Monatsh Chem (2011) 142:773–778
DOI 10.1007/s00706-011-0513-7

ORIGINAL PAPER

Potentiality of partial least-squares multivariate calibration in the spectrophotometric analysis of binary mixtures of purine bases

H. Khajehsharifi · Z. Eskandari

Received: 19 December 2009 / Accepted: 20 April 2011 / Published online: 31 May 2011

Abstract A partial least-squares (PLS) modeling was developed for the simultaneous spectrophotometric determination of adenine (AD) and guanine (GU). The determination of these analytes is pharmacologically necessary. Multivariate calibration is used because of spectral overlapping. The calibration set contained AD and GU in the concentration range of 1.4–20.3 and 1.5–25.7 μg cm^{-3}, respectively. The absorption spectra were recorded from 200 to 300 nm. The predicted residual error sum-of-squares for AD and GU was 0.0500 and 0.4000 for number of principal components 3 and 2, respectively. The root mean square error of prediction for AD and GU was 0.0913 and 0.2582, respectively. The limits of detection were 0.02 and 0.03 μg cm^{-3} for AD and GU, respectively. The proposed method allows the simultaneous determination of AD and GU in spiked real matrixes of human urine, serum, and plasma.

Keywords Adenine · Guanine · Chemometrics · UV/Vis spectroscopy · Simultaneous determination · DNA

Introduction

Deoxyribonucleic acid (DNA) is an important substance in the storage of genetic information and protein biosynthesis. Adenine (AD) and guanine (GU) are important components (bases) found in DNA. Nucleosides and their metabolic products such as purine bases play fundamental roles in life processes [1]. Abnormal changes of the bases in an organism suggest the deficiency and mutation of the immune system and may indicate the presence of various diseases. Such changes have widespread effects on coronary and cerebral circulation, control of blood flow, prevention of cardiac arrhythmias, inhibition of neurotransmitter release, and modulation of adenylate cyclase activity [2, 3].

The separation and determination of AD and GU is an interesting and challenging task because these bases are involved in a variety of biochemical processes. In the ruminant nutrition field, the quantification of microbial protein synthesis and protein degradation in the rumen is critical in all the protein evaluation systems used for these animals. Purine bases play a very important role in the study of the nutrition of ruminants, because AD and GU are used, among others, like internal microbial markers to determine microbial protein synthesis [4]. Determining individual concentrations of GU and AD or their ratio in DNA is important in the measurement of nucleic acid concentration itself. Indeed, nucleic acid components in physiological fluids, tissues, and cells are related to the catabolism of nucleic acids, enzymatic degradation of tissues, dietary habits, and various salvage pathways. Therefore, detection of elevated levels of these substances could be indicative of certain diseases [5]. Their concentration levels are considered as important parameters for the diagnosis of cancers, AIDS, myocardial cellular energy status, disease progress, and therapy responses [6]. Therefore, the determination of nucleosides and their metabolic products is important in physiology and clinical fields. The level of AD and GU in plasma or urine has been suggested to be related to carcinoma or liver diseases [7], so it is important to establish sensitive methods for the detection of AD and GU.

H. Khajehsharifi (✉) · Z. Eskandari
Department of Chemistry, Yasouj University, Yasouj, Iran
e-mail: khajeh_h@yahoo.com; haka@mail.yu.ac.ir

In recent years, a series of methods, including mass spectrometry (MS) [8], high-performance liquid chromatography (HPLC) [9], ion-pairing liquid chromatography (IPIC) [10], capillary electrophoresis (CE) [11], flow-injection chemiluminescence (CL) [12], electrochemical methods [7, 13–16], and spectroscopic methods [17, 18], have been developed for the determination of nucleosides in biological fluids. Although these methods are sensitive, complicated instruments and time-consuming sample pretreatment are required. A spectrophotometric technique is always an acceptable alternative chemical analysis method owing to its acceptable precision and accuracy together with its lower cost compared with other techniques.

