Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. 78 (4) 555–564 (2013)
JSCS–4438
UDC 637.12+520.8.07+546.226+547.867.2:543.4/.5
Original scientific paper

Genetic algorithm-based wavelength selection in multicomponent spectrophotometric determinations by partial least square regression: application to a sulfamethoxazole and trimethoprim mixture in bovine milk

MOHAMMAD HADI GIVIANRAD*, MOHAMMAD SABER-TEHRANI and SABER ZARIN

Department of Chemistry, Science and Research Branch, Islamic Azad University, Tehran, Iran

(Received 3 March, revised 26 May 2012)

Abstract: The simultaneous determination of sulfamethoxazole (SMX) and trimethoprim (TMP) mixtures in bovine milk by spectrophotometric method is, due to spectral interferences, a difficult problem in analytical chemistry. By means of multivariate calibration methods, such as partial least square (PLS) regression, it is possible to obtain a model adjusted to the concentration values of the mixtures used in the calibration range. A genetic algorithm (GA) is a suitable method for selecting the wavelengths for PLS calibration of mixtures with almost identical spectra without the loss of prediction capacity using a spectrophotometric method. In this study, a calibration model based on the absorption spectra in the 200–400 nm range for 25 different mixtures of SMX and TMP. Calibration matrices were formed from samples containing 0.25–20 and 0.3–21 μg mL^{-1} for SMX and TMP, at pH 10, respectively. The root mean squared error of deviation *(*RMSED*)* for SMX and TMP with PLS and genetic algorithm partial least square (GAPLS) were 0.242 and 0.066 μg mL^{-1}, and 0.074 and 0.027 μg mL^{-1}, respectively. This procedure allowed the simultaneous determination of SMX and TMP in synthetic and real samples and good reliability of the determination was proved.

Keywords: sulfamethoxazole; trimethoprim; partial least square; simultaneous determination; bovine milk.

INTRODUCTION

Sulfamethoxazole, 4-amino-*N*-(5-methylisoxazol-3-yl)-benzenesulfonamide, (SMX) is a sulfonamide antibiotic that is a highly effective chemotherapeutic agent, which competitively inhibits the bacterial enzyme dihydropteroate synthetase.[1,2] Trimethoprim, 5-(3,4,5-trimethoxybenzyl)pyrimidine-2,4-diamine,

* Corresponding author. E-mail: givianradh@yahoo.com
doi: 10.2298/JSC120303080G

(TMP), is one of the most widely used antibacterial additive. In addition, TMP is a dihydrofolat-reductase inhibitor.[3] A combination of TMP/SMX is an effective antimicrobial agent that is commonly used in dairy cattle for the treatment or prevention of respiratory infections and mastitis.[4] The use of this combination may lead to the presence of residual levels in milk and meat. Residues of this combination can cause several risks for human health, such as allergic reaction in hypersensitive individuals and induction of resistance of strains of pathogenic bacteria. Various methods have been published for the determination of SMX and TMP in milk and biological fluids.[5–14] Most of these methods employ separation methods, such as online solid phase extraction–liquid chromatography with UV (SPE-LC–UV) and mass spectrometry detection (SPE-LC–MS/MS),[3] a HPLC method using an on-line clean-up column coupled with amperometric detection employing a boron-doped diamond (BDD),[4] or micellar electrokinetic capillary chromatography.[6] The majorities of these methods are expensive and time consuming. Besides these methods, UV–Vis spectrophotometry could be considered a rapid method. Moreover, a flow-through optosensor combined with photochemically induced fluorescence was used for the simultaneous determination of binary mixtures of sulfonamides in pharmaceuticals, milk and urine.[15] However, spectrophotometric methods are sensitive and a hand scanner is inexpensive, available in most work offices and is easy to operate by non-skilled users. The simultaneous determination of SMX and TMP in milk by this method could be a difficult task, because their absorption spectrums overlap in this region and superimposed curves are not suitable for quantitative evaluation. Hence, multivariate calibration methods are playing an important role in the multicomponent analysis of mixtures SMX/TMP by UV–Vis spectrophotometry.[8]

