Asian Journal of Chemistry Vol. 18, No. 4 (2006), 2445–2457

Simultaneous Voltammetric Determination of Cysteine, Tyrosine and Tryptophan by Using Principal Component-Artificial Neural Networks

MIR REZA MAJIDI*†, ABOLGHASEM JOUYBAN‡, HAMID ABDOLLAHI** and KARIM ASADPOUR-ZEYNALI

Department of Analytical Chemistry, Faculty of Chemistry
University of Tabriz, Tabriz 51664, Iran
Tel: (98)(411)3393111; E-mail: majidi@tabrizu.ac.ir

Voltammetry was used for the simultaneous determination of cysteine, tyrosine and tryptophan. Each analyte has a distinctive response at the glassy carbon electrode in a nearly neutral solution. The main difficulty encountered in their simultaneous determination is the high degree of overlapping. Extraction of individual analyte concentration from voltammetric responses was achieved using artificial neural networks (ANNs), principal component artificial neural networks (PC-ANNs), principal component regression and partial least squares regression methods. The calibration set was orthogonally designed in order to obtain maximum information from the calibration procedure. The calibration set was used as traning set in artificial neural networks analysis. The different models were used to predict the concentrations of test set. The root mean square error of calibration and root mean square error of prediction were calculated for all models. The results showed that better prediction was achieved with PC-ANN. The number of hidden neurons, learning rate, momentum and the epochs of training were investigated. The combined technique using cyclic voltammetry, PCR and ANN is helpful in the simultaneous detection of mixtures of oxidizable amino acids.

Key Words: Voltammetry, Artificial neural networks, Principal component regression, Partial least squares regression, Amino acid.

INTRODUCTION

Rapid determination of organic compounds in complex samples is frequently required in chemical, food and pharmaceutical industries and also in medical and environmental studies. Although amino acids are readily separated from matrix interference by using high-performance liquid chromatography (HPLC), their

†Research Institute for Fundamental Sciences, Tabriz, Iran.
‡School of Pharmacy, Tabriz University of Medical Sciences, Tabriz 51664, Iran.
**Department of Chemistry, Institute for Advanced Studies in Basic Sciences, Zanjan 45195-159, Iran.

detection is hindered in the absence of a strong chromophore and formation of derivatives is required to enhance the absorbtivity. Indeed, photometric detection of amino acids after chemical derivatization with *o*-phthaladehyde[1], ninhydrin[2] and phenyl isothiocyanate[3] are very sensitive. Nevertheless, different detection strategies, without using derivatizing reagents, when available, are generally preferred for simplicity and economical convenience. Electrochemical detection following liquid chromatography (LC-ED) represents a very attractive detection possibility for underivatized amino acids[4, 5]. Alternatively, several transition metal-based electrodes have been proposed as amperometric sensors for the determination of amino acids at a constant applied potential[6–11].

Meanwhile, the sensitivity is lower due to the higher background signal with electrochemical techniques. In order to improve the sensitivity, selectivity and reproducibility, some chemically modified electrodes (CMEs) have been used for the determination of amino acids[12, 13]. However, applicability of CMEs for these compounds is often limited due to the stability of CMEs and adsorption of chemicals on the electrodes.

A graphite-methacrylate composite and Pt electrodes have been employed for the simultaneous determination of cysteine (Cys), tyrosine (Tyr) and tryptophan (Trp) by voltammetry with PLS and PCR[14, 15]. The interactions of these compounds on electrode surface lead to responses that are not additive. Almost by using high number of principal components these systems can be modelled by using principal component regression (PCR) and partial least squares regression (PLS). It seems that by using non-linear techniques better results may be obtained. This paper employs two non-linear techniques, *i.e.*, artificial neural networks (ANNs) and principal component-ANNs (PC-ANNs), to simultaneous quantification of the three analytes in mixtures.

EXPERIMENTAL

Cysteine, tyrosine, trypotophan and other reagents were of analytical grade supplied by Merck. Deionized water was used for the preparation of all solutions. The background electrolyte solution was prepared from potassium chloride. The buffer solution was prepared from potassium monohydrogen phosphate and potassium dihydrogen phosphate. The pH of these solutions was adjusted to 7.5. A glassy carbon disk electrode with 2 mm diameter was used as working electrode. A platinum wire was employed as counter electrode and a saturated calomel electrode served as the reference electrode and all potentials in the text refer to it (all electrodes obtained from Azar Electrode Co., Iran). Voltammograms were obtained with PGSTAT 20 Autolab potentiostat from ECO Chemie (The Netherland). The GPES software (version 4.5) was used for saving voltammograms. Currents were registered at 5 mV intervals. Data of voltammograms were firstly converted to EXCEL file;, in the next step they moved to MATLAB format. Neural networks were implemented in Matlab (The Math Works, USA) version 6.1, using the Neural Network Toolbox version 4.0.1. PCR and PLS methods were available from the PLS_Toolbox[16].

