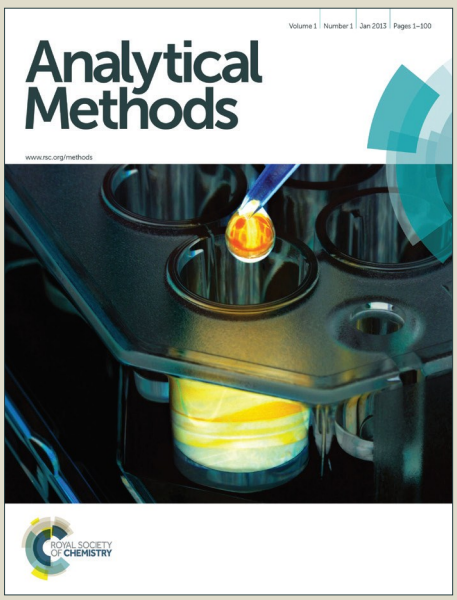
Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Rapid Screening of Critical Process Parameters Based on Near Infrared Spectroscopy: a Case Study of Ethanol Precipitation Process

Binjun Yan, Haibin Qu*

Pharmaceutical Informatics Institute, Zhejiang University, Hangzhou, 310058, China.

* Correspondence to: Haibin Qu

E-mail: quhb@zju.edu.cn

Tel: +86-571-88208428

Fax: +86-571-88208428

Abstract:

In process development, design of experiments (DoE) approach has been widely used for screening the key influential factors and obtaining optimal operating conditions. However, measuring the quality attributes of many products produced in the DoE is time and cost intensive. Therefore, an innovative method using near infrared spectroscopy (NIRS) for rapid screening of CPPs was proposed in this work, taking the ethanol precipitation process of Danhong Injection as a case study. In the work, seven process parameters were studied with a fractional factorial DoE. The supernatants produced were analyzed by NIRS. The spectra were preprocessed and then decomposed by principal component analysis into a few scores and loadings. The scores were taken as the response variables in analysis of variance (ANOVA) to

identify the CPPs. The result showed that water content in the concentrate (p<0.0001), ethanol concentration (p<0.0001) and ethanol consumption (p<0.0001) were the three CPPs. To validate the proposed method, the intermediates were analyzed by HPLC. The concentrations of two representative compounds were taken as the response variables in ANOVA and the same three CPPs were identified. The proposed method was successfully applied and the advantages of integrating spectroscopic techniques into screening DoE was demonstrated.

Keywords: Design of experiments; Near infrared spectroscopy; Critical process parameter; Rapid screening; Ethanol precipitation.

1. Introduction

The concept of quality by design (QbD)[1,2] is widely accepted by the pharmaceutical manufacturing industry and extensively used in the process development of pharmaceuticals. In the QbD implementation, design of experiments (DoE) approach[3-8] has been widely used to build the relationship between material attributes, process parameters and quality attributes for screening the most influential factors and obtaining optimal operating conditions. Usually a screening DoE is conducted before the optimization DoE to narrow down the extensive list of possible impacting factors to a few critical process parameters (CPPs)[9]. However, since the number of DoE experiments increases rapidly with the number of factors, it could be time and cost intensive to measure the quality attributes of many products or intermediates with traditional methods like HPLC. Therefore, a method that enables rapid screening of CPPs is desired.

Moreover, in some traditional industries like foods and herbal drugs, many manufacturing processes have not been well studied and the critical quality attributes of the intermediates are not well-defined or can not be easily measured. As the compositions of the intermediates are complicated, characterizing them by determining only a few compounds is insufficient.

Near infrared spectroscopy (NIRS) is a real-time analysis technique which allows nondestructive measurement without sample preparation[10], and widely used for rapid identification, on-line determination and process monitoring[11,12]. The spectrum can provide versatile information about the sample, including physical properties and chemical composition. Therefore, NIRS can be used for fast and comprehensive

characterization of complicated samples[13-16], which may provide a good solution for the two issues discussed above. In this work, an innovative method for rapid screening of CPPs based on NIRS was proposed and applied to the ethanol precipitation process.

