

Research Article

Synergy Interval Partial Least Square (siPLS) with Potentiometric Titration Multivariate Calibration for the Simultaneous Determination of Amino Acids in Mixtures

Xiaobo Zuo, Sheng Fang and Xianli Liang

School of Food Science and Biotechnology, Zhejiang Gongshang University, Xuezheng Street No. 18, Hangzhou 310018, China

Abstract: The simultaneous determination of amino acids in solution is important for food and nutrient industries. This study investigated the feasibility of potentiometric titration with synergy interval Partial Least Square (siPLS) for the simultaneous determination of glycine, glutamic acid and phenylalanine in aqueous solution. The methods used were compared with the traditional Partial Least Square (PLS) that based on full pH range. The performance was evaluated by the Root Mean Square Error of Cross-Validation (RMSECV), the Root Mean Square Error of Prediction (RMSEP) and the correlation coefficient (R). By optimizing the pH region with siPLS, a good linear model was produced for the calibration set with correlation coefficient R of 0.9947, 0.9956 and 0.9913 for glycine, glutamic acid and phenylalanine, respectively. A single set of synergy mixtures was tested independently and good results were obtained. The results show that the siPLS method can locate the informative region by using a graphically-oriented interface which is more easily to use and interpret. The study proves the feasibility of potential titration with chemometrics in the simultaneous determination of amino acid mixtures without preliminary separation.

Keywords: Amino acids, chemometrics, partial least square, potentiometric titration

INTRODUCTION

The simultaneous determination of different amino acids in solution is especially important for food and biopharmaceutical manufacturing as well as other related industries (Wakayama *et al.*, 2010). Until now, HPLC based method is the most popular methodology for analyzing amino acid components. A wide range of pre- or post-column fluorescent labeling reagents such as fluorescamine, ninhydrin, 7-fluoro-4-nitrobenz-2, 1, 3-oxodiazole and dansyl chloride etc are developed for the HPLC analysis of different amino acids by using fluorescence detector (Zhao *et al.*, 2013; Redruello *et al.*, 2013; Chen *et al.*, 2005). In addition to the HPLC method, gas chromatography, capillary electrophoresis and liquid chromatography combined with mass spectrometry are also developed for the determination of amino acids (Mudiam *et al.*, 2012; Mohabbat and Drew, 2008; Smith, 1997). These methods are quite well suited for the simultaneous determination of amino acids in complex system, especially when unknown substances are presented in the mixture. However, there are some cases that the amino acids in mixtures are with the same type but in different concentrations, such as some amino acids nutrient solutions (Ohtani *et al.*, 2006; Cerdán *et al.*, 2013) and amino acid mixtures that extracted from protein hydrolysates (Liebster *et al.*,

1961; Moore and Stein, 1949) in food industry. For these samples, a simple and efficient way other than the expensive and time consuming HPLC methods, for the simultaneous determination of different amino acids will be beneficial.

The potentiometric titration method is always used for determining concentration of single amino acid in solution, but can hardly be used for the simultaneous determination of several amino acids in the mixture (Michalowski *et al.*, 2005; Ni and Kokot, 2008). Nowadays, the application of chemometrics in food industries has been an interesting area in the resolution of multi-component mixtures offers the advantages of fast, minimizing preliminary separation steps and eliminating the use of chemical reagents (Munck *et al.*, 1998; Christensen *et al.*, 2006). For potentiometric titration, the application of chemometrics was first introduced by Lindberg and Kowalski (1988) for the simultaneous analysis of acids with Partial Least Squares (PLS) regression. After that, several papers have been published using PLS calibration method to acid-base titration (Ni, 1998; Shamsipur *et al.*, 2002), potentiometric precipitation titration (Ni and Wu, 1999) and complexometric titration (Zhang *et al.*, 2005). Recently, many new algorithms such as Artificial Neural Network (ANN) and orthogonal signal correction have been applied to process the

Corresponding Author: Sheng Fang, School of Food Science and Biotechnology, Zhejiang Gongshang University, Xuezheng Street No. 18, Hangzhou 310018, China, Tel.: +86 571 88071024

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

potentiometric titration of different acid mixtures (Song *et al.*, 1993; Aktaş and Yaşar, 2004; Ghorbani *et al.*, 2006). For example, the orthogonal signal correction has been applied to the simultaneous conductometric titration of mixtures of acetic acid, monochloroacetic acid and trichloroacetic acid (Ghorbani *et al.*, 2006). However, all these multivariate calibration methods in titration used the whole information pH region to build a calibration model and very little attention has been given to the optimized variable selection in regression.

It has been demonstrated that the performance of optimal variable selection before regression in multivariate calibration can improve the accuracy and robustness of the model (Li *et al.*, 2014; Du *et al.*, 2004; Jiang *et al.*, 2002). Many algorithms that used for the optimal variable selection have been developed and applied, such as moving windows based methods (Du *et al.*, 2004; Jiang *et al.*, 2002; Fang *et al.*, 2009), genetic algorithm (Goicoechea and Olivieri, 2003), synergy interval PLS (siPLS) (Nørgaard *et al.*, 2000; Leardi and Nørgaard, 2004) and CLoVA (Hemmateenejad *et al.*, 2013; Hemmateenejad and Karimi, 2011). Very recently, the genetic algorithm and Competitive Adaptive Reweighted Sampling (CARS) method were used to select the effective wavelengths in the simultaneous determination of three branched-amino acids (leucine, isoleucine and valine) by Fourier transform near-infrared spectral technique (Wei *et al.*, 2014). In these methods, the iPLS and siPLS (Nørgaard *et al.*, 2000; Leardi and Nørgaard, 2004) have been shown as efficient tools to choose the optimal sub regions in a graphical manner which can provide an overall picture of the model performance in different subintervals. Many papers have indicated that the optimal subintervals selected by iPLS or siPLS methods could provide more precision prediction results than traditional PLS model which based on full-spectrum region. For example, the spectrophotometric methods combined with iPLS or siPLS have successfully applied for the determination of total volatile basic nitrogen content of pork (Cai *et al.*, 2011), quality parameters of biodiesel/diesel blends (Ferrão *et al.*, 2011), contents in vinegar (Chen *et al.*, 2012a, b), antioxidant activity in dark soy sauce (Ouyang *et al.*, 2012). However, to the best of our knowledge, the application of iPLS and siPLS to potentiometric titration multivariate calibration to select the optimal pH subintervals has not yet been explored (Fang *et al.*, 2009).