Electrode preparation

Prior to each experiment, the glassy carbon electrode was polished with 0.05 μm alumina in water slurry using a polishing cloth and deionizied water. The polished electrode was placed in a phosphate buffer solution (pH = 7.5) and the electrochemical activation of the electrode was performed by continuous potential cycling from 0.3 to 0.9 V at a scan rate of 0.1 V s^{-1} until a stable voltammogram was obtained. This electrode was used for electrochemical methods. This procedure was repeated for each voltammogram. Meanwhile, polishing renewed the surface of the electrode.

Multivariate calibrations

The voltammograms were recorded from 0.3 to 0.9 V with a scan rate of 0.050 V s^{-1}. In each experiment, the 15th cycle was recorded. The currents from forward scan were used as multivariate data for further analysis. The first step in the simultaneous determination of three analytes by multivariate calibration methods involves constructing the calibration matrix for the ternary mixture. A calibration set was prepared by using orthogonal array design method, in order to extract maximum quantitative information efficiently. A five level orthogonal array design, denoted by $AO_{25}(5^3)$ was selected in this experiment[17]. Composition of the calibration set is given in Table-1 and composition of test set is given in Table-2. The calibration set was used as traing set in ANN analysis.

Although multivariate calibration methods have been extensively described elsewhere[18–29], a brief description is given below.

Artificial Neural Networks (ANNs)

ANNs have been utilized to solve a wide variety of problems ranging from character and voice recognition to modelling and data mapping. Applications of ANNs in analytical chemistry include modelling nonlinear calibration curves, quantitative analysis of multicomponent systems and data reduction or mapping. ANNs are a parallel computational technique composed of groups of highly interconnected processing elements called neurons. Neurons are arranged in a series of layers. The first layer, which is termed the input layer, receives the experimental information, experimental parameters, topological descriptors, etc. as input. The last layer is the output layer and its neurons produce the output of the network. One or multi-layers of neurons between the input and output layers are called hidden layers. The first step in ANNs is to define the number of neurons in the input and output layers based on the characteristics of the system. The number of neurons in the hidden layer is an adjustable parameter so that it should be optimized. In the next step, the network is trained using experimental data and, in the final step, the network is used for prediction. Among different learning methods in neural network computing, the most popular method is the back-propagation (BP) method and it is often used in chemical studies. In BP feedforward networks, sigmoidal transfer function is used in the hidden layer and sigmoidal and/or linear transfer functions are usually used in the output layer. The mean square error (MSE) is usually used as a criterion for finalizing the learning process and is computed using Eq. (1).

$$MSE = \frac{1}{P \times M} \sum_{p=1}^{P} \sum_{m=1}^{M} (O_{pm} - T_{pm})^2 \qquad (1)$$

where M is the number of neurons in the output layer and P is the number of experimants. The most serious problem in BP-ANN is overfitting where the data points in the training set can be fitted very well (*i.e.*, MSE for training set is very low), but the prediction for data points other than training set is poor and produces high prediction error. Therefore, to evaluate the quality of fitting, training and test sets are used in this work. The MSE from the training set will decrease monotonously with an increase in the number of epochs in the learning procedure. The optimizing process of the number of epochs will be finalized when the prediction error for the test set becomes constant or elevated. Generally, the MSE for the test set become larger if there is overfitting in the learning procedure. In this work, the training process is stopped manually when the mean square error of the test set remains constant after successive epochs.

Principal Component Regression (PCR)

