ELSEVIER

Journal of Chromatography A, 927 (2001) 211–218

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Prediction of electrophoretic mobilities of sulfonamides in capillary zone electrophoresis using artificial neural networks

M. Jalali-Heravi*, Z. Garkani-Nejad

Department of Chemistry, Sharif University of Technology, PO Box 11365-9516, Tehran, Iran

Received 6 February 2001; received in revised form 22 May 2001; accepted 5 July 2001

Abstract

Artificial neural networks (ANNs) were successfully developed for the modeling and prediction of electrophoretic mobility of a series of sulfonamides in capillary zone electrophoresis. The cross-validation method was used to evaluate the prediction ability of the generated networks. The mobility of sulfonamides as positively charged species at low pH and negatively charged species at high pH was investigated. The results obtained using neural networks were compared with the experimental values as well as with those obtained using the multiple linear regression (MLR) technique. Comparison of the results shows the superiority of the neural network models over the regression models.

Keywords: Neural networks, artificial; Electrophoretic mobility; Regression models; Chemometrics; Quantitative structure–activity relationships; Mathematical modelling; Sulfonamides

1. Introduction

Capillary electrophoresis (CE) has become a powerful separation technique and is widely applied to a variety of analytical samples. During method development in CE, the analysts generally have to employ a large number of experiments, which is often time-consuming. Quantitative structure–retention relationships (QSRRs) have been extensively used to explain separation mechanisms and to predict retention behavior in analytical chemistry. However, only a few research groups have investigated the quantitative correlation between the analytical parameters and the responses in CE. For example, Fu and Lucy have developed empirical expressions for the prediction of electrophoretic mobility of monoamines and aliphatic carboxylic acids [1,2]. They have correlated the mobility of the analytes with the molecular mass, molar volume and dissociation constant using non-linear equations. Also, Liang and coworkers have studied the correlation between the electrophoretic mobility of 13 flavonoids and topological indices [3]. The separation of sulfonamides by CE was usually conducted in the mode of capillary zone electrophoresis (CZE) or micellar electrokinetic chromatography (MEKC) [4,5]. These compounds can be separated at an optimum pH using various types of buffers. For sulfonamides, two dissociation equilibria exist as shown in Fig. 1. Therefore, depending on the pH of the buffer employed, sulfonamides can be separated by CZE either

*Corresponding author. Tel.: +98-21-600-5718; fax: +98-21-601-2983.

E-mail address: jalali@sina.sharif.ac.ir (M. Jalali-Heravi).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)01099-8

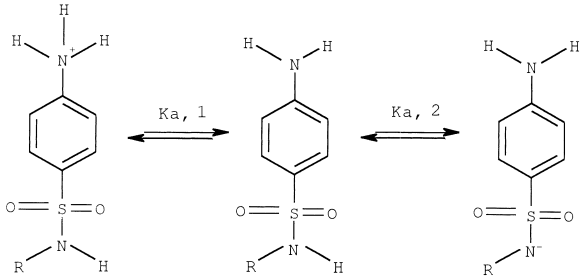

Fig. 1. Dissociation equilibria of sulfonamides involving $pK_{a,1}$ and $pK_{a,2}$.

as negatively charged, deprotonated species or as positively charged, protonated species [6].

Since sulfonamides are commonly used to treat bacterial infections related to the respiratory, intestinal and urinary tracts, theoretical study of these compounds seems to be useful. On the other hand, the electrophoretic mobility of an ion is of fundamental importance in CE, therefore the main goal of this work was investigation of the influences of different parameters on electrophoretic mobility of sulfonamides as negatively and positively charged species. As a first step in this work, three separate multiple linear regression (MLR) models were developed for the prediction of electrophoretic mobility of different forms of sulfonamides. Then, for inspection of non-linear interactions between different parameters in the MLR models, two separate artificial neural network (ANN) models were generated for the prediction of electrophoretic mobility of some sulfonamides.

2. Theory

A detailed description of the theory behind a neural network has been adequately described by different researchers [7–10]. Among different methods of training of neural networks, the back-propagation (BP) technique is the most popular and is often used in analytical applications [11–14]. An artificial neural network consists of a number of ''neurons'' or ''hidden units'' that receive data from the outside, process the data, and output a signal. A ''neuron'' is essentially a regression equation with a non-linear output. When more than one of these neurons is used, non-linear models can be fitted. The back propagation network receives a set of inputs, which are multiplied by each neuron's weights. These products are summed for each neuron and a non-linear transfer function is applied. The transformed sums are then multiplied by the output weights where they are summed a final time, transformed, and interpreted. Since a back-propagation network is a supervised method, the desired output must be known for each input vector so an error can be calculated. This error is propagated backwards through the network, adjusting the weights so that the next time the network sees the same input patterns, it will come closer to the desired output. The patterns are repeated many times until the network learns the relationship.