The aim of this study was to explore the potential of siPLS algorithm for the optimal subintervals selection in potentiometric titration multivariate calibration and its use for the simultaneous determination of amino acids mixtures in aqueous solution. The ternary mixtures of glycine, glutamic acid and phenylalanine in aqueous solution are used as a model system. The performance was evaluated according to the Root Mean Square Error of Cross-Validation (RMSECV), the Root Mean Square Error of Prediction (RMSEP) and the correlation coefficient (*R*).

MATERIALS AND METHODS

Materials: All amino acids are analytical-reagent grade chemicals. Stock solutions of hydrogen chloride (0.1 M), *L*-glycine (0.05 M), *L*-glutamic acid (0.05 M) and *L*-phenylalanine (0.05 M) were prepared according to classical method with ultra-pure water throughout. The resistivity of ultra-pure water is ≥ 18.2 M Ω -cm. Since no standardization procedure for potentiometric titration multivariate calibration was necessary, the prepared 0.1 mol/L NaOH solution was used both for calibration and prediction. Solutions of sodium chloride (NaCl) with 1.0 M and Hydrogen Chloride (HCl) with 0.1 M were prepared to adjust the ionic strength and initial pH of the sample solutions, respectively.

Equipments and apparatus: The titrations were conducted in batch mode with magnetic stirrer, syringe pumps and the glass vessels, which were standard equipments. A syringe pump (Harvard apparatus) was used for the precise addition of the titrant. Measurements of pH (± 0.001) were carried out with a Metrohm pH meter by using a combined glass electrode. All experiments were performed at room temperature about 22°C. All calculations were performed on a PC with the Windows operating system, which was equipped with the Excel and Matlab programs. The iToolbox (Nørgaard *et al.*, 2000; Leardi and Nørgaard, 2004) for Matlab was used for the variable selection and multivariate models of iPLS and siPLS.

Procedure: In a typical procedure, a suitable amount of three amino acids stock solution was placed in a glass vessel. Then 1.0 mL of 1.0 M sodium chloride (NaCl) solution and 0.4 mL of 0.1 M Hydrogen Chloride (HCl) solution were added to the vessel for adjusting the ionic strength and initial pH of solution, respectively. Since the same amount of hydrogen chloride was added to all the samples in calibration and prediction sets, the effect of hydrogen chloride in the calibration was neglectable. The solution was finally diluted up to 5.0 mL with ultra pure water. The mixture was stirred and then titrated by the precise addition of sodium hydroxide with flow rate ranged between 0.15 and 0.35 mL/min. The pH meter was used to monitor the solution pH data during titration and the titrant volumes added to reach the predetermined pH values were recorded. Finally, the experimental data were processed and two matrices of titration data were obtained; the volume of titrant at each pH point (0.1 pH interval from 3.0 to 12.0) formed the first matrix *V* and the concentration of three amino acids formed the second matrix *C*.

Statistics analysis: For the construction of chemometrics model, two common statistical parameters were chosen to evaluate the model

performance in the prediction ability of glycine, glutamic acid and phenylalanine in aqueous mixtures. The first parameter is the root mean square error of cross validation (RMSECV) and Root Mean Square Error of Prediction (RMSEP) as shown in Eq. (1). The RMSECV or RMSEP parameter is an expression of the average error in the analysis of each amino acid in the calibration or prediction set:

$$RMSECV/P = \left[\frac{1}{n} \sum_{i=1}^n (c_i - c_{i,pred})^2 \right]^{1/2} \quad (1)$$

Another commonly used parameter is the relative error of prediction (REP) that indicates the predictive ability of established chemo metrics model for each component, as calculated from the following equation:

$$REP\% = \sqrt{\frac{\sum_{i=1}^n (c_i - c_{i,pred})^2}{\sum_{i=1}^n (c_i)^2}} \times 100 \quad (2)$$

where, c_i is the reference concentration and $c_{i,pred}$ represents the predicted concentration of the analyte in i sample by chemometrics model and n is the number of sample used. Correlation coefficient R was calculated by equation:

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (c_{i,pred} - c_i)^2}{\sum_{i=1}^n (c_{i,ave} - c_i)^2}} \quad (3)$$

where, $c_{i,ave}$ is the mean of the reference results for all samples in calibration or prediction set.

RESULTS AND DISCUSSION

In the traditional acid-base titration system, the endpoints are always determined by visual indicators. However, for acid mixtures when the ΔpK between any two acids is less than 4 logarithmic units, the titration curve will be overlapped and it will make the determination of endpoint more difficult. For amino acids used in this study, the pK values are listed in the Table 1. As can be seen in the Table 1 that pK of all these three species are very close and far less than 4 units. In the traditional titration method, the titration curve of these amino acids will be overlapped and affected each other. Under that condition, multivariate calibration methods can be used to the pH titration for the simultaneous determination of amino acids in mixtures and offer the advantage of eliminating preliminary separation steps.

Experimental design of the calibration sets: The multivariate calibration process requires a training data

Table 1: The pK_a values of glycine, glutamic acid and phenylalanine from reference (Henchoz *et al.*, 2007)

Amino acid	pK_1 (α -COOH)	pK_2 (α -NH ₃ ⁺)	pK_R (β -COOH)
Glycine	2.43	9.60	
Glutamic acid	2.29	9.87	4.10
Phenylalanine	2.18	9.15	

Table 2: Concentration data of different mixtures in the calibration set
Concentration (mmol/L)

Samples	<i>L</i> -glycine (Gly)	<i>L</i> -glutamic acid (Glu)	<i>L</i> -phenylalanine (Phe)
M1	4.000	4.000	4.000
M2	4.000	8.000	8.000
M3	4.000	12.00	12.00
M4	4.000	16.00	16.00
M5	8.000	4.000	8.000
M6	8.000	8.000	4.000
M7	8.000	12.00	16.00
M8	8.000	16.00	12.00
M9	12.00	4.000	12.00
M10	12.00	8.000	16.00
M11	12.00	12.00	4.000
M12	12.00	16.00	8.000
M13	16.00	4.000	16.00
M14	16.00	8.000	12.00
M15	16.00	12.00	8.000
M16	16.00	16.00	4.000

set, which includes the known concentration of each substances and the corresponding pH titration spectrum of that sample. Then, some of the multivariate calibration algorithm such as principal component regression or partial least squares PLS are applied to construct a mathematical model. The established model can be efficiently used to predict the concentrations of unknown mixtures. In this study, a calibration set contains 16 standard solutions was prepared according to four-level orthogonal design. It is known that the using of orthogonal design to construct a training set, maximum information can be obtained from only a few samples (Ni, 1998). In Table 2, all the concentrations of the ternary mixtures in the calibration set are summarized.