Most measurements are not selective for an analyte of interest in a mixture; in addition, the data also contain noises. In principal component analysis (PCA), the measured data are reduced to contain only the information that is relevant to the system. Its systematic variations are extracted and the information in the many variables is concentrated into a few underlying (latent) variables called principal components. The first step in PCR is to decompose the data matrix into an orthonormal basis set:

$$D_{n,m} = S_{n,q} L^{T}_{q,m} \quad (2)$$

where $D_{n,m}$ contains the n recorded response as rows, each digitized into m data points, $S_{n,q}$ is the score matrix which relates to samples composition, $L_{q,m}$, in which T denotes transpose, is the loading matrix and q is the least of n and m, which usually is n. The second step in PCR is to separate the eigen vectors that account for the systematic variation from those corresponding to noise:

$$D_{n,m} = S_{n,r} L^{T}_{r,m} + E_{n,m} = \hat{D} + E_{n,m} \quad (3)$$

where $\hat{D}$ is the predicted data matrix, $E_{n,m}$ is the residual matrix and r is the number of significant components. It corresponds to the number of compounds that contribute significantly to the measured voltammograms. The third step in PCR is to correlate score matrix S with the concentration matrix C using the following expression:

$$C_{n,l} = S_{n,q} B_{q,l} + F_{n,l} \quad (4)$$

where B is the matrix of regression coefficients which is resolved by using a least square procedure and l is the number of components in the mixture. $F_{n,l}$ is the residual matrix of concentration matrix. The final step in PCR is to predict the concentration of unknown samples from the following equations:

$$S_{unk} = D_{unk} L \quad (5)$$

and

$$C_{unk} = S_{unk} B \quad (6)$$

where the subscript unk refers to the unknown samples.

Principal component-Artificial neural networks

Reducing the number of inputs to a network reduces the training time, repetition and redundancy in the input data and so potentially giving a more accurate network. Principal component analysis is often used to reduce the large number of data to too much smaller PCs. In this work, scores were used as network inputs instead of original data.

Partial least square (PLS)

Partial least square regression can be employed as an alternative to PCR. The PLS method is carried out by decomposition of the data matrix $D^{n, m}$ and concentration matrix $C_{n, l}$:

$$D_{n, m} = S_{n, q} L^{T}_{q, m} + E_{n, m} \tag{7}$$

$$C_{n, l} = U_{n, p} O^{T}_{p, l} + F_{n, l} \tag{8}$$

where n, m, q, l, S, L^T discussed later and U denotes the composition score matrix with n rows and p columns and O stands for the $p \times l$ loading matrix.

The relationship between scores and concentration is obtained from

$$C = SBU^{T} + F \tag{9}$$

where B is the matrix of the regression coefficient obtained by a least square procedure. Once the model is built, it can be used to predict the concentration of unknown samples.

Leave-one-out cross-validation method was used to select the number of latent variables to be used in PCR and PLS. The optimum number of latent variables was chosen when a minimum value of PRESS function calculated using the following expression is achieved:

$$PRESS(k) = \sum_{i=1}^{Samples} (C_{i_{true}} - \hat{C}_{i_{cal}}(k))^2 \tag{10}$$

where k refers to the number of latent variables considered $C_{i_{true}}$ is the real concentration of analyte in the sample i and $\hat{C}_{i_{cal}}$ is the calculated concentration by multivariate calibration methods using k factors.

The relative performance of different models was evaluated by the parameters root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) which were calculated using the following expressions:

$$RMSEC = 100 \times \left(\frac{\Sigma (C_{exp\,cal} - C_{cal})^2}{\Sigma (C_{cal})^2} \right)^{1/2} \tag{11}$$

$$RMSEP = 100 \times \left(\frac{\Sigma (C_{exp\,pre} - C_{pre})^2}{\Sigma (C_{pre})^2} \right)^{1/2} \tag{12}$$

where $C_{exp\,cal}$ and C_{cal} are the experimental and predicted concentrations for calibration set, $C_{exp\,pre}$ and C_{pre} are the experimental and predicted concentrations for test set, respectively.

RESULTS AND DISCUSSION

First experimental studies indicate that repeated voltammograms shows a decrease in the current intensity from scan to scan and after approximately 10 successive cycles, a steady state was attained. From this point, the cyclic voltammograms were highly reproducible from cycle to cycle. Therefore the cyclic voltammograms of each particular sample were taken after 15 cycles in order to ensure the reproducibility of data. The individual voltammograms obtained for three analytes were shown in Fig. 1. The precision of measurements was studied