In this report, a spectrophotometric method coupled with partial least-squares (PLS) was developed for the simultaneous determination of AD and GU. The method was also applied for the determination of the analytes in human spiked serum, plasma, and urine.

Theory

PLS was originally developed to achieve two main objectives: to approximate X and Y accurately, and to model the relationship between X and Y. Details on the PLS algorithms can be found in Ref. [19]. PLS models a linear relationship between a set of input variables (predictor), $x_i \in R^N$, $i = 1 \ldots n$ and a set of output variables (responses), $x_i \in R^M$, $i = 1 \ldots n$ by means of latent variables. In the classical PLS method, the nonlinear iterative PLS (NIPALS) [19] algorithm is applied to the PLS regression in order to sequentially extract the latent vectors t, u and the weight vectors w, c from the X and Y matrices in decreasing order of their corresponding singular values. As a result, PLS decomposes X ($n \times N$) and Y ($n \times M$) matrices with mean zero into the form:

$$X = TP^{\mathrm{T}} + E \tag{1}$$

$$Y = UQ^{\mathrm{T}} + F \tag{2}$$

where T and U are ($n \times k$) matrices of the extracted k score vectors, P ($N \times k$) and Q ($M \times k$) are matrices of loadings, and E ($n \times N$) and F ($n \times M$) represent matrices of residuals. Unlike the classical PLS algorithm, the modified PLS algorithm normalizes the latent vectors t, u rather than the weight vectors w, c [20, 21]. The PLS regression model can be expressed with regression coefficient B and residual matrix R as follows:

$$Y = XB + R \tag{3}$$

$$B = W\left(P^{\mathrm{T}}W\right)^{-1}C^{\mathrm{T}} \tag{4}$$

where P ($N \times k$) is the matrix consisting of loading vectors $p_i = X^{\mathrm{T}}t_i/(t_i{}^{\mathrm{T}}t_i)$ $i = 1 \ldots k$. Due to the fact that $p_i{}^{\mathrm{T}}w_j = 0$ for $i > j$ and in general $p_i{}^{\mathrm{T}}w_j \neq 0$ for $i < j$, the matrix $P^{\mathrm{T}}W$ is upper triangular and thus invertible [22]. Moreover, using the fact that $t_i{}^{\mathrm{T}}t_j = 0$ for $i \neq j$ and $t_i{}^{\mathrm{T}}u_j = 0$ for $j > i$, Rannar et al. [23] derived the following equalities:

$$W = X^{\mathrm{T}}U \tag{5}$$

$$P = X^{\mathrm{T}}T\left(T^{\mathrm{T}}T\right)^{-1} \tag{6}$$

$$C = Y^{\mathrm{T}}T\left(T^{\mathrm{T}}T\right)^{-1} \tag{7}$$

Substituting Eqs. 5–7 into Eq. 4 using the orthogonality of the matrix T columns, we can write the matrix B in the following form:

$$B = X^{\mathrm{T}}U\left(T^{\mathrm{T}}XX^{\mathrm{T}}U\right)^{-1}T^{\mathrm{T}}Y \tag{8}$$

which will be used to make predictions in PLS regression.

Results and discussion

Absorption spectra

The electronic absorption spectra of AD and GU are shown in Fig. 1. As can be seen the spectra are overlapped. Thus, these compounds cannot be measured in the presence of each other by a simple calibration procedure without prior separation. Therefore multivariate calibration was used to resolve the spectra and for determination of each compound in the mixtures. The composition data of the solutions are listed in Table 1. Regarding spectral data (see Fig. 1), spectra were recorded in the region between 200 and 300 nm (in 1.00-nm steps).