Many works have pointed out that it was necessary to preprocess the raw spectrum in order to develop stable and reliable calibration models (Ni, 1998; Shamsipur *et al.*, 2002; Ni and Wu, 1999). In this study, the simple smoother developed by Savitzky and Golay (1964) was applied. The SG Smoothing could give balance in such a way that the noise is maximally removed, while the signal features are kept intact as much as possible (Savitzky and Golay, 1964). It is known that the SG algorithm contains two parameters, the polynomial order N and the window size W . Here, we use the Root Mean Square Error (RMSE) for the calibration set by full spectrum PLS were compared as a criterion, under different polynomial order N and window size W . Finally, an optimized combination of $N = 3$ and $W = 5$ was chosen, which is a good compromise in practice. The titration curve of calibration mixtures after SG smoothing are presented in Fig. 1. From Fig. 1,

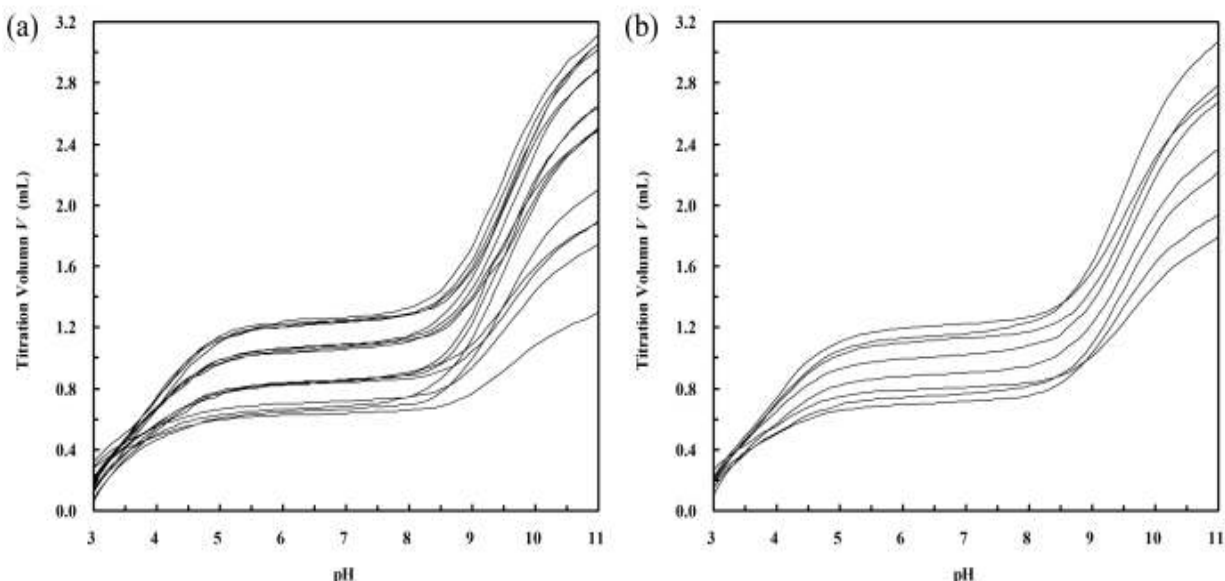


Fig. 1: The titration curve for samples after Savitzky and Golay (1964) smoothing: (a) calibration samples and (b) prediction samples

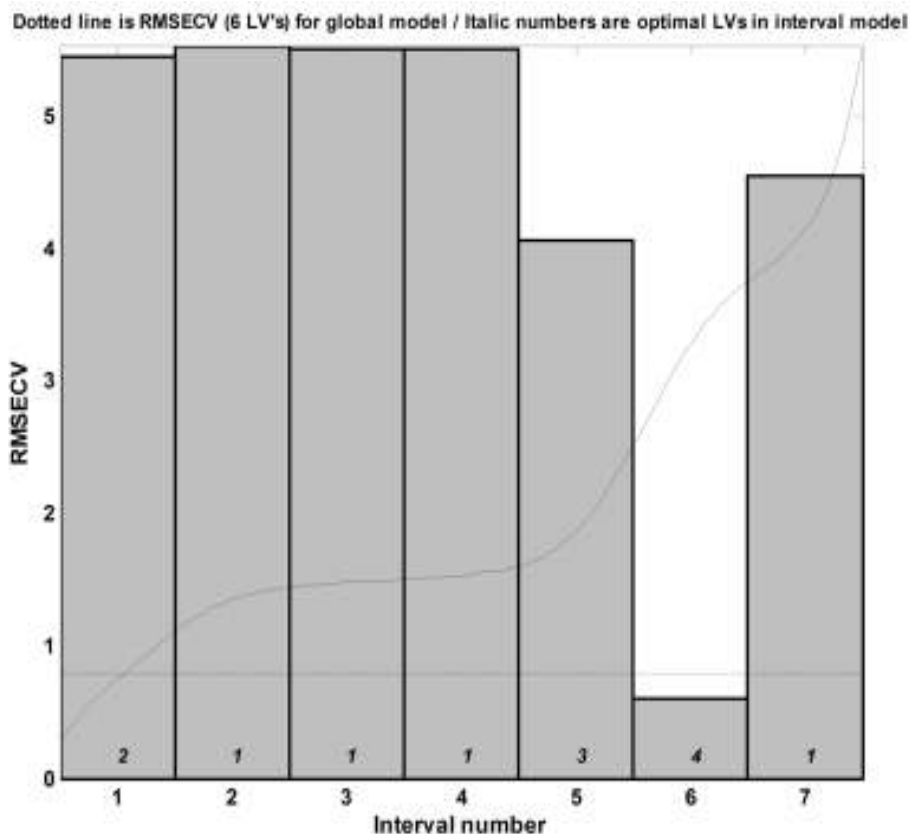


Fig. 2: Cross-validated prediction errors (RMSECV) values to full-pH model and interval models (bars) to glycine determination using PLS and iPLS algorithms (dotted line and numbers above interval numbers refers to full-spectrum RMSECV and latent variables using in each model, respectively)

it can be found that all the titration curves have a large jump in the pH range 5-8 which is corresponding to the

acid group of amino acids with pK between 2 and 4. It also can be seen that the y-coordinates of titration

volume V and x-coordinates of pH resemble the absorbance and wavelength in traditional spectrophotometry, respectively. From this point of view, the titration curve could be seen as the titration spectrum and any chemometrics algorithm used in spectrum can be applied to titration multivariate calibration.