Ethanol precipitation is a widely used purification process in the manufacturing of herbal drugs[17-19]. In the process, with the addition of ethanol, large amounts of impurities of poor solubility in ethanol are precipitated and thus removed. Many process parameters are considered as possible impacting factors for ethanol precipitation[18,20] and the CPPs should be identified using screening DoE before process optimization.

Danhong Injection is an herbal drug produced from Danshen (Radix *Salvia Miltiorrhizae*) and Honghua (Flos *Carthami*), which is widely used for clinical treatment in China. The main bioactive compounds in Danhong Injection are phenolic acids and flavones[21,22], and shown to be beneficial to the cardio-cerebral vascular system of human body. The ethanol precipitation process is an important purification process in the production of Danhong Injection, and in the work, this process was taken as an application case of the proposed method. Seven process parameters were studied with a fractional factorial DoE. The intermediates produced during the DoE experiments were analyzed by NIRS and the spectra were decomposed by principal component analysis (PCA) into a few scores and loadings. The scores were taken as the response variables in analysis of variance (ANOVA) and regressed on the process parameters to identify the CPPs. To validate the method proposed, the intermediates

were analyzed by HPLC and the concentrations of two representative compounds were taken as the response variables in ANOVA. This work has shown the applicability and advantages of integrating spectroscopic techniques into DoE for rapid screening of CPPs.

2. Experimental

2.1. Materials

Danshen was purchased from Nepstar Drugstore (Hangzhou, China) and Honghua was purchased from Daily Healthy Drugstore (Hangzhou, China). The raw materials of Danshen and Honghua (3:1, w/w) were extracted together with boiling water at a temperature of 100 ^{O}C. Then the extract is concentrated under reduced pressure at a temperature of 70 ^{O}C to obtain the concentrate. The concentrate were adjusted to different water content (0.44, 0.49 and 0.54 g/g) by mixing them with certain amounts of deionized water. Standard substance of danshensu (DSS) was purchased from Shanghai Winherb Medical Technology Co., Ltd. (Shanghai, China). Standard substance of hydroxysafflor yellow A (HSYA) was purchased from Aladdin Reagent Corporation (Shanghai, China). The analytical grade ethanol was purchased from Lingfeng Chemical Reagent Co., Ltd. (Shanghai, China). Deionized water was produced using a Milli-Q academic water purification system (Millipore Corp., Milford, MA, USA). The HPLC-grade acetonitrile were purchased from Merck (Darmstadt, Germany). The HPLC-grade formic acid was purchased from ROE Scientific Inc. (Newark, DE, USA). The HPLC-grade ammonium formate was

purchased from Alfa Aesar China Co., Ltd (Tianjin, China).

2.2. Design of experiments

The DoE was carried out as reported in previous publication [20]. A fractional factorial DoE with three center points (Table 1) was conducted using the Design Expert 8.0 software (State-Ease Inc., Minneapolis, MN, USA). Seven process parameters were studied, including water content in the concentrate (g/g), ethanol concentration (v/v), ratio of ethanol consumption to concentrate (ml/g), flow rate of the ethanol solution (ml/min), stirring rate, refrigerant temperature (^{O}C) and refrigerant time (h).