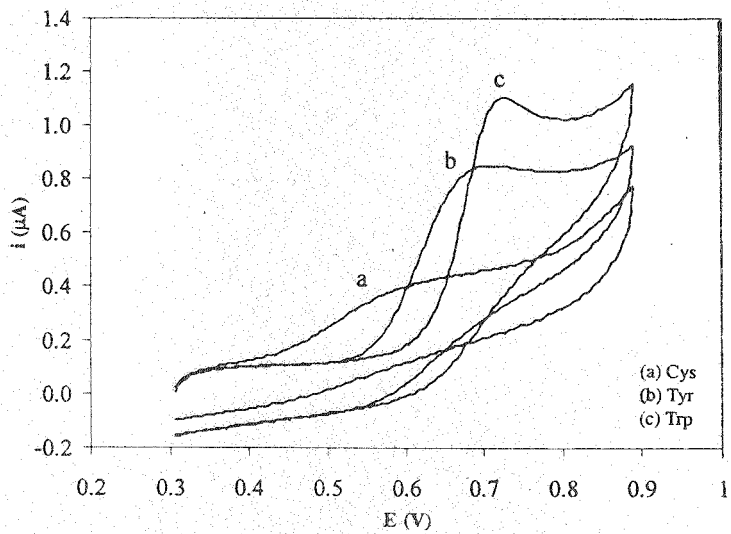

Fig. 1. Cyclic voltammograms of Cys, Tyr, Trp in 0.1 M KCl and phosphate buffer 0.1 M (pH = 7.5). (Concentrations of amino acids were 10^{-4} M)

with 5 replicates of 10^{-4} M for each analyte. The relative standard deviations for E_p and i_p were 1.3 and 4.7% for Cys, 1.5 and 8.7% for Tyr and 1.3 and 1.6% for Trp, respectively, which were sufficient for these kinds of measurements. Fig. 2 shows cyclic voltammograms of amino acid solutions at different concentrations from 7.80 to 122.2 μM in 0.1 M KCl and phosphate buffer (pH = 7.5). As shown in Fig. 1, the voltammograms were highly overlapped and a selective potential is not able to determine such amino acids by using classical univariate calibration. The additivity of the cyclic voltammograms was investigated from the study of various mixtures of amino acids. The results showed that the voltammogram of the mixture (Fig. 3a) is quite different from that of the sum of the voltammetric signals of three single amino acids (Fig. 3b). This means that the additivity of three compounds can be ascribed to the analyte-analyte interaction effects during oxidation. Therefore, the position of peaks and/or shape of voltammograms could be changed with these effects. This is typical fouling during measurement which could be caused by, for example, products of cystein oxidation at the first peak being adsorbed to the electrode and thereby reducing the surface area for the oxidation of tyrosin and tryptophan at higher potentials.

Combination of cyclic voltammetry and chemometrics may allow the resolution and simultaneous determination of Cyc, Tyr and Trp in mixtures without using a physical separation. The used chemometric techniques in this work were PCR, PLS, ANN and PC-ANN. The calibration sets for these methods were the same as