siPLS model for selecting optimal regions: For the building of iPLS and siPLS models, the full pH range between pH 3.0 and 12.0 was first divided into 7 equidistant subintervals, with 11 pH points in each interval. For every subinterval, a PLS model with different numbers of latent variables was established. The RMSECV for every model was calculated as a critical value for the comparison with the whole pH region model. Figure 2 shows the RMSECV obtained by iPLS for each subinterval and latent variables for each model that represented by number in the bars. The RMSECV for the full pH region using 6 latent variables are shown by the dotted line for the purpose of comparison. It can be found that the iPLS model with interval number 6 at pH between 9.5 and 10.8 give the smallest RMSECV values with 0.6085. This result is better than 0.7936 of the full-pH region PLS model as shown in Table 3.

The iPLS method can find out the optimal subinterval, however, the different combinations of

subintervals may result in models with better predictive abilities. Therefore, based on the results obtained above, variables selection by siPLS was implanted to test different combinations of intervals. The principle of siPLS method is to calculate all different combinations of two, three or four subintervals that obtained by iPLS model. The results of different combinations obtained by PLS, iPLS and siPLS were shown in Table 3. As can be seen in the table, many combinations of subintervals give better results than full region and subinterval 6. The combination of 2, 6 and 7 subintervals with 6 latent variables gives the lowest RMSECV with 0.4740 that was far better than the subinterval 6. The results demonstrate that the siPLS algorithm can avoid the loose of relevant information region which will improve the performance of calibration model.

For a full view of model comparisons, a graphic test of PLS, iPLS and siPLS model for glycine was shown in Fig. 3. The results in Table 3 and Fig. 3 clearly show that the selection of optimal pH region in potentiometric titration multivariate calibration improve the model performance. The correlation coefficient R improves from 0.9844 for PLS to 0.9947 for siPLS model. As shown in Fig. 3, the combined subintervals 2, 6 and 7 with pH region 4.3-5.6 and 9.5-12 were finally selected to construct the calibration model for glycine. It is interesting to find out that this pH

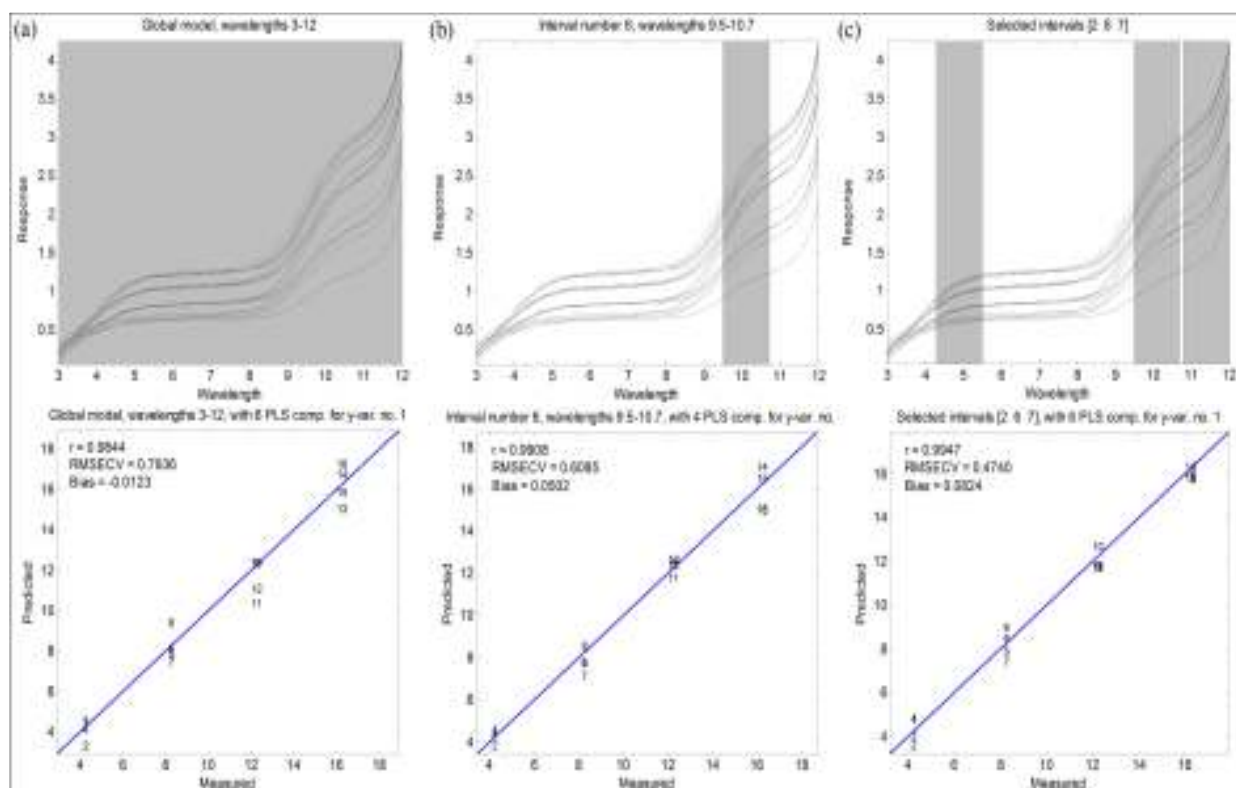


Fig. 3: The graphic and statistic results of PLS, iPLS and siPLS for glycine

Table 3: Results of PLS, iPLS and siPLS models for glycine determination

Intervals	pH range	LVs	Method	R	RMSECV
Full	3-12	6	PLS	0.9844	0.7936
6	9.5-10.8	4	iPLS	0.9908	0.6085
2, 6	4.3-5.6 and 9.5-10.8	4	siPLS	0.9925	0.5478
1, 6	3-4.3 and 9.5-10.8	5	siPLS	0.9918	0.5722
4, 6	4.3-5.6 and 9.5-10.8	5	siPLS	0.9907	0.6081
2, 6, 7	4.3-5.6 and 9.5-12	6	siPLS	0.9947	0.4740
1, 6, 7	3-4.3 and 9.5-12	7	siPLS	0.9939	0.4978
2, 3, 6	4.3-6.9 and 9.5-10.8	4	siPLS	0.9923	0.5553