One-component calibration

Individual calibration curves were constructed with several points (Fig. 2), as absorbance versus analyses concentration in the linear range of AD and GU. The wavelengths used to generate calibration curves were 200–300 nm. Linear regression results, line equations, and R^2 are also

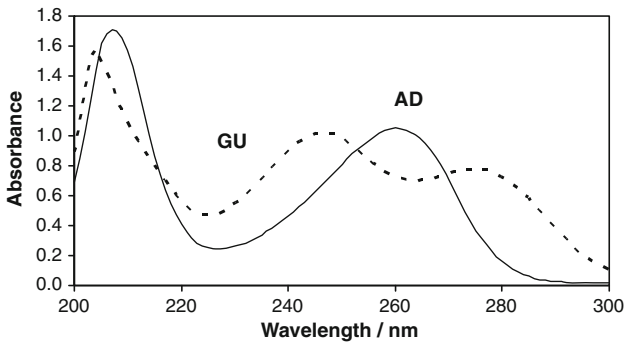

Fig. 1 Absorption spectra of AD (10.4 μg cm^{-3}) and GU (12.0 μg cm^{-3}) at pH = 7.0 and 273 K

Table 1 Concentration data of mixtures used in the calibration set for the determination of AD and GU

Mixture	AD ($\mu g\ cm^{-3}$)	GU ($\mu g\ cm^{-3}$)
M1	1.4	1.5
M2	6.1	1.5
M3	10.8	1.5
M4	15.5	1.5
M5	20.3	1.5
M6	1.4	7.1
M7	6.1	7.1
M8	10.8	7.1
M9	15.5	7.1
M10	20.3	7.1
M11	1.4	12.8
M12	6.1	12.8
M13	10.8	12.8
M14	15.5	12.8
M15	20.3	12.8
M16	1.4	18.4
M17	6.1	18.4
M18	10.8	18.4
M19	15.5	18.4
M20	20.3	18.4
M21	1.4	24.2
M22	6.1	24.2
M23	10.8	24.2
M24	15.5	24.2
M25	20.3	24.2

shown in Fig. 2. The limits of detection were 0.02 and 0.03 $\mu g\ cm^{-3}$ for AD and GU, respectively, and these were calculated according to calibration line characteristics.

Multicomponent calibration and prediction

Multivariate calibration methods such as PLS require a suitable experimental design of the standard belonging to the calibration set in order to provide good prediction. In this study, a mixture design was used for the experimental design. It is important that we use a method of design that does not create an underlying correlation among the concentrations of the components. The calibration matrix used for the analysis is shown in Table 2. For the prediction step six prepared mixtures that were not included in the previous set were employed as an independent test.

Number of factors

Determining how many factors to be used in the calibration is a key step in factor-based methods. Only those factors