In the present study, a multivariate calibration method, *i.e.*, partial least square (PLS) regression, was applied to the simultaneous spectrophotometric determination of SMX and TMP in milk. The major advantage of multi-component analysis using PLS are speed and expense. A genetic algorithm (GA) is a very useful technique with variable selection problems, because the relationship between the presence/absence of the variables in a calibration model and the prediction ability of the model, specifically for PLS models, is very complex and the mathematical properties are unknown.

EXPERIMENTAL

Apparatus and software

The spectra were obtained with an Agilent UV–Vis spectrophotometer, PerkinElmer (Lambda 25) in the wavelength range of 200 to 400 nm. A 1-cm path length quartz cell was used. The spectra were blank-corrected. The pH values of the solutions were measured with a Metrohm, model 827 instrument, using a combined glass electrode. All the absorbance spectra were digitized and stored at wavelengths from 200 to 350 nm in steps of 1 nm and transferred (in ASCII format) to a Pentium IV computer with Windows XP operating system. Data

processing (GAPLS) was performed by a laboratory-written program in MATLAB, version 7.8.0 (R2009a).

PLS and GA

The application of quantitative chemometrics methods, particularly partial least squares (PLS), to multivariate chemical data is becoming more widespread, owing to the availability of digitized spectroscopic data and commercial software for laboratory computers. The advantage of multicomponent analysis along with the partial least squares method in mixtures is the demonstration of fast separation steps. The theory and application of partial least squares (PLS) in spectrometry were previously discussed by several workers.[16,17] Genetic algorithm (GA) is a very useful technique in variable selection problems, because of the relationship between the presence/absence of the variables in the calibration model and the prediction ability of the model. The algorithm used in this paper is an evolution of the algorithm described in the literature,[18] the parameters of which are tabulated in Table I.

TABLE I. Parameters of the genetic algorithms

Parameter	Value
Population size	30 Chromosomes; on average, 5 variables per chromosome in the original population
Regression method	PLS
Response	Cross-validated percent explained variance (5 deletion groups; the number of components is determined by cross-validation)
Maximum number of variables selected in the same chromosome	30
Probability of mutation	1 %
Probability of cross-over	50 %
Maximum number of components	Determined by cross-validation on the model containing all the variables (not more than 15)
Number of runs	100, backward elimination after every 100th evaluation and at the end (if the number of evaluation is not a multiple of 100)
Window size for smoothing	3

Reagents and solutions

All experiments were performed with analytical grade reagents and used directly without further purification. Methanol was obtained from Merck. Doubly distilled deionized water was used to prepare the reagent solutions. The pharmaceutical substances (SMX and TMP) were of analytical grade and were obtained from Sigma. The stock solutions (50 ppm) of SMX and TMP were prepared by dissolving the sample powders in NaOH 0.001 M and deionized water, respectively. Universal buffer solutions (pH 2–12) were prepared by mixing 50 mM phosphoric acid, 50 mM boric acid, 50 mM acetic acid and a sufficient amount of CO_2-free NaOH solution.

One-component calibration

For each component, the linear dynamic concentration range was found by one-component calibration. For the preparation of each solution, different volumes of SMX and TMP stock solutions (50 μg mL^{-1}) were added to 2 ml universal buffer (pH 10) in a 10 ml volu-

metric flask and then diluted to the mark with deionized water. The concentration ranges of SMX and TMP for the construction of the calibration graphs were 5–40 μg mL^{-1}. The absorbance spectra were achieved over the 200–400 nm spectral range *vs.* the solvent blank. The linear dynamic range for SMX (0.25–20 μg mL^{-1}) and TMP (0.3–21 μg mL^{-1}) was determined by regression of the absorbance at the corresponding λ_{max} *vs.* concentration.