Table 4: Results of PLS, iPLS and siPLS models for glutamic acid determination

Intervals	pH range	LVs	Method	R	RMSECV
Full	3-12	3	PLS	0.9915	0.5866
2	4.3-5.6	3	iPLS	0.9956	0.4189
2, 3	4.3-6.9	3	siPLS	0.9956	0.4258
1, 2	3-5.6	3	siPLS	0.9940	0.4918
3, 4	5.6-8.2	2	siPLS	0.9936	0.5076
1, 3, 7	3-4.3 and 5.6-6.9 and 10.8-12	3	siPLS	0.9935	0.5128
1, 2, 3	3-6.9	3	siPLS	0.9933	0.5206
1, 3, 6	3-4.3 and 5.6-6.9 and 9.5-10.8	3	siPLS	0.9933	0.5213
1, 3, 5	3-4.3 and 5.6-6.9 and 8.2-9.5	3	siPLS	0.9932	0.5265

range was exactly locate in the information regions of the glycine with pK_1 of 2.43 and pK_2 of 9.60. This result further confirms that siPLS can automatically locate the informative region for the given substance in multicomponent potentiometric titration. It can also be seen that the use of a graphically-oriented interface as shown in Fig. 3 in the iPLS and siPLS toolbox make it more easy to use and interpret the obtained results.

Table 4 shows the model results of PLS, iPLS and siPLS for glutamic acid. As can be seen that the full pH region PLS calibration model result a RMSECV of 0.5866 with 3 latent variables. In addition, the selected subintervals and their combinations all give better results compare with the full pH region PLS model. However, it is found that the combination of two or three subintervals does not give better results than the single subinterval 2.

As shown in Table 4, the subinterval 2 with pH range 4.3-5.6 give the best result with a correlation coefficient R of 0.9956 and RMSECV of 0.4189. So the intervals were selected to build the final iPLS model for glutamic acid. The above results for glutamic acid seem interesting, since the addition of other subintervals does not improve the model performance as in the case of glycine. It is known that the glutamic acid molecule contains another β -COOH group with pK of 4.1 by compare with glycine or phenylalanine. The optimal pH region of 4.3-5.6 obtained by iPLS may correspond to information area of this acid group. The different pK of β -COOH group in glutamic acid molecule make it sufficient to characterize glutamic acid in the known mixture. So the addition of information area of α -COOH and α -NH₃⁺ that are very close to glycine and phenylalanine will in turn bring more interference and perturbation factors. This result demonstrated the needs of pH or wavelength selection in the potentiometric

titration multivariate calibration to build a robust model (Fig. 4).

The siPLS was also applied for the quantification of phenylalanine. In the same way, the whole pH was divided into 7 equidistant subintervals. The correlation coefficient R and RMSECV are calculated and shown in Table 5. It can be seen that the best iPLS model using number 5 subintervals with pH region of 8.2-9.5 does not produce better results than the full pH region PLS model. On the other hand, combinations of two or three subintervals can give lower RMSECV values than the full pH region PLS model. These combinations always contain the number 6 or 7 subintervals with pH regions of 9.5-10.8 or 10.8-12, respectively. Take into account that the phenylalanine contains α -NH₃⁺ group that with lowest pK of 9.15 in all the three amino acids. So the information area for phenylalanine may locate at high pH regions. We found that the combination of 5, 6 and 7 subintervals can give a RMSECV value of 0.6178 with only 4 LVs. Although it gives a little larger RMSECV than the combination of 3, 6 and 7 with 0.5809, the model with combination of 5, 6 and 7 contains less number of LVs. It is known that the over fitting problem may occur at large number of LVs. As pointed out by Norgaard *et al.* (2000) the over fitting problem is not only case for siPLS toolbox but goes for all variable selection methods. So for variable selection in potentiometric titration or spectrum multivariate calibration, carefully selection with knowledge of molecular information will be beneficial to build a robust model. Here, the siPLS models using the pH subintervals 5, 6 and 7 were developed and used for the quantification of aspartame.

Prediction of synthetic mixtures of amino acids: In order to evaluate the predictive ability of different models in the analysis of real samples, the analysis of eight three-component of synthetic three-component