Table 1 Fractional factorial design of experiments

Standard order	Water content (g/g)	Ethanol concentration (v/v)	Ethanol consumption (ml/g)	Flow rate (ml/min)	Stirring rate	Refrigerant temperature (^{O}C)	Refrigerant time (h)
1	0.44	0.92	1.7	15	slow	5	20
2	0.54	0.92	1.7	15	fast	5	40
3	0.44	0.98	1.7	15	fast	15	20
4	0.54	0.98	1.7	15	slow	15	40
5	0.44	0.92	2.5	15	fast	15	40
6	0.54	0.92	2.5	15	slow	15	20
7	0.44	0.98	2.5	15	slow	5	40
8	0.54	0.98	2.5	15	fast	5	20
9	0.44	0.92	1.7	25	slow	15	40
10	0.54	0.92	1.7	25	fast	15	20
11	0.44	0.98	1.7	25	fast	5	40
12	0.54	0.98	1.7	25	slow	5	20
13	0.44	0.92	2.5	25	fast	5	20
14	0.54	0.92	2.5	25	slow	5	40
15	0.44	0.98	2.5	25	slow	15	20
16	0.54	0.98	2.5	25	fast	15	40
17	0.49	0.95	2.1	20	medium	10	30
18	0.49	0.95	2.1	20	medium	10	30
19	0.49	0.95	2.1	20	medium	10	30

2.3. Ethanol precipitation

Ethanol precipitation experiments were conducted as Table 1 in random order. During the process, ethanol solution was pumped into 20 g of concentrate at a designed flow rate under continuous magnetic stirring. The ethanol addition was stopped when ethanol consumption reached the designed amount. Then the mixture was refrigerated at a designed temperature for a designed time. Finally, the supernatant was collected for NIRS and HPLC analysis.

2.4. Near infrared spectroscopy

The NIR spectra of the 19 supernatants were collected with an Antaris MX Fourier-Transform NIR spectrometer (Thermo Fisher Scientific Inc., Madison, WI, USA). Each spectrum was obtained by averaging 32 scans over the spectral region 4000 ~ 10,000 cm^{-1} at 4 cm^{-1} interval in transmission mode with optical length of 2 mm.

2.5. HPLC analysis

The concentrations of DSS and HSYA (two representative active compounds in Danshen and Honghua) in the supernatants, reported in previous publication [20], were used to validate the result of the CPPs screening method. And These concentrations were determined by HPLC according to the method in the literature[23] with an HPLC system HP1100 series (Agilent Technologies Inc., Waldbronn, Germany).

2.6. Data analysis

Two strong absorption bands (4335 ~ 4435 cm^{-1} and 5110 ~ 5270 cm^{-1}) beyond the detection range of the spectrometer were discarded. Then the spectra were pretreated by Savitzky–Golay first-order derivative to eliminate the baseline drift caused by scatter effects[24]. As attribution of all the absorption bands was not feasible and the importance of each wavelength was not easy to be determined, all the variables were considered of equal importance and standardized to zero mean and unit variance. The pretreated spectra of the 19 supernatants were decomposed by PCA into a few scores and loadings. The scores were taken as the response variables in ANOVA and regressed on the process parameters to identify the CPPs. To validate the method proposed, the concentrations of DSS and HSYA determined by HPLC were also taken as the response variables in ANOVA. The data analysis was conducted in the SIMCA-P+ 13.0 software (Umetrics Inc., Umeå, Sweden) and the Design Expert 8.0 software.

3. Results and Discussion

3.1. Near infrared spectra and principal component analysis

The NIR spectra of the supernatants are shown in Fig. 1. As almost all kinds of organic compounds have their absorption peaks in the NIR range, these spectra provided a comprehensive characterizations of the complex compositions.

PCA was conducted on the 19 NIR spectra and two significant principal components (PCs) were retained. PC1 explained 81.5 % of the total variations in the

spectra and PC2 explained 7.0 %. The scores plot (t_1-t_2) of the 19 supernatants is shown in Fig.2 and the three center points of DoE are closely distributed. DoE 2 and DoE 10 have small t_1 scores and from Table 1 we see that these two supernatants were produced with concentrates of high water content, low ethanol concentration and low ethanol consumption. On the contrary, DoE 7 and DoE 15 have large t_1 scores, which were produced with concentrates of low water content, high ethanol concentration and high ethanol consumption. It is shown that the PCs can reflect the impacts of process parameters on the quality of the supernatants.