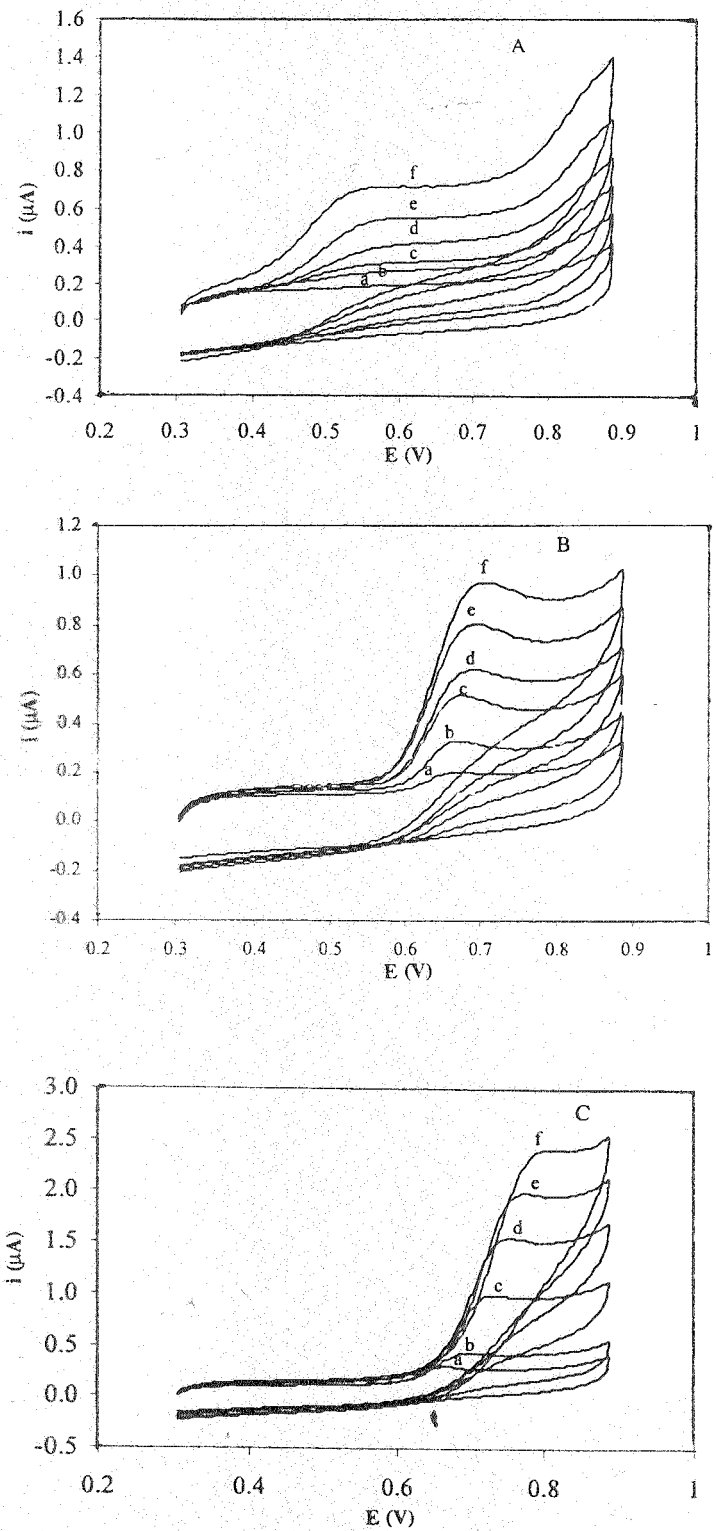

Fig. 2. Cyclic voltammograms of amino acid solutions at different concentrations: (A) Cys, (B) Tyr, (C) Trp in 0.1 M KCl and phosphate buffer 0.1 M (pH = 7.5): (a) 7.80, (b) 22.06, (c) 49.10, (d) 77.40, (e) 101.50, (f) 122.20 μM

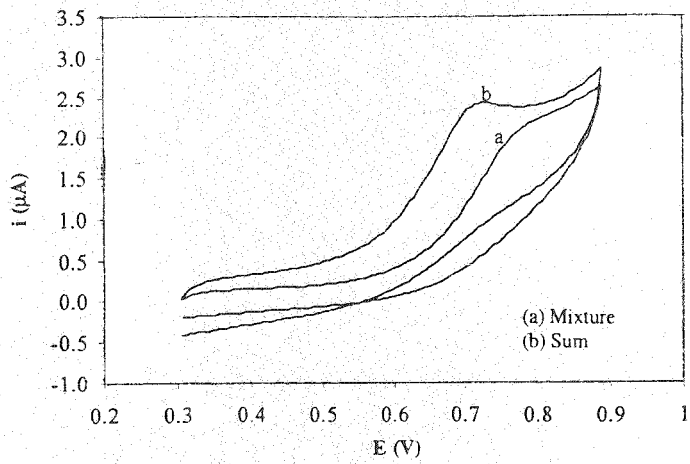

Fig. 3. Cyclic voltammograms of a mixture of amino acids: (a) The sum of the signals of each component, (b) ternary mixture in 0.1M KCl and phosphate buffer 0.1 M (pH = 7.5)

those given in Table-1. The compositions were orthogonally designed in order to obtain maximum information from calibration procedure.

TABLE-1
COMPOSITIONS OF THE CALIBRATION SAMPLES ($\times 10^{-4}$)

Sample No.	Cys	Tyr	Trp	Sample No.	Cys	Tyr	Trp
1.	0.60	0.20	0.12	14.	2.40	5.00	0.24
2.	0.60	0.60	0.24	15.	2.40	7.50	0.48
3.	0.60	1.00	0.48	16.	5.00	0.20	0.48
4.	0.60	5.00	0.80	17.	5.00	0.60	0.80
5.	0.60	7.50	0.88	18.	5.00	1.00	0.88
6.	1.00	0.20	0.88	19.	5.00	5.00	0.12
7.	1.00	0.60	0.12	20.	5.00	7.50	0.24
8.	1.00	1.00	0.24	21.	7.50	0.20	0.24
9.	1.00	5.00	0.48	22.	7.50	0.60	0.48
10.	1.00	7.50	0.80	23.	7.50	1.00	0.80
11.	2.40	0.20	0.80	24.	7.50	5.00	0.88
12.	2.40	0.60	0.88	25.	7.50	7.50	0.12
13.	2.40	1.00	0.12				