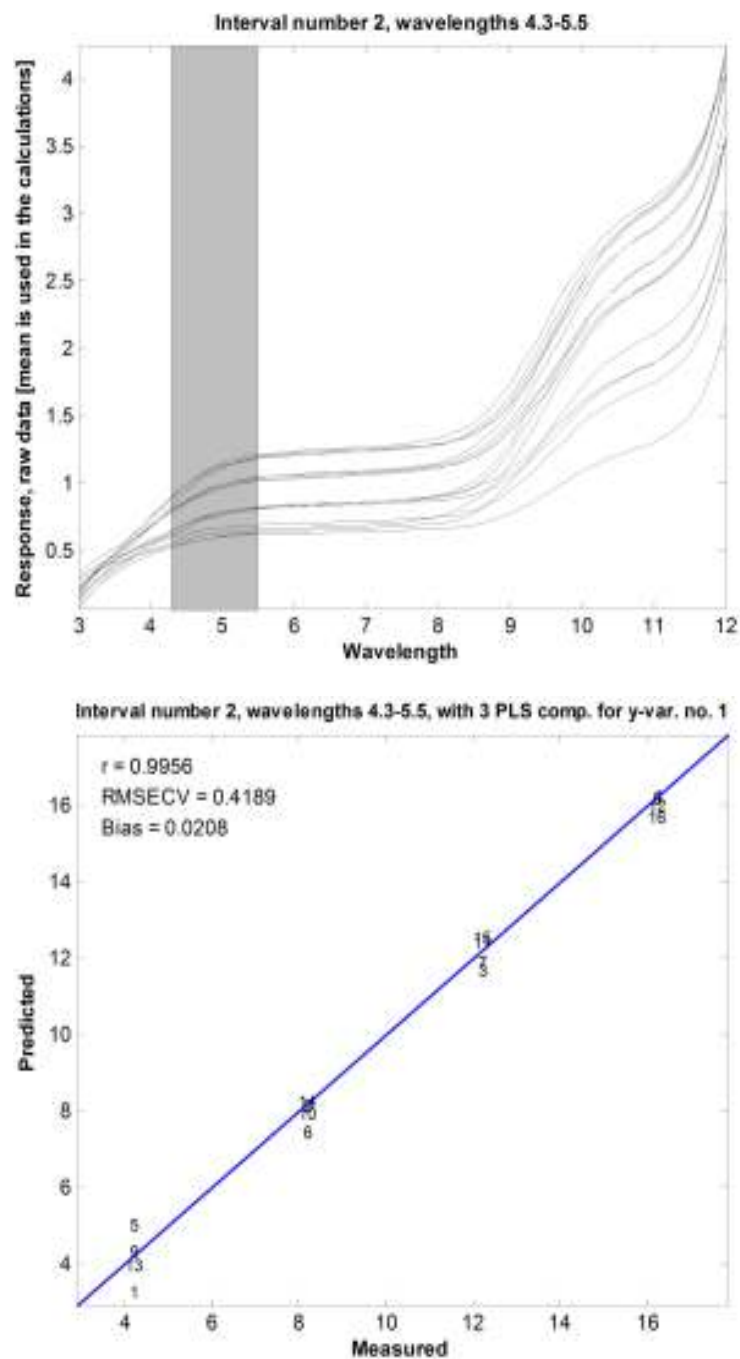


Fig. 4: Regions selected to build models with subinterval 2 at pH 4.3-5.6 and results for glutamic acid

Table 5: Results of PLS, iPLS and siPLS models for phenylalanine determination

Intervals	pH range	LVs	Method	R	RMSECV
Full	3-12	4	PLS	0.9831	0.8033
5	8.2-9.5	3	iPLS	0.8425	2.6736
2, 6	4.3-5.6 and 9.5-10.8	3	siPLS	0.9900	0.6537
5, 6	8.2-10.8	4	siPLS	0.9880	0.7159
4, 6	6.9-8.2 and 9.5-10.8	4	siPLS	0.9869	0.7249
3, 6	5.6-6.9 and 9.5-10.8	3	siPLS	0.9867	0.7422
3, 6, 7	5.6-6.9 and 9.5-12	6	siPLS	0.9923	0.5809
1, 6, 7	3-4.3 and 9.5-12	7	siPLS	0.9917	0.6167
5, 6, 7	8.2-12	4	siPLS	0.9913	0.6178
4, 6, 7	6.9-8.2 and 9.5-12	6	siPLS	0.9909	0.6289

Table 6: Added and found results of the synthetic mixture of glycine (Gly), glutamic acid (Glu) and phenylalanine (Phe) by the PLS and siPLS methods

Added			PLS found			Recovery%			siPLS found			Recovery%		
Gly	Glu	Phe	Gly	Glu	Phe	Gly	Glu	Phe	Gly	Glu	Phe	Gly	Glu	Phe
6.00	5.000	13.00	5.700	4.880	12.69	95	98	98	5.870	4.94	12.70	98	99	98
7.00	7.000	4.800	6.790	6.950	5.010	97	99	104	6.910	6.95	4.520	99	99	94
13.30	13.00	6.000	13.69	13.44	6.090	103	103	102	13.51	13.04	6.560	102	100	109
5.20	14.80	9.800	4.940	15.33	9.240	95	104	94	5.110	14.26	9.640	98	96	98
10.60	9.200	9.000	10.80	9.080	9.320	102	99	104	10.71	8.92	9.160	101	97	102
11.30	5.800	12.00	11.05	6.020	10.96	98	104	91	11.16	5.90	11.39	99	102	95
15.00	11.00	7.200	14.44	11.17	7.860	96	102	109	14.28	10.90	7.830	95	99	109
9.00	14.20	15.20	9.380	13.87	14.97	104	98	98	8.960	13.91	15.59	100	98	103

Table 7: Statistical parameters of the optimal results by using PLS and siPLS to determine glycine (Gly), glutamic acid (Glu) and phenylalanine (Phe) in synthetic mixtures

Parameters	PLS			siPLS		
	Gly	Glu	Phe	Gly	Glu	Phe
pH region	3.0-12.0	3.0-12.0	3.0-12.0	4.3-5.6 and 9.5-12	4.3-5.6	8.2-12
RMSEP	0.34	0.29	0.52	0.28	0.24	0.43
RSEP%	3.3	2.8	5.1	2.8	2.3	4.2
R ²	0.990	0.995	0.983	0.994	0.998	0.985

glycine, glutamic acid and phenylalanine mixtures was carried out. The synthetic mixtures used were first titrated by the same concentration solution of NaOH and the obtained data are then pretreated by the Savitzky-Golay smoothing algorithm with the same parameters as in the calibration set. The titration spectrum of prediction set after Savitzky-Golay smoothing was also shown in Fig. 2. For each amino acid, the concentrations were predicted by the calibration model based on the full pH region PLS and the iPLS or siPLS optimized above. Table 6 shows the concentrations of each amino acid and the obtained results by applying different algorithm to all eight synthetic samples. Table 6 also shows the recovery of glycine, glutamic acid and phenylalanine by different methods in the predict series. As can be seen in the table, the recovery of compounds was also quite acceptable.

Table 7 summarized the obtained values of statistical parameters including RMSEP, RSEP% and *R* for each amino acid in the prediction set. It can be found that all the chemometrics methods that based on factor analysis can achieve good results in resolving the overlapping potentiometric titrations curves of glycine, glutamic acid and phenylalanine in their ternary mixtures, although the solution equilibrium in the acid-base titration procedure is complex. The results also show that the using of selected variables (pH) by iPLS and siPLS method presents better predictions with lower errors in relation to the full pH region PLS model. The results clearly show the successful application of iPLS and siPLS variable selection method to potentiometric titration multivariate calibration as a pre-processing method before the traditional PLS regression.