Binary standard solution

Two set of solutions (calibration and prediction sets) were prepared (Table II). The calibration set contained 25 standard mixtures, and 6 mixtures were used as the validation set. Concentrations of SMX and TMP in the standard mixtures were 1–15 μg mL^{-1} and 0.2–8.0 μg mL^{-1}, respectively. For preparation of each solution, the required volumes of stock solution and 2 mL universal buffer were added to 10 mL volumetric flask and diluted to the mark with 0.001 M NaOH. Then absorbance spectra of the mixture were recorded *vs.* the blank solvent in the wavelength range of 200–400 nm (at 1.0 nm intervals).

TABLE II. Concentration data of the different mixtures used in the calibration (M1–M25) and the prediction (M26–M31) set for the determination of SMX and TMP

Mixture	SMX[a]	TMP[b]	Mixture	SMX	TMP	Mixture	SMX	TMP
M1	1.0	0.2	M14	8.0	6.0	M27	5.0	1.0
M2	1.0	1.6	M15	8.0	8.0	M28	6.0	1.6
M3	1.0	3.0	M16	11.0	0.2	M29	8.0	5.0
M4	1.0	6.0	M17	11.0	1.6	M30	10.0	6.0
M5	1.0	8.0	M18	11.0	3.0	M31	15.0	2.4
M6	5.0	0.2	M19	11.0	6.0			
M7	5.0	1.6	M20	11.0	8.0			
M8	5.0	3.0	M21	15.0	0.2			
M9	5.0	6.0	M22	15.0	1.6			
M10	5.0	8.0	M23	15.0	3.0			
M11	8.0	0.2	M24	15.0	6.0			
M12	8.0	1.6	M25	15.0	8.0			
M13	8.0	3.0	M26	4.0	0.8			

[a]Concentrations of sulfamethoxazole in μg mL^{-1}; [b]concentrations of trimethoprim in μg mL^{-1}

Preparation of real samples

In this study, milk was selected as a real sample. In order to determine concentration of SMX and TMP in milk, an extraction method was applied. In this case, the binary standard solutions were prepared in skimmed milk. The milk samples were filtered through a filter paper to remove solid particles. Then 10 ml of filtered milk and 15 g of anhydrous sodium sulfate were transferred into a 50 mL polypropylene centrifuge tube, blended with 15 ml of ethyl acetate for 1 min using a high-speed blender, and then centrifuged for 2 min at 3000 rpm. The organic phase was collected into clean centrifuge tubes, and the same amount of organic phase was added to the remainder of the sample in the centrifuge tubes and the extraction process was repeated. The combined extracts were evaporated to dryness under reduced pressure at 40 °C. The residue was dissolved in 2 ml of 0.001M NaOH and used for spectrophotometric analysis. An average extraction recovery in the range of 92–98 % was achieved using this developed method. The blank sample was prepared in the same method as for the analytes except that no analyte was added to the milk.

RESULTS AND DISCUSSION

The absorbance spectra of 15 mg L^{-1} SMX and TMP under certain experimental condition were recorded.[8] It could be inferred that the spectra of SMX and TMP were highly overlapped. Therefore, these compounds cannot be determined in presence of each other and it is necessary to use a chemometrics method to solve this problem. This problem alleviated by use of multivariate calibration methods.

Effect of pH

The influence of the medium pH on the absorption spectra of the two compounds was studied over the range of 2 to 10. The optimum pH (pH 10) was selected at which value the minimum overlap between two species, higher selectivity and maximum sensitivity were evidenced as mentioned previously.[8]

Variable selection

In this work, the calibration set had 121 variables. GA was run for these 121 variables (in the 200–320 nm range), using a PLS regression method. The maximum number of factors which, determined by cross-validation on the model consisted of all the variables. To find optimum the wavelength set for the determination of SMX and TMP, the GA procedure was repeated 10 times. Finally, a wavelength was selected if the percent of selection for that variable exceeded a critical value. The wavelengths selected by GA were 277, 239, 240, 241, 242, 217, 216 and 215 nm for SMX and 238, 237, 236, 235, 234, 256, 255 and 312 nm. Selected variables were used for PLS modeling.