To obtain optimum number of principal components a plot of the PRESS against the number of factors for each individual component indicates a minimum value of optimal number of factors. For finding the smallest model (fewest number of factors) the F statistics was used to carry out the significance determination[30]. Basically, in PCR and PLS three principal components should be sufficient for these three component systems. However, it cannot be analyzed for Cys, Tyr and Trp by PCR and PLS using only three principal components. Improvements were obtained when 7, 9 and 10 principal components were

included for Cys, Tyr and Trp respectively in PCR analysis. But in PLS analysis 7, 8 and 9 latent variables were obtained. Three factors were attributed to each analyte and additional factors might be attributed to interactions and/or shifts in the shapes of voltammograms. Obtained results from PCR and PLS are given in Table 3.

TABLE-2
COMPOSITIONS OF THE TEST SAMPLES ($\times 10^{-4}$)

Sample No.	Cys	Tyr	Trp
1.	1.4	6.0	0.16
2.	2	2.0	0.60
3.	2	0.4	0.30
4.	6	0.4	0.60
5.	6	0.8	0.40
6.	7	2.0	0.40
7.	7	6.0	0.16

TABLE-3
CALIBRATION AND PREDICTION ERRORS FOR PCR, PLS, ANN AND PC-ANN ANALYSIS

	PCR	PLS	ANN	PC-ANN
Cys:				
RMSEC	4.67	4.34	1.57	0.23
RMSEP	5.20	5.01	25.69	2.06
Tyr:				
RMSEC	2.87	2.52	1.31	0.32
RMSEP	8.73	7.69	5.85	4.35
Trp:				
RMSEC	5.28	3.98	3.87	1.72
RMSEP	3.36	3.52	10.68	3.09

From this point, what seems a non-linear multivariate calibration technique is necessary. Therefore an alternative approach is to use neural networks. In this work feed forward neural networks were used. The used networks consisted of an input layer. The currents at the 10 mV intervals for forward scan were used as inputs. A hidden layer with sigmoidal transfer function was selected. The output was simply the estimated concentration. Sigmoidal transfer function was used in output layer. Before training the input (*i.e.*, currents) and target values (*i.e.*, concentrations) were normalized between 0.1 and 0.9.[31] In order to reduce the size of network each concentration was separately estimated, so that all calculations only had one output, to be comparable to the PLS and PCR results, which were separately performed for each compound. During the training phase for neural network analysis, extensive experimentation was performed to define appropriate network parameters. Although the selection of neurons in the hidden layers in a back-propagation network is empirical, it is recognized that this choice can have a significant effect on network performance. A large number of hidden

neurons can provide more predicting power, but the network will require more computation time and may also suffer in terms of its ability to generalize for an unknown data set[32]. Different numbers of neurons in the hidden layer (from 1 to 10) were tested at an arbitrary learning rate and momentum. The performance of the network stabilized after the inclusion of an adequate number of hidden neurons (more than 5) when the training reached sufficient epochs. Therefore 5 neurons in hidden layer were selected as optimum. Then, learning rate and momentum were optimized in a similar way. Optimum values of learning rate and momentum were 0.1 and 0.8, respectively.

The optimum number of epochs for each component was investigated and after 2500 epochs was found as optimum (Fig. 4). The results obtained from

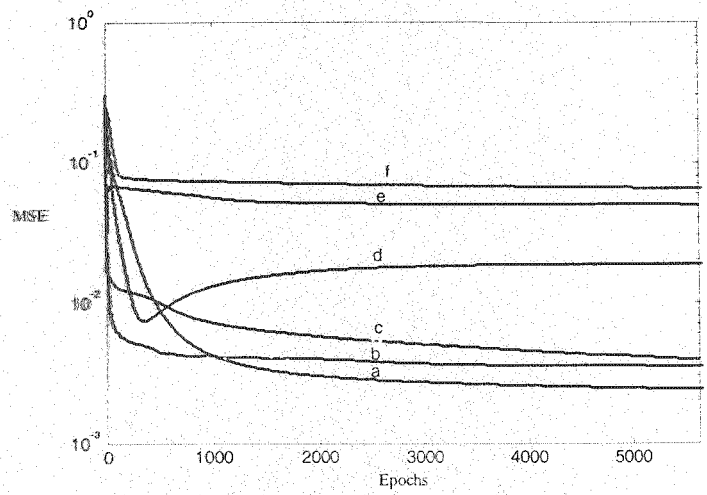

Fig. 4. Plot of MSE as the function of the number of epochs for three amino acids: (a, d) Tyr, (b, e) Cys, (c, f) Trp. Number of neurons in hidden layer were five. (a–c) and (d–f) are training and prediction respectively. (MSE values are for normalized data)

simultaneous analysis of amino acids by ANNs method are given in Table-3.