The purpose of this study is to determine the feasibility of applying siPLS method to potentiometric titration multivariate calibration. It can be seen that the

siPLS method that based a graphically-oriented interface make it more easily to use and interpret the obtained results in potentiometric titration multivariate calibration. The successful application of wavelength selection iPLS and siPLS method that based on graph interface for potentiometric titration indicated that this method may have general implications in other types of titration calibration, such as potentiometric precipitation titration (Ni and Peng, 1995) and complexometric titration (Ni and Wu, 1997). Finally, it must be pointed out that as a calibration method (Assefa *et al.*, 2013; Xu *et al.*, 2012), when large noise and other interference substance are presented in the mixtures which are not included in the calibration set, the method may lead to large error.

CONCLUSION

This study investigated the applicability of iPLS and siPLS algorithm for the variable selection in potentiometric titration multivariate calibration. The potentiometric titration multivariate calibration combined with variable selection by the iPLS and siPLS was tested by the simultaneous determination of glycine, glutamic acid and phenylalanine in aqueous solution. The results show that the using of selected variables by iPLS and siPLS method presents better predictions with lower errors in relation to the full pH region PLS model. The toolbox can automatically locate the informative region for the given substance by using a graphically-oriented interface, make it easily to use and interpret the obtained results. The study proves the feasibility of potentiometric titration with chemometrics method in the simultaneous determination of amino acids in mixtures without the need for preliminary separation steps and also demonstrated the utility of siPLS in titration multivariate calibration for optimal variable selections.

ACKNOWLEDGMENT

The authors acknowledge financial support from the Science and Technology Planning Project of Zhejiang Province (Project no. 2012C37032).

REFERENCES

- Aktaş, A.H. and S. Yaşar, 2004. Potentiometric titration of some hydroxylated benzoic acids and cinnamic acids by artificial neural network calibration. *Acta Chim. Slov.*, 51(2): 273-282.
- Assefa, M.D., A.K. Chaubey, A. Taddess and A.Y. Hibstie, 2013. Relative determination of micronutrients of different species of *Teff* (*Eragrostis*) seeds of Ethiopia origin by calibration free laser induced breakdown spectroscopy techniques. *Adv. J. Food Sci. Technol.*, 5(3): 332-338.
- Cai, J.R., Q.S. Chen, X.M. Wan and J.W. Zhao, 2011. Determination of Total Volatile Basic Nitrogen (TVB-N) content and Warner-Bratzler Shear Force (WBSF) in pork using Fourier transform near infrared (FT-NIR) spectroscopy. *Food Chem.*, 126(3): 1354-1360.
- Cerdán, M., A. Sánchez-Sánchez, J.D. Jordá, M. Juárez and J. Sánchez-Andreu, 2013. Effect of commercial amino acids on iron nutrition of tomato plants grown under lime-induced iron deficiency. *J. Plant Nutr. Soil Sc.*, 176(6): 859-866.
- Chen, H.P., L. Chen, Y.M. Wei and Y.L. Zheng, 2005. Simultaneous determination of amino acids in wheat by reversed phase high performance liquid chromatography with precolumn derivatization. *Chin. J. Anal. Chem.*, 33(12): 1689-1692.
- Chen, Q.S., J. Ding, J.R. Cai and J.W. Zhao, 2012a. Rapid measurement of Total Acid Content (TAC) in vinegar using near infrared spectroscopy based on efficient variables selection algorithm and nonlinear regression tools. *Food Chem.*, 135(2): 590-595.
- Chen, Q.S., J. Ding, J.R. Cai, Z.B. Sun and J.W. Zhao, 2012b. Simultaneous measurement of total acid content and soluble salt-free solids content in Chinese vinegar using near-infrared spectroscopy. *J. Food Sci.*, 77(2): 222-227.
- Christensen, J., L. Nørgaard, R. Bro and S.B. Engelsen, 2006. Multivariate auto fluorescence of intact food systems. *Chem. Rev.*, 106(6): 1979-1994.
- Du, Y.P., Y.Z. Liang, J.H. Jiang, R.J. Berry and Y. Ozaki, 2004. Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares. *Anal. Chim. Acta*, 501(2): 183-191.
- Fang, S., M.Q. Zhu and C.H. He, 2009. Moving window as a variable selection method in potentiometric titration multivariate calibration and its application to the simultaneous determination of ions in Raschig synthesis mixtures. *J. Chemometr.*, 23(3): 117-123.
- Ferrão, M.F., M.S. Viera, R.E.P. Pazos, D. Fachini, A.E. Gerbase and L. Marder, 2011. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. *Fuel*, 90(2): 701-706.
- Ghorbani, R., J. Ghasemi and B. Abdollahi, 2006. Conductometric simultaneous determination of acetic acid, monochloroacetic acid and trichloroacetic acid using orthogonal signal correction-partial least squares. *J. Hazard. Mater.*, 131(1-3): 13-18.
- Goicoechea, H.C. and A.C. Olivieri, 2003. A new family of genetic algorithms for wavelength interval selection in multivariate analytical spectroscopy. *J. Chemometr.*, 17(6): 338-345.
- Hemmateenejad, B. and S. Karimi, 2011. Construction of stable multivariate calibration models using unsupervised segmented principal component regression. *J. Chemometr.*, 25(4): 139-150.
- Hemmateenejad, B., S. Karimi and N. Mobaraki, 2013. Clustering of variables in regression analysis: A comparative study between different algorithms. *J. Chemometr.*, 27(10): 306-317.
- Henchoz, Y., J. Schappler, L. Geiser, J. Prat, P.A. Carrupt and J.L. Veuthey, 2007. Rapid determination of pKa values of 20 amino acids by CZE with UV and capacitively coupled contactless conductivity detections. *Anal. Bioanal. Chem.*, 389(6): 1869-1878.
- Jiang, J.H., R.J. Berry, H.W. Siesler and Y. Ozaki, 2002. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near infrared spectroscopic data. *Anal. Chem.*, 74(14): 3555-3565.
- Leardi, R. and L. Nørgaard, 2004. Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions. *J. Chemometr.*, 18(11): 486-497.
- Liebster, J., K. Kopoldová and M. Dobiášová, 1961. A simplified method for separating amino-acid mixtures from protein hydrolysates on one ion-exchange column during preparation. *Nature*, 191: 1198.
- Lindberg, W. and B. Kowalski, 1988. Evaluation of potentiometric acid-base titrations by partial-least-squares calibration. *Anal. Chim. Acta*, 206: 125-135.
- Li, W., L. Lin and G. Li, 2014. Wavelength selection method based on test analysis of variance: application to oximetry. *Anal. Method.*, 6: 1082-1089.