Data processing and PLS modeling

PLS and GAPLS regression was run on the calibration data, and the concentrations of the analytes were predicted at the optimum number of factors. The selection of the number of factors in factor analysis-based methods is very important for achieving the best prediction. The model refinement procedure used the predicted residual errors sum of squares (PRESS) of the leave-one-out cross-validation to select the optimal number of PLS factors. In the case of each number of factors used, given 25 calibration standards, n–1 solutions were used in the model-building step (calibration) and the resulting model was used to predict the concentration of the desired analyte in the solution. This procedure was repeated n times until the concentrations of the analyte in all solutions were predicted. When each number of factors was entered, *PRESS* was calculated by comparing the predicted concentration of compounds in each sample with a known concentration of compounds in the standard solutions, (Eq. (1)):

$$PRESS = \left[\sum_{i=1}^{N} \left(\hat{c}_i - c_j \right) \right]^2 \quad (1)$$

where N is the total number of calibration samples; c_j the reference concentration for the j^{th} sample; $\hat{c}_i$ represents the estimated concentration. The GAPLS and PLS *PRESS* values are minimum in a number of factors for TMP and SMX, respectively. Then these numbers of factors were selected as the optimum for the calibration models (Figs. 1 and 2).

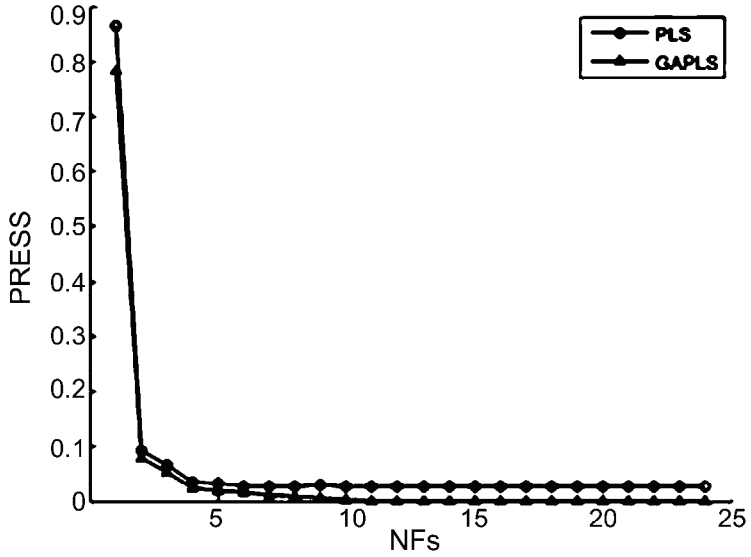

Fig. 1. Plots of *PRESS vs.* number of factors (NFs) by PLS and GAPLS for SMX.

Statistical parameters

For the constructed models, two parameters were selected to evaluate the prediction ability of the models for the simultaneous determination of TMP and SMX, *i.e.*, the root mean square error of deviation *(RMSED)* (Eq. (2)) and the relative error of prediction *(REP)* (Eq. (3)), calculated for each component as follows:

$$RMSED = \sqrt[2]{\frac{\sum_{i=1}^{n} \left(\hat{c}_i - c_i \right)^2}{n}} \quad (2)$$

$$REP(\%) = \frac{100}{\bar{c}} \sqrt[2]{\frac{1}{n} \sum_{i=1}^{n} (\hat{c}_i - c_i)^2} \quad (3)$$

where c_i is the true analyte concentration in the sample i, $\hat{c}_i$ represents the estimated analyte concentration in the sample i, $\overline{c}$ is the mean of the true concentration in the prediction set and n is the total number of samples used in the prediction set.