In this work, the electrophoretic mobility of sulfonamides as negatively and positively charged species and negatively charged species in the presence of the micelle of sodium dodecyl sulfate (SDS) was calculated using ANNs. The inputs of the networks were selected by developing MLR models.

3. Experimental

3.1. Data set

The electrophoretic mobility of 13 sulfonamides as positively charged species ($H2A^+$) and negatively charged species (A^-) was taken from Ref. [6] and electrophoretic mobility of anionic sulfonamides in the presence of SDS micelle was taken from Ref. [5]. The structure of compounds studied in this work is given in Fig. 2 and the electrophoretic mobilities are given in Tables 1 and 2, respectively. As can be seen from Fig. 2, the substituents consist of five and six member heterocyclic compounds with different atoms of oxygen, nitrogen and sulfur. In addition, there are different electron donor and electron acceptor species, such as methyl and chlorine groups on these heterocyclic rings. Electrophoretic mobilities for cationic sulfonamides ranged from 0.71×10^{-4} to 1.94×10^{-4} ($cm^2\ V^{-1}\ s^{-1}$), for anionic sulfonamides from -2.13×10^{-4} to -2.63×10^{-4} ($cm^2\ V^{-1}\ s^{-1}$), and for anionic sulfonamides in the presence of the SDS from -2.01×10^{-4} to -2.45×10^{-4} ($cm^2\ V^{-1}\ s^{-1}$).

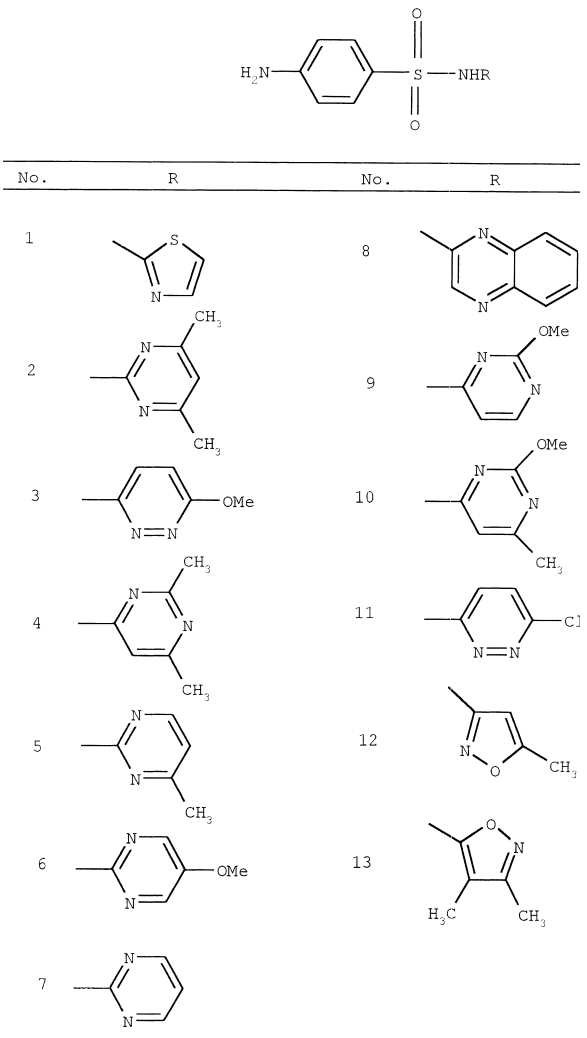

Fig. 2. Structures of the sulfonamides studied in this work.

3.2. Regression analysis

As first step for developing of the regression model the numerical parameters (descriptors) should be generated. A total of 47 descriptors were calculated for each sulfonamide as cationic and anionic forms. These descriptors can be classified into four major groups of topological, geometric, electronic and physico-chemical parameters. Topological descriptors include fragment descriptors and molecular connectivity indexes that were estimated from two-dimensional representations of the molecules (15 descriptors). Geometric descriptors consist of van der Waals volume, surface area, shadow areas, shape factor, etc. (18 descriptors). Electronic descriptors include partial charges of the most negative and the most positive atoms, dipole moment, ionization potential (12 descriptors), and finally physico-chemical descriptors include polarizability and p-function of dissociation constant (two descriptors). In order to calculate the geometric and electronic descriptors, the three-dimensional structures of compounds as anionic (A^-) and cationic ($H2A^+$) were optimized using AM1 Hamiltonian implemented in the MOPAC program [15] and Hyperchem package [16]. For calculating of the volume of the molecules, the algorithm given by Stouch and Jurs was used and its program was written in FORTRAN 77 in our laboratory [17]. Some of the descriptors encoded similar information for the compounds. It was therefore desirable to test each descriptor and eliminate some of those that show high correlation ($R>0.90$) with each other. By using this criterion, ten out of 47 original descriptors were eliminated. Then, multivariate linear models were generated using SPSS/PC software package [18]. The best MLR model is one that has high R- and F-values, low standard deviation and high ability for prediction. Three of the best models for different forms of sulfonamides (two anionic and one cationic forms) are presented in Table 3.