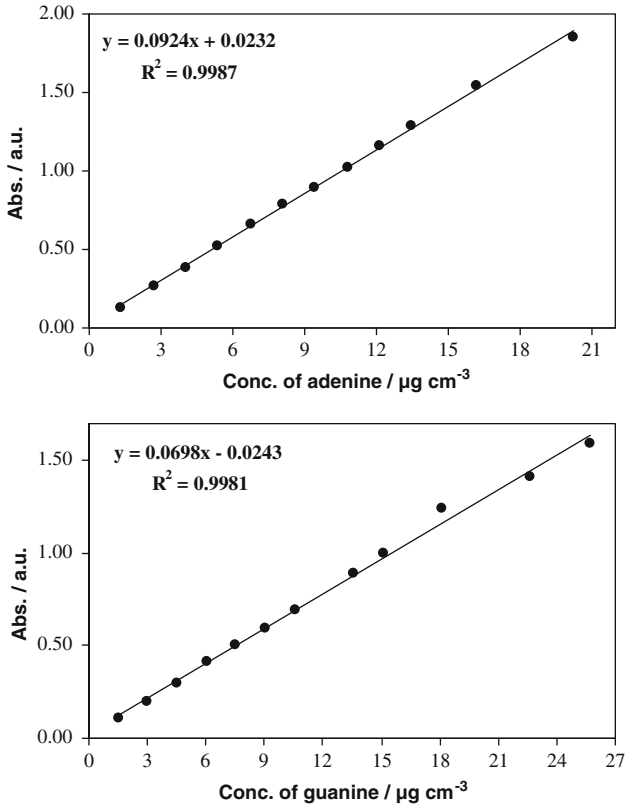

Fig. 2 Analytical curves for univariate determination of AD and GU

that contain analytical information must be kept. The discarded factors should contain only noise. The cross-validation procedure leaving out one sample at a time is used for this purpose and the predicted residual error sum-of-squares (PRESS) is calculated:

$$\mathrm{PRESS} = \sum_{i=1}^{n} (\hat{c}_i - c_i)^2 \tag{9}$$

where c_i is the reference concentration for the i-th sample, $\hat{c}_i$ represents the estimated concentration, and n is the total number of samples. The optimum number of factors was determined rather than the selection of the model, which yields a minimum in prediction error variance or PRESS; the model selected is the one with the fewest number of factors such that PRESS for that model is not significantly greater than the minimum PRESS. In our case, 13 factors (half the standards + 1) were used as the maximum number of initial factors. One reasonable choice for the optimum number of factors would be that number which yielded the minimum PRESS. Since there are a finite number of samples in the training set, in many cases the minimum PRESS value causes overfitting for unknown samples that were not included in the model. A solution to this problem was suggested by Haaland and Thomas [26] in which the PRESS values for all previous factors are

Table 2 Added and found results for the synthetic mixtures of AD and GU

Mixture	Added (μg cm^{-3})		Found[a] (μg cm^{-3})		Recovery (%)	
	AD	GU	AD	GU	AD	GU
1	2.0	7.6	1.9 ± 1.6	7.7 ± 1.9	95.3	101.0
2	4.7	3.0	4.6 ± 1.3	3.0 ± 1.5	98.7	100.7
3	10.1	21.9	10.2 ± 1.9	21.4 ± 1.9	101.0	97.7
4	13.5	2.3	13.5 ± 1.1	2.4 ± 1.9	100.2	104.5
5	17.6	16.6	17.5 ± 0.7	16.4 ± 1.8	99.5	99.0
6	6.1	9.8	6.2 ± 1.7	9.9 ± 1.8	101.5	101.5
					99.4 ± 2.2	100.7 ± 2.3

[a] Mean ± RSD% ($n = 3$)

Table 3 Statistical parameters of the optimized matrix using PLS

Component	NPC	PRESS	RMSEP	RSEP (%)
AD	2	0.0500	0.0913	0.8732
GU	3	0.4000	0.2582	2.1047

NPC number of principal components, *RMSEP* root mean square error of prediction, *RSEP* relative standard error of prediction

compared to the PRESS values at the minimum. The statistical F test can be used to determine the significance of PRESS values greater than the minimum. Table 3 shows the optimum number of factor and PRESS values for AD and GU. Plots of PRESS versus number of factors by PLS are shown in Fig. 3.

Determination of AD and GU in synthetic mixtures

The predictive ability of the method was determined using six two-component mixtures. The results obtained by applying the PLS algorithm to six synthetic samples are listed in Table 2 which also shows the recoveries for the synthetic series of AD and GU mixtures.

Model quality criteria

For the evaluation of the predictive ability of a multivariate calibration model, the root mean square error of prediction (RMSEP) and relative standard error of prediction (RSEP) can be used [27]:

$$\mathrm{RMSEP} = \frac{1}{n}\left[\sum_{i=1}^{n}(\hat{c}_i - c_i)^2\right]^{0.5} \quad (10)$$

$$\mathrm{RSEP}(\%) = 100\left[\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{\sum (c_i)^2}\right] \quad (11)$$

where c_i and $\hat{c}_i$ are the predicted concentration and the observed value of the concentration in the sample, respectively, and n is the number of samples in the prediction set.