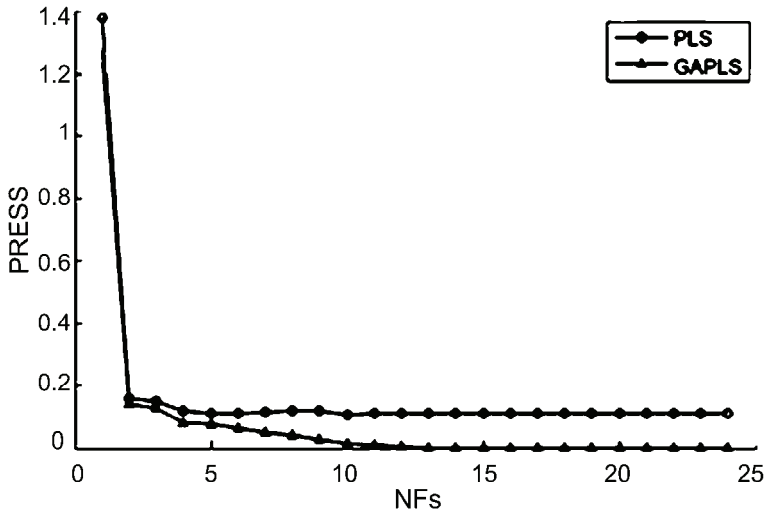

Fig. 2. Plots of *PRESS vs.* number of factors (NFs) by PLS and GAPLS for TMP.

Determination of SMX and TMP in the test set mixtures

The PLS and GAPLS methods were applied to the spectrophotometric concurrent monitoring of SMX and TMP. The predictive abilities of the methods were determined using 6 test mixtures (Tables III and IV). It can be seen that the values of *PRESS*, *RMSED* and *REP* (%) for the GAPLS method were lower than those for the PLS method.

TABLE III. Prediction results for SMX in the synthetic samples using the PLS and GAPLS methods

Sample	PLS[a]			GAPLS[b]	
	Added[c]	Predicted[c]	Recovery, %	Predicted	Recovery, %
1	4.00	3.88	97.00	3.98	99.50
2	5.00	5.35	107.00	5.11	102.20
3	6.00	6.15	102.50	6.03	100.50
4	8.00	8.19	102.37	7.90	98.75
5	10.00	10.39	103.90	9.98	99.80
6	15.00	14.94	99.60	15.05	100.33
RMSED[d]	–	0.242	–	0.066	–
REP[e] / %	–	3.023	–	0.829	–

[a]Partial least square regression; [b]genetic algorithm partial least square; [c]concentrations of sulfamethoxazole in μg mL^{-1}; [d]root mean squared error of deviation in μg mL^{-1}; [e]relative error of prediction

TABLE IV. Prediction results for TMP in the synthetic samples using the PLS and GAPLS methods

Sample	PLS[a]			GAPLS[b]	
	Added[c]	Predicted[c]	Recovery, %	Predicted	Recovery, %
1	0.80	0.74	92.50	0.81	101.25
2	1.00	1.03	103.00	1.01	101.00
3	1.60	1.51	94.37	1.59	99.37
4	5.00	5.03	100.60	4.98	99.6
5	6.00	6.13	102.16	6.06	101.00
6	2.40	2.35	97.91	2.39	99.58
RMSED[d]	–	0.074	–	0.027	–
REP[e] / %	–	2.648	–	0.964	–

[a]Partial least square regression; [b]genetic algorithm partial least square; [c]concentrations of trimethoprim in μg mL^{-1}; [d]root mean squared error of deviation in μg mL^{-1}; [e]relative error of prediction