- Michalowski, T., M. Toporek and M. Rymanowski, 2005. Overview on the Gran and other linearisation methods applied in titrimetric analysis. *Talanta*, 65: 1241-1253.
- Mohabbat, T. and B. Drew, 2008. Simultaneous determination of 33 amino acids and dipeptides in spent cell culture media by gas chromatography-flame ionization detection following liquid and solid phase extraction. *J. Chromatogr. B*, 862(1-2): 86-92.
- Moore, S. and W.H. Stein, 1949. Chromatography of amino acids on starch columns. Solvent mixtures for the fractionation of protein hydrolysates. *J. Biol. Chem.*, 178(1): 53-77.
- Mudiam, M.K.R., C. Ratnasekhar, R. Jain, P.N. Saxena, A. Chauhan and R.C. Murthy, 2012. Rapid and simultaneous determination of twenty amino acids in complex biological and food samples by solid-phase microextraction and gas chromatography-mass spectrometry with the aid of experimental design after ethyl chloroformate derivatization. *J. Chromatogr. B*, 907: 56-64.
- Munck, L., L. Nørgaard, S.B. Engelsen, R. Bro and C.A. Andersson, 1998. Chemometrics in food science—a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometr. Intell. Lab.*, 44(1-2): 31-60.
- Ni, Y.N., 1998. Simultaneous determination of mixtures of acids by potentiometric titration. *Anal. Chim. Acta*, 367(1-3): 145-152.
- Ni, Y.N. and A.G. Wu, 1999. Simultaneous determination of halide and thiocyanate ions by potentiometric precipitation titration and multivariate calibration. *Anal. Chim. Acta*, 390(1-3): 117-123.
- Ni, Y.N. and S. Kokot, 2008. Does chemometrics enhance the performance of electroanalysis? *Anal. Chim. Acta*, 626: 130-146.
- Ni, Y.N. and Y.L. Wu, 1997. Simultaneous determination of mixed metal ions by complexometric titration and multivariate calibration. *Anal. Chim. Acta*, 354: 233-240.
- Ni, Y.N. and Z.H. Peng, 1995. Determination of mixed metal ions by complex metric titration and nonlinear partial least squares calibration. *Anal. Chim. Acta*, 304: 217-222.
- Norgaard, L., A. Saudland, J. Wagner, J.P. Nielsen, L. Munck and S.B. Engelsen, 2000. Interval Partial Least Squares Regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Appl. Spectrosc.*, 54(3): 413-419.
- Ohtani, M., M. Sugita and K. Maruyama, 2006. Amino acid mixture improves training efficiency in athletes. *J. Nutr.*, 136(2): 538-543.
- Ouyang, Q., J.W. Zhao, Q.S. Chen, H. Lin and Z.B. Sun, 2012. Rapid measurement of antioxidant activity in dark soy sauce by NIR spectroscopy combined with spectral intervals selection and nonlinear regression tools. *Anal. Method.*, 4(4): 940-946.
- Redruello, B., V. Ladero, I. Cuesta, J.R. Álvarez-Buylla, M.C. Martín, M. Fernández and M.A. Alvarez, 2013. A fast, reliable, ultra high performance liquid chromatography method for the simultaneous determination of amino acids, biogenic amines and ammonium ions in cheese, using diethyl ethoxymethylenemalonate as a derivatising agent. *Food Chem.*, 139(1-4): 1029-1035.
- Savitzky, A. and M.J.E. Golay, 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36(8): 1627-1639.
- Shamsipur, M., B. Hemmateenejad and M. Akhond, 2002. Multicomponent acid base titration by principal component-artificial neural network calibration. *Anal. Chim. Acta*, 461(1): 147-153.
- Smith, J.T., 1997. Developments in amino acid analysis using capillary electrophoresis. *Electrophoresis*, 18(12-13): 2377-2392.
- Song, X.H., J. Xu and R.Q. Yu, 1993. Artificial neural networks applied to potentiometric titration of multi-component polybasic acid mixtures. *Microchim. Acta*, 111(4-6): 199-206.
- Wakayama, M., N. Aoki, H. Sasaki and R. Ohsugi, 2010. Simultaneous analysis of amino acids and carboxylic acids by capillary electrophoresis-mass spectrometry using an acidic electrolyte and uncoated fused-silica capillary. *Anal. Chem.*, 82(24): 9967-9976.
- Wei, X., N. Xu, D. Wu and Y. He, 2014. Determination of branched-amino acid content in fermented cordyceps sinensis mycelium by using FT-NIR Spectroscopy technique. *Food Bioprocess Tech.*, 7(1): 184-190.
- Xu, L., Z.H. Ye, H.F. Cui, X.P. Yu, C.B. Cai and H.W. Yang, 2012. Calibrating the shelf-life of Chinese flavored dry Tofu by FTIR spectroscopy and Chemometrics: Effects of data preprocessing and nonlinear transformation on multivariate calibration accuracy. *Food Anal. Method.*, 5(6): 1328-1334.
- Zhang, Y., J. Sun, X.T. Yu and T.H. Li, 2005. Simultaneous determination of iron and aluminum with constant potential complexometric titration under non-equilibrium state. *Chin. J. Anal. Chem.*, 33(12): 1764-1766.
- Zhao, M., Y. Ma, L.L. Dai, D.L. Zhang, J.H. Li, W.X. Yuan, Y.L. Li and H.J. Zhou, 2013. A high-performance liquid chromatographic method for simultaneous determination of 21 free amino acids in tea. *Food Anal. Method.*, 6(1): 69-75.