The loadings plot is shown in Fig.3, and the absorbance peak near 4800 cm^{-1} (Fig.1) has a small value in loading *p1*. DoE 2 and DoE 10 have small t_1 scores, which is caused by their large absorbance near 4800 cm^{-1}. Therefore, these two samples may have higher contents of some compounds with absorbance peak near 4800 cm^{-1}. However, as the compositions of herbal intermediates are complicated, further interpretation of the NIR spectra is not available.

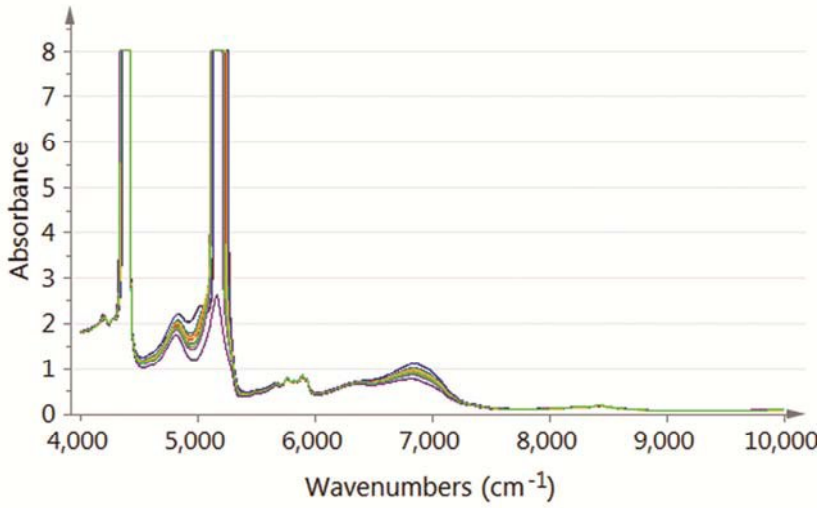


Fig.1 NIR spectra of the supernatants

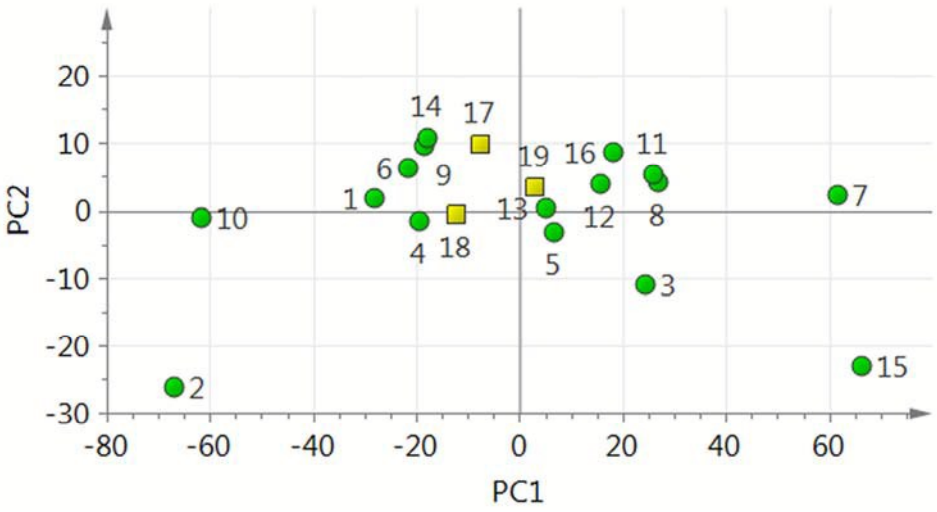


Fig.2 PCA scores plot (t_1-t_2) of the 19 supernatants (marked by their standard orders). The three center points of DoE are represented by yellow squares and the rest are represented by green circles