Determination of SMX and TMP in bovine milk samples

The real samples in this study were bovine milk from the Kadkhoda dairy, Iran. Three spiking levels (3, 6 and 10 μg mL^{-1} for SMX, and 3, 5 and 8 μg mL^{-1} for TMP) in bovine milk samples were determined, based on the GAPLS and PLS methods. Accordingly, the accuracy of the GAPLS method was recognized as being better than that of the PLS method. Each measurement was repeated 6 times and the obtained mean recovery values are tabulated in Table V (found values calculated based on $\bar{X} \pm (tS.D.)/\sqrt{N}$, for N = 6 measurements and $t_{(N-1=5)}$ = = 2.571). The obtained values for the recovery rate varied between 98 and 108 % and the obtained values for the *RSD* ranged from 1.2 to 5.5 %. It could be observed that for all mixtures, the calculated values were in satisfactory agreement with those of the declared values. Consequently, the performance characteristics are recognized as acceptable and therefore the method is considered suitable for the intended purpose.

TABLE V. Recovery yields and relative standard deviations at different spiking levels in bovine milk samples

Sample	PLS[a]			GAPLS[b]	
	Spiking level[c]	Recovery, %	*RSD*[d] / %	Recovery, %	*RSD*[d] / %
SMX[e]	3.00	108.10	4.10	101.90	3.50
	6.00	100.40	3.80	99.60	1.20
	10.00	97.50	2.50	98.20	1.60
TMP[f]	3.00	84.40	4.80	108.40	4.70
	5.00	101.10	5.90	100.70	5.50
	8.00	105.40	4.80	101.80	3.90

[a]Partial least square regression; [b]genetic algorithm partial least square; [c] different spiking levels in bovine milk samples in μg mL^{-1}; [d]relative standard deviation; [e]sulfamethoxazole; [f]trimethoprim

CONCLUSIONS

This was the first spectrophotometric determination study that was accomplished on bovine milk in this manner. For overcoming the high spectral overlap observed between the absorption spectra, PLS was successfully applied to the absorption spectra for the concurrent analysis of SMX and TMP in their synthetic mixtures. In addition, the present study showed that GA could be regarded as a good method for feature selection in the spectral data sets. The proposed techniques are precise, inexpensive, simple to perform and do not require separation steps. The resulting data for the TMP and SMX mixture data set demonstrated that the predictive ability of the models, obtained with the wavelengths selected by the algorithm, was frequently enhanced. The good agreement between the obtained and spiked results demonstrated the utility of this procedure for the simultaneous detection of SMX and TMP in complex synthetic and bovine milk samples without tedious pretreatment. The superiority of the developed GAPLS method, which included the application of a chemometric method to milk samples and biological fluid samples, as well as the selection of wavelength by the GA method, over other official methods was demonstrated. In addition, it is very simple and accurate.

Acknowledgement. The authors are so grateful to the Laboratory Complex of I.A.U. for valuable technical assistance.

И З В О Д

ИЗБОР ТАЛАСНИХ ДУЖИНА БАЗИРАН НА ГЕНЕТИЧКОМ АЛГОРИТМУ У ВИШЕКОМПОНЕНТНОМ ОДРЕЂИВАЊУ МЕТОДОМ НАЈМАЊИХ КВАДРАТА: ПРИМЕНА НА СМЕШИ СУЛФАМЕТОКСАЗОЛА И ТРИМЕТОПРИМА У МЛЕКУ ГОВЕДА

MOHAMMAD HADI GIVIANRAD, MOHAMMAD SABER-TEHRANI И SABER ZARIN

Department of Chemistry, Science and Research Branch, Islamic Azad University, Tehran, Iran

