*. 2005;45:170-176.
24. Ferreira DS, et al. Fourier transforms near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean (*Glycine max* (L.) Merril) composition, *Food Research International*. 2013;1:51.