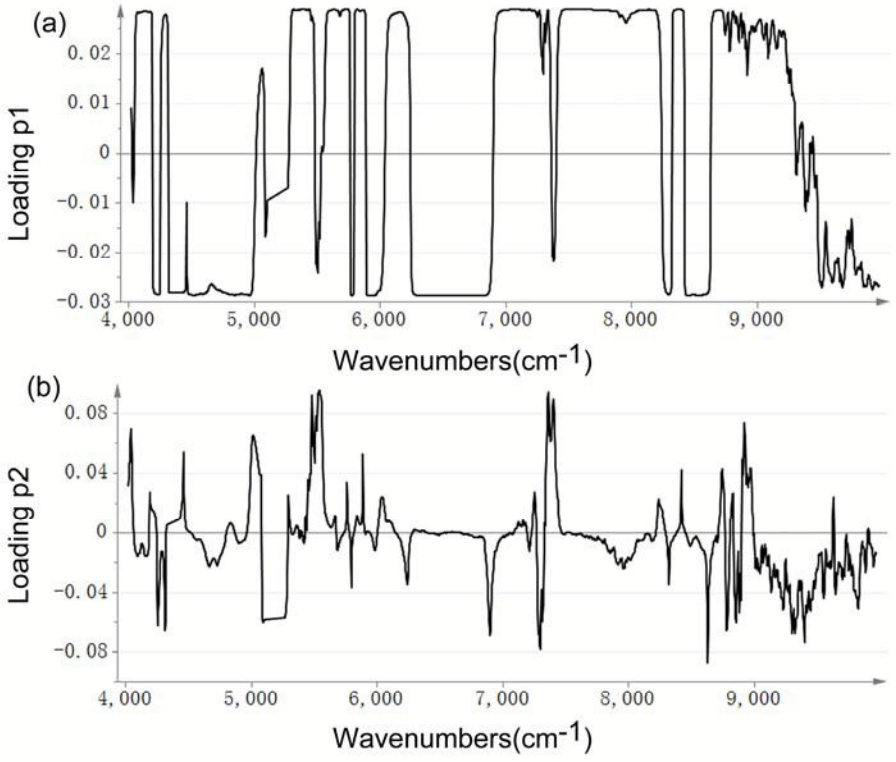


Fig.3 PCA loadings plot. (a) Loading *p1*. (b) Loading *p2*.

3.2. Identification of critical process parameters

To identify the CPPs, the t_1 and t_2 scores were taken as the response variables in ANOVA and regressed on the seven process parameters (coded values). From Table 2, we know that water content, ethanol concentration and ethanol consumption were three significant factors for score t_1, and a large t_1 was the result of low water content, high ethanol concentration and high ethanol consumption. No significant process parameter was found for score t_2. Therefore, water content, ethanol concentration and ethanol consumption were identified as the CPPs for the ethanol precipitation process.

Table 2 Result of analysis of variance on scores t_1 and t_2

Factors	t_1		t_2	
	Regression coefficient	*p*-value (Prob>*F*)	Regression coefficient	*p*-value (Prob>*F*)
Water content	-16.9	<0.0001[a]	1.5	0.6325
Ethanol concentration	26.4	<0.0001[a]	-0.6	0.8336
Ethanol consumption	17.2	<0.0001[a]	1.6	0.6059
Flow rate	3.1	0.1504	2.6	0.3959
Stirring rate	-3.7	0.0909	-2.1	0.5022
Refrigerant temperature	-1.7	0.4046	-1.1	0.7259
Refrigerant time	-2.3	0.2716	1.5	0.6132

[a] Significant at 0.05 level.

3.3. Validation of the proposed method

DSS is an important phenolic acid in Danshen and HSYA is an important flavone in Honghua. To validate the proposed method, the supernatants were analyzed by HPLC and the concentrations of DSS and HSYA (Table 3) were taken as the response variables in ANOVA (Table 4). For DSS and HSYA, the same three CPPs were found, i.e., water content, ethanol concentration and ethanol consumption.