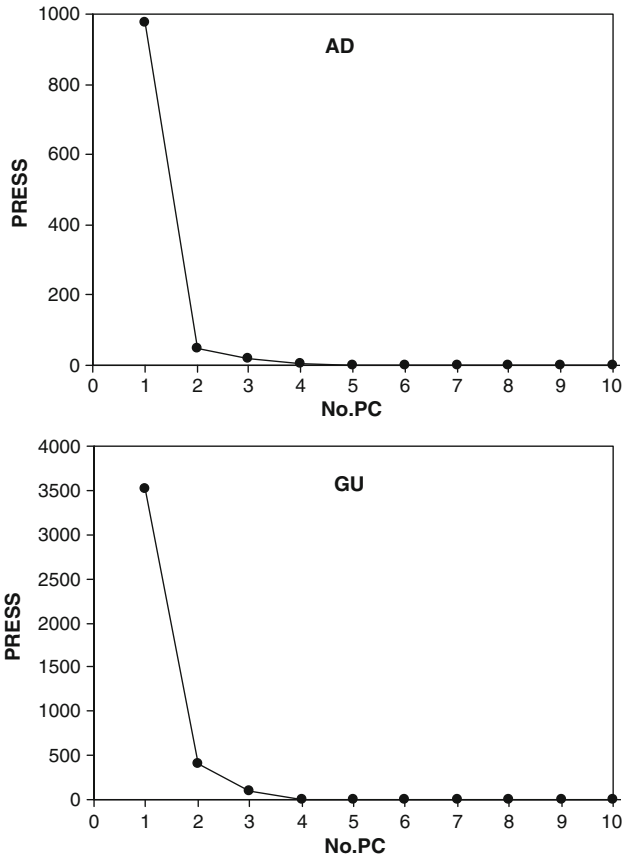

Fig. 3 Plots of PRESS versus number of factors by PLS

The values of RMSEP and RSEP (%) for AD and GU are summarized in Table 3.

Determination of AD and GU spiked in real samples

To assess the reliability of the method, prepared real samples were analyzed. Table 4 shows the results and the composition of the real samples.

The validation of the method was carried out by comparing with the labeled amounts. As can be seen, the

Table 4 Recoveries of AD and GU in spiked real samples by PLS

Real samples	Mixtures	Added (μg cm^{-3})		Found[a] (μg cm^{-3}) ± RSD %		Recovery (%)	
		AD	GU	AD	GU	AD	GU
Serum	S1	4.8	14.2	4.6 ± 1.5	13.9 ± 0.9	95.8	97.9
	S2	16.6	9.5	16.8 ± 0.5	9.3 ± 1.1	101.2	97.9
	S3	8.4	7.1	8.3 ± 1.7	7.3 ± 1.7	98.8	102.8
Plasma	P1	4.8	14.2	4.9 ± 2.2	13.9 ± 0.7	102.1	97.9
	P2	16.6	9.5	16.8 ± 0.6	9.6 ± 0.8	101.2	101.1
	P3	8.4	7.1	8.0 ± 1.9	7.8 ± 1.9	95.2	109.9
Urine	U1	4.8	14.2	4.9 ± 1.1	13.7 ± 1.8	102.1	96.5
	U2	16.6	9.5	16.8 ± 0.6	9.8 ± 1.6	101.2	103.2
	U3	8.4	7.1	8.1 ± 1.7	7.7 ± 1.7	96.4	108.5
Mean recovery ± SD						99.3 ± 2.8	101.7 ± 4.8

[a] Mean ± RSD% ($n = 3$)

recovery was quantitative and there were no significant differences between the amounts obtained from this method and labeled amounts. The plots of the predicted concentration versus actual values are shown in Fig. 4 for AD and GU (line equations and R^2 values are also shown).