3.3. Neural network generation

The ANN program was written in FORTRAN 77 in our laboratory. The descriptors appearing in the MLR models were used as inputs for generation of the networks. A three-layer network with a sigmoidal transfer function was designed. The initial weights were randomly selected between -0.3 and $+0.3$. Before training, the input and output values were normalized between 0.1 and 0.9. The number of neurons in the hidden layer, learning rate and momentum were optimized. The standard error of training (SET) was plotted versus the number of iterations for different number of neurons at the hidden layer. The number of neurons at the hidden layer with the minimum value of SET was selected as the optimum number. Then, learning rate and momentum were optimized in a similar way.

Table 1
Experimental and ANN and MLR calculated values of electrophoretic mobilities together with the values of the descriptors appearing in the model for the cationic sulfonamides

No.	Compound	Descriptors[a]			Electrophoretic mobility[b]		
		ΔH	pK	SA	μ_{ANN}	μ_{MLR}	μ_{EXP}
1	Sulfathiazole	243.691	2.08	243.882	1.535	1.319	1.530
2	Sulfamethazine	163.060	2.28	292.104	1.472	1.499	1.500
3	Sulfamethoxypyridazin	155.183	2.09	278.172	1.333	1.279	1.320
4	Sulfisomidine	173.891	2.68	292.428	1.940	2.182	1.940
5	Sulfamerazine	170.180	2.17	270.828	1.471	1.383	1.460
6	Sulfamete	140.830	1.87	280.008	0.981	0.905	0.960
7	Sulfadiazine	177.159	2.10	249.282	1.334	1.320	1.330
8	Sulfaquinoxaline	197.635	1.86	293.400	0.908	1.045	0.900
9	Sulfamonomethoxine	142.627	1.98	278.010	1.166	1.158	1.200
10	Sulfadimethoxine	133.860	1.87	309.546	0.897	0.827	0.900
11	Sulfachloropyridazine	183.379	1.90	263.700	0.996	1.074	1.000
12	Sulfamethoxazole	167.916	1.83	260.676	0.783	1.007	0.740
13	Sulfisoxazole	157.043	1.66	278.442	0.707	0.651	0.710

[a] Definitions of the descriptors are given in the text.
[b] μ is mobility in (10^4 cm^2 V^{-1} s^{-1}).

4. Results and discussion

The experimental and calculated values of the electrophoretic mobilities using the MLR and ANN models for a collection of 13 cationic and anionic sulfonamides ($H2A^+$, A^-) and also anionic sulfonamides in the presence of SDS micelle, are given in Tables 1 and 2, respectively. The values of the

Table 2
Experimental and ANN and MLR calculated values of electrophoretic mobilities together with the values of the descriptors appearing in models 1 and 2 for the anionic sulfonamides

No.[a]	Descriptors[b]			Electrophoretic mobility[c]					
	ΔH	$PPCH$	SA	Anionic (1)[d]			Anionic (2)[e]		
				μ_{ANN}	μ_{MLR}	μ_{EXP}	μ_{ANN}	μ_{MLR}	μ_{EXP}
1	−38.676	2.994	232.650	−2.634	−2.570	−2.630	−2.452	−2.352	−2.450
2	−33.983	2.998	283.950	−2.155	−2.193	−2.170	−2.050	−2.044	−2.050
3	−49.012	2.993	273.366	−2.295	−2.314	−2.250	−2.114	−2.142	−2.130
4	−39.583	2.992	286.758	−2.163	−2.204	−2.130	−2.034	−2.039	−2.030
5	−28.375	2.996	265.536	−2.292	−2.299	−2.310	−2.131	−2.176	−2.120
6	−58.408	2.997	276.768	−2.266	−2.290	−2.280	−2.112	−2.115	−2.110
7	48.732	2.989	247.176	−2.332	−2.313	−2.330	−2.242	−2.277	−2.240
8	−8.605	2.984	288.918	−2.207	−2.191	−2.210	−2.027	−1.989	−2.030
9	−61.776	2.991	275.796	−2.381	−2.329	−2.390	−2.133	−2.146	−2.120
10	−95.875	2.990	304.686	−2.195	−2.200	−2.190	−2.014	−1.953	−2.010
11	−27.341	2.989	259.056	−2.353	−2.388	−2.360	−2.182	−2.223	−2.180
12	−39.995	2.976	254.142	−2.471	−2.608	−2.470	−2.286	−2.281	−2.280
13	−51.892	2.996	265.158	−2.427	−2.368	−2.430	−2.196	−2.191	−2.220