Истовремено одређивање сулфаметоксазола и триметоприма у смеши, у бивољем млеку, сектрофотометријском методом представља велики проблем у аналитичкој хемији због спектралних сметњи. Коришћењем мултивариантне калибрационе методе, као што је парцијална регресија најмањих квадрата (PLS), могуће је добити модел прилагођен концентрацијама у смеши, коришћеним у калибрационом опсегу. Генетички алгоритам (GA) је погодна метода за селектовање таласних дужина за PLS калибрацију смеша са скоро идентичним спектрима, без губитка капацитета предвиђања, при коришћењу спектрофотометријске методе. У овом раду је калибрациони модел базиран на апсорпционом спектру у опсегу 200–400 nm за 25 различитих смеша сулфаметоксазола и триметоприма range. Калибрационе матрице су формиране за узорке који садрже 0,25–20 и 0,3–21 μg mL^{-1} сулфаметоксазола и триметоприма на pH 10. Вредности *RMSED* за сулфаметоксазол и триметоприм, базиране на PLS и GAPLS, износиле су 0,242 и 0,066 μg mL^{-1}, и 0,074 и 0,027 μg mL^{-1}. Ова процедура омогућава истовремено одређивање сулфаметоксазола и триметоприма у синтетичким и реалним узорцима и потврђена је доба поузданост методе.

(Примљено 3. марта, ревидирано 26 маја 2012)

REFERENCES

1. G. L. Mandel, W. A. Petri, *Goodman and Gilman's the Pharmacological Basis of Therapeutics*, 11th ed, McGraw Hill, New York, USA, 2006, p. 1057
2. AOAC, method 993.32, *Multiple sulfonamide residues in raw bovine milk, Handbook of analytical methods of analysis*, 16th ed., V. A. Arlington, USA, 1995, p. 13
3. L. S. Andrade , M. C. de Moraes, R. C. Rocha-Filho, O. Fatibello-Filho, Q. B. Cass, *Anal. Chim. Acta* **654** (2009) 127
4. D. C. G. Bedor, T. M. Gonçalves, M. L. L. Ferreira, C. E. M. de Sousa, A. L. Menezes, E. J. Oliveira, D. P. de Santana, *J. Chromatogr., B* **863** (2008) 46
5. S. E. Hismiogullar, E. Yarsan, *Hacet. Univ. J. Fac. Pharm.* **29** (2009) 95
6. R. Injac, J. Kac, K. Ovic-Rajic, B. Strukelj, *J. Food Drug Anal*. **16** (2008) 18
7. M. L. Fernández de Córdova, P. Ortega Barrales, G. Rodríguez Torné, A. Molina Díaz, *J. Pharm. Biomed. Anal*. **31** (2003) 669
8. M. H. Givianrad, M. Saber-Tehrani, P. Abroomand Azar, M. Mohagheghian, *Spectrochim. Acta* **78** (2011) 1196
9. V. Srivatsan, A. K. Dasgupta, P. Kale, R. R. Datla, D. Soni, M. Patel, R. Patel, C. Mavadhiya, *J. Pharm. Biomed. Anal.* **1031** (2007) 307
10. M. Goodarzi, P. Shahbazikhah, M. R. Sohrabi, M. Fathabadi, A. Hassannouri, *J. Chil. Chem. Soc.* **54** (2009) 309
11. F. E. Silva, M. F. Ferrao, G. Parisotto, E. I. Müller, E. M. Flores, *J. Pharm. Biomed. Anal.* **49** (2009) 800
12. Y. Wen, M. Zhang, Q. Zhao, Y. Q. Feng, *J. Agric. Food Chem.* **53** (2005) 8468
13. H. Amini, A. Ahmadiani, *J. Pharm. Biomed. Anal*. **43** (2007) 1146
14. F. C. C. R. de Paula, A. C. de Pietro, Q. B. Cass, *J. Chromatogr., A* **1189** (2008) 221
15. J. L. Flores, M. L. F. de Cordova, A. M. Diaz, *Anal. Chim. Acta* **600** (2007) 164
16. S. Wold, M. Sjostrom, L. Eriksson, *Chemometr. Intell. Lab*. **58** (2001) 109
17. R. G. Brereton, *Analyst* **125** (2000) 2125
18. R. Leardi, A. L. Gonzalez, *Chemometr. Intell. Lab.* **41** (1998) 195.