Conclusion

The PLS method was employed to model the spectrophotometric data of AD and GU. A mixture of AD and GU is a difficult complex system due to the high spectral overlapping observed between their absorption spectra. To overcome the drawback of spectral interferences PLS multivariate calibration approaches were applied. According to the obtained results, application of the UV spectrophotometric method with multivariate calibration is an effective and accurate approach for the simultaneous determination of purine bases in binary mixtures of synthetic mixtures and simultaneous determination of these bases in human serum, human urine, and plasma samples.

Experimental

Chemicals

AD and GU were purchased from Fluka, whereas trichloroacetic acid was supplied by Merck. The stock solutions of AD and GU were prepared by dissolving them in a buffer solution (pH = 7.0) that was prepared from KH_2PO_4 and NaOH (Merck). All used chemicals were of analytical reagent grade. Throughout the experiments, doubly distilled water was used.

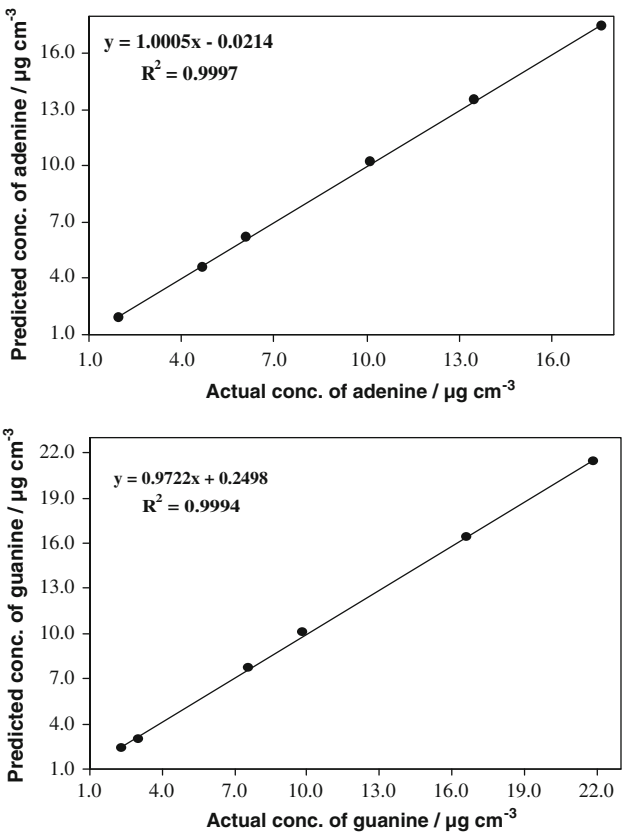

Fig. 4 Plots of predicted concentration versus actual concentration of AD and GU by PLS

Instrumentation and software

Electronic absorption measurements were carried out on a Jasco v-570 spectrophotometer (slit width 1 nm, scan rate 2,000 cm/min) using 1.0-cm quartz cells. Measurements of pH were made with a Metrohm 692 pH meter using a combined electrode. All absorption spectra were digitized

and stored at wavelengths from 200 to 300 nm in steps of 1 nm and then transferred in TXT format to an Pentium 4, 2.4 GHz computer using MATLAB software, Version 7 (MathWorks). PLS calculus was carried out in the PLS Toolbox (Eigenvector Company, Version 2.5).

Analysis of the calibration set

A mixture design for the two-component mixtures was used as the calibration set in order to provide good prediction in the PLS method. A training set of 25 samples was taken (Table 1). The concentrations of AD and GU were varied in the ranges 1.4–20.3 and 1.5–24.2 μg cm^{-3}, respectively. The mixed standard solutions were placed in a 10-cm^3 volumetric flask and completed to final volume with buffer solution (pH = 7.0). Finally the absorption spectra of all prepared solutions were recorded between 200 and 300 nm against a blank of buffer.

Analysis of the prediction set

For the prediction set, six series of mixtures were prepared randomly, but because they were employed as an independent test the concentrations used were not present in the previous set (Table 2). The concentration added was is the range 2.0–17.6 and 2.3–16.6 μg cm^{-3} for AD and GU, respectively.