[a] The numbers refer to the compounds given in Table 1.
[b] Definitions of the descriptors are given in the text.
[c] μ is mobility in (10^4 cm^2 V^{-1} s^{-1}).
[d] Anionic sulfonamides in the absence of the micelle.
[e] Anionic sulfonamides in the presence of the micelle.

Table 3
Specifications of the multiple linear regression models

Model	Descriptor	Notation	Coefficient	Mean effect
Cationic	Heat of formation	ΔH	+0.001 (±0.001)	0.184
	P-function of dissociation constant	pK	+1.335 (±0.124)	2.708
	Surface area	*SA*	−0.002 (±0.002)	−0.611
	Constant		−1.091 (±0.714)	
Anionic (1)	Heat of formation	ΔH	+0.002 (±0.001)	−0.077
	Most positive partial charge	*PPCH*	+3.655 (±2.453)	10.928
	Surface area	*SA*	+0.008 (±0.001)	2.076
	Constant		−15.252 (±7.330)	
Anionic (2)	Heat of formation	ΔH	+0.001 (±0.000)	−0.039
	Most positive partial charge	*PPCH*	+1.800 (±1.537)	5.382
	Surface area	*SA*	+0.007 (±0.001)	1.855
	Constant		−9.351 (±4.591)	

descriptors appearing in the models are also given in these tables.

4.1. Regression analysis

Multiple linear regressions were performed using all of compounds in the data set. After regression analysis, three of the best models for cationic and anionic sulfonamides were chosen and are presented in Table 3. It can be seen from this table that three descriptors have appeared in each model. The first model for positively charged species consists of surface area (*SA*), heat of formation of cations (ΔH) and p-function of dissociation constant (pK). The second and third models for negatively charged species consist of surface area (*SA*), heat of formation of anions (ΔH) and maximum positive partial charge on the anions (*PPCH*). Mean effects of these parameters on the mobilities are also given in Table 3. It can be seen from this table that for cationic sulfonamides, pK has a larger mean effect than *SA* and ΔH, whereas for anionic sulfonamides, the parameter of *PPCH* has the largest mean effect. Mean effect and coefficient of *SA* in the models indicate that for cationic sulfonamides, positive charge density on the surface decreases as surface area increases and therefore the motion of cation toward the cathode and its electrophoretic mobility decreases. However for anionic sulfonamides, with larger surface area, negative charge density on the surface is less and motion of the anion toward the anode is slower and its electrophoretic mobility is lower. On the other hand, in agreement with the Offord, the mobility is inversely proportional to the surface area [19]. This is due to the fact that the frictional coefficient may arise from the shear across a small element of liquid close to the migration molecule that can be a function of the surface area of the molecule. Also, it is generally assumed that electrophoretic mobility is given by $\mu_e = q/6\pi\eta r$, where r is the hydrodynamic radius of the analyte [20]. It can be seen from this equation that electrophoretic mobility is inversely proportional to the radius of the molecules. Heat of formation of the species (ΔH) has also appeared in all three models. Inspection of Table 3 reveals that the mean effect of the parameter of ΔH is very small in all three models compared with the remaining parameters appearing in the models. This is due to the fact that ΔH has no significant effect on the mobilities. It is noteworthy that inclusion of this parameter considerably improves the statistics of the models. The third parameter in the anionic models is the most positive partial charge on the analyte (*PPCH*). This parameter plays an important role in migration of anions in the presence of an electric field. As the value of *PPCH* increases, the electrophoretic mobility of anions toward the anode decreases. In agreement with the experiment, p-function of dissociation constant for the cationic sulfonamides shows a linear relationship with the electrophoretic mobility. It is noteworthy that a number of topological descriptors were investigated for the generation of the models in this work, but these parameters, in agreement with previous

work [1], did not contribute to the mobilities and did not appear in the models.