Two advantages of the CPPs screening method based on NIRS should be noted. First is the significantly reduced time consumption. Less than one minute is needed for analyzing one sample with NIRS, and tens of minutes are required in HPLC analysis. Therefore, much time can be saved, considering usually a large number of samples to be analyzed during DoE experiments. Second, NIRS can give a more comprehensive characterization of the intermediate, compared with the limited number of compounds determined by HPLC, which may be more suitable for complex samples like herbal drugs.

Table 3 Concentrations of DSS and HSYA in the supernatants

Standard order	DSS (mg/g)	HSYA (mg/g)
1	0.333	0.636
2	0.370	0.747
3	0.306	0.518
4	0.359	0.660
5	0.242	0.464
6	0.264	0.546
7	0.208	0.344
8	0.226	0.443
9	0.344	0.652
10	0.370	0.748
11	0.296	0.500
12	0.256	0.486
13	0.227	0.451
14	0.259	0.530
15	0.193	0.326
16	0.226	0.457
17	0.276	0.539
18	0.289	0.556
19	0.281	0.535

Table 4 Result of analysis of variance on the concentrations of DSS and HSYA

Factors	DSS		HSYA	
	Regression coefficient	*p*-value (Prob>*F*)	Regression coefficient	*p*-value (Prob>*F*)
Constant	0.280	/	0.534	/
Water content	0.011	0.0195[a]	0.046	0.0001[a]
Ethanol concentration	-0.021	0.0003[a]	-0.065	< 0.0001[a]
Ethanol consumption	-0.049	< 0.0001[a]	-0.087	< 0.0001[a]
Flow rate	-0.009	0.0624	-0.013	0.1180
Stirring rate	0.003	0.4943	0.009	0.2557
Refrigerant temperature	0.008	0.0771	0.015	0.0840
Refrigerant time	0.008	0.0761	0.012	0.1342

[a] Significant at 0.05 level.

3.4. Impacts of critical process parameters

As NIR spectra are difficult to be interpreted, the proposed method identified the CPPs without showing the specific impacts of the CPPs. This shortcoming is not critical, because usually an optimization DoE will be carried out following the screening DoE to better reveal the impacts of CPPs. However, if the impacts of the CPPs are required in the screening DoE stage, it would not be necessary to analyze all the supernatants produced in the DoE, as the CPPs have been identified already. To reduce the time consuming HPLC analysis, we can only analyze the supernatants produced in DoE 1 to DoE 8. These eight experiments are chosen here because they compose a full factorial design for the three CPPs when the differences in the levels of the four insignificant process parameters are neglected. The quantitative relationship established from the eight experiments are as follows:

$$Y_{DSS} = 0.288 + 0.016\,a - 0.014\,b - 0.054\,c \quad (1)$$

$$Y_{HSYA} = 0.545 + 0.054\,a - 0.054\,b - 0.096\,c \quad (2)$$

where Y_{DSS} and Y_{HSYA} are the concentrations of DSS and HSYA in the supernatant, and

a, *b*, *c* are the coded values of the three CPPs (water content, ethanol concentration and ethanol consumption) respectively. The regression coefficients in Equations (1) and (2) are very close to those in Table 4, considering that only a small number of experiments are used.

The CPPs screening method based on NIRS is shown to be capable of reducing the number of samples subject to other time consuming analysis. This method may be applied in the industrial production situation, where a large number of samples can be collected but only a very small portion can be analyzed by time and/or cost intensive methods. The proposed method can identify those samples of large differences that need further analysis.

4. Conclusions

In this work, a method for rapid screening of CPPs was proposed and applied in the case study of the ethanol precipitation process of Danhong Injection. NIRS was used for fast and comprehensive characterization of the supernatants, and the three CPPs were identified rapidly. The method can, in some extent, replace other time and cost intensive measurements, which shows the advantages of integrating spectroscopic techniques into screening DoE and is worthy of application in process development.

Acknowledgments

This work was financially supported by National Natural Science Foundation (No.

81273992). The authors would like to thank Zhengtai Guo for his valuable assistance in experimental works.