Analysis of real samples

Serum and plasma

The serum and plasma samples were homogenized. For the deproteinization, 1 cm^3 of 24% w/v trichloroacetic acid was added to 1 cm^3 of serum and 1 cm^3 of plasma, and after 15 min the resulting mixture was centrifuged at 3,000 rpm [24]. The pH of the supernatant solution was fixed at 7.0 by adding NaOH solution. Afterwards, the appropriate amounts of the stock solutions of AD and GU were added to 0.5 cm^3 of the final prepared serum and plasma, then filled to the final volume (10 cm^3) with buffer solution to obtain the desired concentration. The electronic absorption spectrum was recorded in the range 200–300 nm against a blank solution of serum.

Urine

The urine sample was diluted 1:3 with distilled water. Then cell debris and the particulate matter were removed from the urine using low-speed centrifugation for 5 min at 1,500 rpm [25]. Afterwards the pH of sample was fixed at 7.0 by adding NaOH solution. Then the appropriate amounts of the stock solutions of AD and GU were added to 0.5 cm^3 of the final prepared urine and completed to the final volume (10 cm^3) with buffer solution to obtain the desired concentration. The electronic absorption spectrum was recorded in the range 200–300 nm against a blank of urine.

Acknowledgments We are grateful to Yasouj University for supporting the research.

References

1. Saenger W, Cantor CR (1984) Principles of nucleic acid structure. Springer, New York
2. Li SP, Li P, Dong TTX, Tsim KWK (2001) Electrophoresis 22:144
3. Malathi R, Johnson IM (2001) J Biomol Struct Dyn 18:709
4. Carro MD, Miller EL (2002) Anim Sci 75:315
5. Sheng R, Ni F, Cotton TM (1991) Anal Chem 63:437
6. Yang FQ, Guan J, Li SP (2007) Talanta 73:269
7. Sun W, Li Y, Duan Y, Jiao K (2008) Biosens Bioelectron 24:988
8. Huang YF, Chang HT (2007) Anal Chem 79:4852
9. Gill BD, Indyk HE (2007) Int Dairy J 17:596
10. Ganzera M, Vrabl P, Worle E, Burgstaller W, Stuppner H (2006) Anal Biochem 359:132
11. Yeh CF, Jiang SJ (2002) Analyst 127:1324
12. Liu EB, Xue BC (2006) J Pharm Biomed Anal 41:649
13. Wang Z, Xiao S, Chen Y (2006) J Electroanal Chem 589:237
14. Abbaspour A, Noori A (2008) Analyst 133:1664
15. Xiao F, Zhao F, Li J, Liu L, Zeng B (2008) Electrochim Acta 53:7781
16. Zen JM, Chang MR, Ilangovan G (1999) Analyst 124:679
17. Heisler I, Keller J, Tauber R, Sutherland M, Fuchs H (2002) Anal Biochem 302:114
18. Amri CE, Baron MH, Maurel MC (2003) Spectrochim Acta A 59:2645
19. Geladi P, Kowalski BR (1986) Anal Chim Acta 185:1
20. Rosipal R, Trejo LJ (2001) J Mach Learn Res 2:97
21. Lewi PJ (1995) Chemometr Intell Lab 28:23
22. Hoskuldsson A (1988) J Chemometr 2:211
23. Rannar S, Lindgren F, Geladi P, Wold S (1994) J Chemometr 8:111
24. Khajehsharifi H, Sadeghi M, Pourbasheer E (2009) Monatsh Chem 140:685
25. Khajehsharifi H, Pourbasheer E (2008) J Chin Chem Soc 55:163
26. Haaland DM, Thomas EV (1988) Anal Chem 60:1193
27. Wiberg K, Sterner-Molin A, Jacobsson SP (2004) Talanta 62:567