Including 60 potentials as inputs and 5 neurons in hidden layer would result in a large number (311) of weights when bias and hidden neurons are included, which is clearly unjustified by present data set consisting only of 25 samples. Hence the data was reduced using PCA. As noted above, only 7, 9, 10 principal components were kept as the input to PC-ANN for Cys, Tyr and Trp respectively. In PC-ANNs, different numbers of neurons in the hidden layer (from 1 to 10) were tested using a learning rate of 0.1, a momentum term of 0.8 and 10000 iterations. Fig. 5 illustrates the relationship between the network error for different numbers of neurons in the hidden layer for Cys, Tyr and Trp. The numbers of neurons in the hidden layer at the minimum of these curves were selected as the optimum number (Fig. 5). Therefore 4, 5 and 6 neurons in hidden layer were selected as optimum number of neurons for Cys, Tyr and Trp respectively. The optimum number of epochs for PC-ANN was also obtained. Fig. 6 shows the plots of MSE as a function of number of epochs for Cys, Tyr and Trp components. As observed, by continued training beyond 1000 epochs for analytes Cys, Trp and 10000 epochs for Trp, the performance of the networks was stabilized.

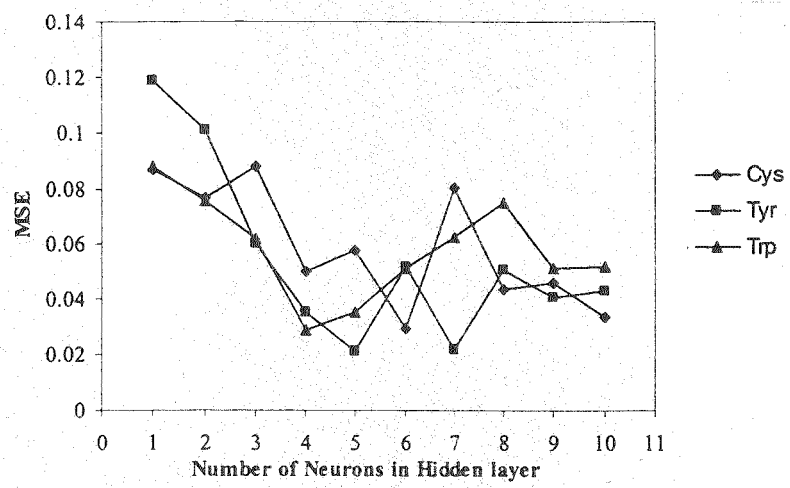

Fig. 5. The relationship between number of neurons in hidden layers *vs.* MSE. (The MSE values are for normalized data)