References

1 ICH, ICH harmonised tripartite guideline: Pharmaceutical development Q8(R2), 2009. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf, (accessed 2012-04-28).

2 L. X. Yu, G. Amidon, M. A. Khan, S. W. Hoag, J. Polli, G. K. Raju and J. Woodcock, *AAPS Journal*, 2014, **16**, 771-783.

3 N. Suriyachai, K. Weerasaia, N. Laosiripojana, V. Champreda and P. Unrean, *Bioresour. Technol.*, 2013, **142**, 171-178.

4 G. Davidov-Pardo, I. Arozarena and M. R. Marin-Arroyo, *Food Bioprocess Technol.*, 2013, **6**, 941-951.

5 A. Alberti, A. A. F. Zielinski, D. M. Zardo, I. M. Demiate, A. Nogueira and L. I. Mafra, *Food Chem.*, 2014, **149**, 151-158.

6 A. K. Das, V. Mandal and S. C. Mandal, *Phytochem. Anal*, 2013, **24**, 230-247.

7 H. Q. Wu, M. Tawakkul, M. White and M. A. Khan, *Int. J. Pharm.*, 2009, **372**, 39-48.

8 H. Q. Wu, M. White and M. A. Khan, *Int. J. Pharm.*, 2011, **405**, 63-78.

9 D. C. Montgomery, *Design and analysis of experiments*, Wiley, New York, 6th edn., 2004.

10 T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J. P. Remon and C. Vervaet, *Int. J. Pharm.*, 2011, **417**, 32-47.

11 M. L. Zhang, G. P. Sheng, Y. Mu, W. H. Li, H. Q. Yu, H. Harada and Y. Y. Li, *Water Res.*, 2009, **43**, 1823-1830.

12 Z. Wu, E. Xu, J. Long, Y. Zhang, F. Wang, X. Xu, Z. Jin and A. Jiao, *Food Control*, 2015, **50**, 405-412.

13 C. Hue, Z. Gunata, A. Bergounhou, S. Assemat, R. Boulanger, F. X. Sauvage and F. Davrieux, *Food Chem.*, 2014, **148**, 240-245.

14 Z. S. Wu, Y. F. Peng, W. Chen, B. Xu, Q. Ma, X. Y. Shi and Y. J. Qiao, *Bioresour. Technol.*, 2013, **137**, 394-399.

15 S. Tankeu, I. Vermaak, G. Kamatou and A. Viljoen, *Phytochem. Anal*, 2014, **25**, 81-88.

16 H. Morita, T. Hasunuma, M. Vassileva, A. Kondo and R. Tsenkova, *Anal. Methods*, 2014, **6**, 6628-6634.

17 H. X. Huang and H. B. Qu, *Anal. Chim. Acta*, 2011, **707**, 47-56.

18 L. Zhang, B. Yan, X. Gong, L. Yu and H. Qu, *AAPS Pharmscitech*, 2013, **14**, 277-286.

19 B. Yan, Y. Li, Z. Guo and H. Qu, *Phytochem. Anal*, 2014, **25**, 59-65.

20 B. J. Yan, Z. T. Guo, H. B. Qu, B. C. Zhao and T. Zhao, *China J. Chin. Mater. Med.*, 2013, **38**, 1672-1675.

21 C. E. He, J. H. Wei, Y. Jin and S. L. Chen, *Ind. Crop. Prod.*, 2010, **32**, 313-317.

22 G. Zhang, M. L. Guo, R. P. Li, Y. Li, H. M. Zhang and Z. W. Su, *Chem. Nat. Compd.*, 2009, **45**, 398-401.

23 Y. Li, Z. T. Guo, X. C. Gong and H. B. Qu, *China J. Chin. Mater. Med.*, 2013, **38**, 1653-1656.

24 L. Eriksson, E. Johansson, N. Kettaneh-Wold and S. Wold, *Multi-and megavariate data analysis: Principles and applications*, Umetrics, Umea, Sweden, 2001.