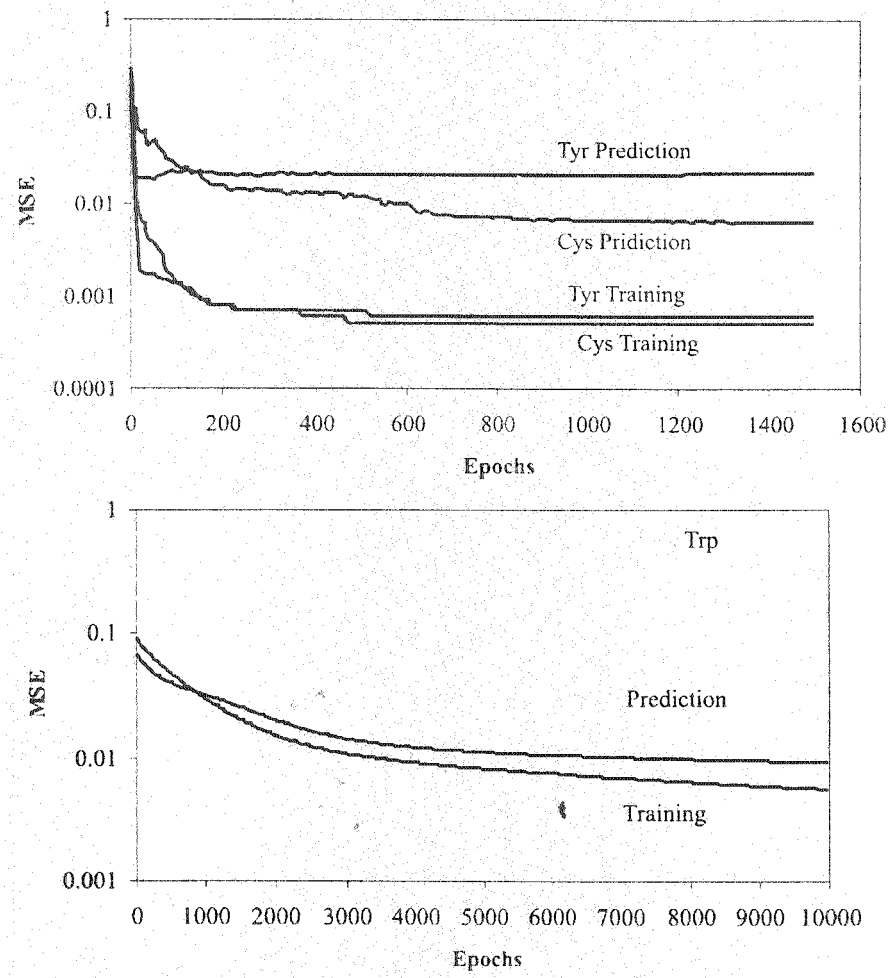

Fig. 6. Plots of MSE as the function of the number of epochs. Number of neurons in hidden layer for Cys, Tyr and Trp were five. (MSE values are for normalized data)

Table-3 gives the prediction and calibration errors of ANN and PC-ANN modeling with optimum numbers of hidden nodes. From Table-3, it is considered that the PC-ANN gives better results than ANN, PCR and PLS analyses.

In cyclic voltammetry, the relationships between currents and concentrations in all potentials except peak potential were not linear. Especially in the presence of interactions between analytes these relationships are very complex.

Conclusion

Overlapping peaks in voltammetric analysis can be handled by multivariate calibration techniques in order to extract quantitative analysis. But in the presence of interactions between analytes the response of the mixture is quite different from the sum of signals. This problem is poorly resolved by using high number of latent variables. For these systems, using of nonlinear techniques leads to better results. In this study, PC-ANN exhibited better prediction ability than conventional multivariate techniques, such as ANN, PLS and PCR. In addition to its formal application in qualitative problems, it can also be used to predict quantitatively the concentrations of mixtures, as shown here. The methods proposed here can serve for the analysis of oxidizable amino acids, where increased speed in obtaining the results is crucial and may become an attractive alternative to chromatographic techniques, which are more tedious and time-consuming.

ACKNOWLEDGEMENTS

This work has been supported by Research Institute for Fundamental Sciences, Tabriz, Iran. The authors wish to thank Professor S.M. Golabi for very useful discussions and suggestions.

REFERENCES

1. J. Abecassis, C. David-Eteve and A. Soun, *J. Liq. Chromatogr.*, **8**, 135 (1985).
2. P. Hamilton, *Anal. Chem.*, **35**, 2055 (1963).
3. S. Black and M. Coon, *Anal. Biochem.*, **121**, 281 (1982).
4. J. Polta and D.C. Johnson, *J. Liq. Chromatogr.*, **6**, 1727 (1983).
5. L.E. Welch, W.R. LaCourse, D.A. Mead (Jr.), D.C. Johnson and T. Hu, *Anal. Chem.*, **61**, 555 (1989).
6. A.P. Clarke, P. Jandick, R.D. Rocklin, Y. Liu and N. Avdalovic, *Anal. Chem.*, **71**, 2774 (1999).
7. W.T. Kok, H.B. Hanekamp, P. Bos and R.W. Frei, *Anal. Chim. Acta*, **142**, 31 (1982).
8. K. Stulik, V. Pacakova, M. Weingart and M. Podolak, *J. Chromatogr.*, **367**, 311 (1986).
9. J.B. Kalif and C.O. Huber, *Anal. Chim. Acta*, **176**, 275 (1985).
10. P. Luo, F. Zhang and R.P. Baldwin, *Anal. Chem.*, **63**, 1702 (1991).
11. Y. Xie and C.O. Huber, *Anal. Chem.*, **63**, 1714 (1991).
12. I.G. Casella, T.R.I. Cataldi, A. Guerrieri and E. Desimoni, *Anal. Chim. Acta*, **335**, 217 (1996).
13. I.G. Casella, M. Gatta and T.R.I. Cataldi, *J. Chromatogr. A*, **878**, 57 (2000).
14. J. Saurina, S. Hernendez-Cassou, E. Fàbregas and S. Alegret, *Anal. Chim. Acta*, **405**, 153 (2000).