The descriptors appearing in the MLR model for anionic sulfonamides in the presence of the SDS micelle and corresponding coefficients are similar to those in the absence of the micelle. It should be noted that this micelle is an anionic one, therefore the direction of the electroosmotic and electrophoretic mobilities are the same as in the absence of the micelle.

Because the number of molecules included in the data set was small, the cross-validation method [21] was used to evaluate the ability of the selected models in predicting the electrophoretic mobility of the sulfonamides. In this method, since three descriptors appeared in the models, three species were removed randomly from the data set each time and the model was generated with the remaining molecules. Then, the electrophoretic mobility of the removed molecules was predicted using the generated model. This procedure was continued until each analyte was predicted once. The values of R^2_{cv} obtained using the cross-validation method for different groups of compounds are given in Table 4. These results in agreement with the experiment, indicate that size of the analyte and its electronic properties play an important role in the migration of the sulfonamides in capillary electrophoresis.

4.2. Neural network analysis

A 3-4-2 ANN was generated using the three descriptors of *SA*, ΔH and *PPCH* appearing in the MLR models of 1 and 2 as inputs for the anionic sulfonamides and a 3-6-1 ANN was generated using the three descriptors of *SA*, ΔH and p*K* appearing in the MLR model for the cationic sulfonamides. In order to optimize the number of nodes in the hidden layer, several training sessions were conducted with different numbers of hidden nodes. The value of SET was calculated for a total of 30 000 iterations. The calculated values of SET were plotted against the number of iterations, from which the number of hidden nodes with minimum value of SET was chosen. The results obtained indicate that four nodes in hidden layer were sufficient for a good performance of the network for the anionic sulfonamides. Learning rate and momentum were optimized in a similar way. Because the three parameters appearing in the two MLR models were identical, the network for anionic sulfonamides has three inputs and two outputs for the two series of data. The number of hidden nodes for the cationic network was obtained to be six. This network also consists of three inputs, the same as three parameters in the MLR model, and one output for one series of data.

Because of the small data set, the cross-validation method was used to evaluate the prediction ability of the generated networks. In this method, similar to that for the MLR model, three species were removed randomly from the data set each time and the network was trained with the remaining data. Then the electrophoretic mobilities of the removed species were predicted using the trained network. It is noteworthy that training of the network was stopped when the standard error of prediction (SEP) started to increase, i.e. when overtraining begins. A typical plot of SET and SEP variations versus the number of iterations is given in Fig. 3 for one group of cationic sulfonamides in the cross-validation procedure. It can be seen that while the SET for the training set continues to decrease during the progression of

Table 4
R^2_{cv} values of cross-validation for MLR and ANN methods

Group	No. of removed compounds	Cationic		Anionic (1)		Anionic (2)	
		R^2_{MLR}	R^2_{ANN}	R^2_{MLR}	R^2_{ANN}	R^2_{MLR}	R^2_{ANN}
1	3	0.952	0.999	0.513	0.991	0.828	0.995
2	3	0.890	0.995	0.792	0.972	0.946	0.966
3	3	0.625	0.994	0.661	0.905	0.775	0.885
4	2	0.833	0.945	0.897	0.994	0.689	0.999
5	2	0.686	0.999	0.913	0.999	0.502	0.999

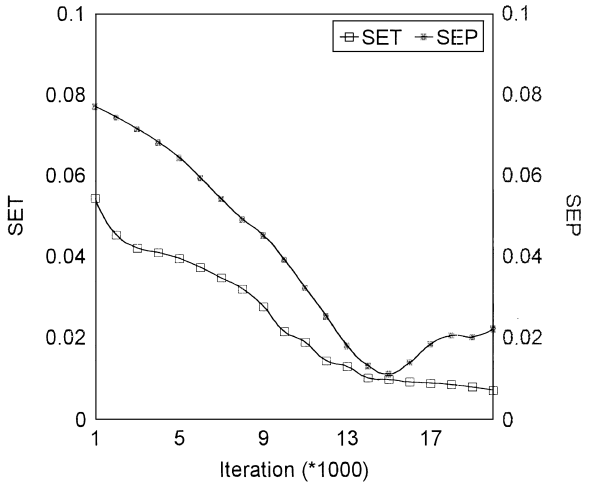

Fig. 3. Variations of SET and SEP versus the number of iterations for one group of the cationic sulfonamides.

iteration, the SEP for the prediction set initially decreases and then starts to increase after ~15 000 iterations. This situation, called overtraining, causes the ANN to loose its predictive power. Therefore, during training of the networks, it is desirable that iterations are stopped when overtraining begins.

The values of R^2_{cv} obtained using the cross-validation method for different groups of compounds are shown in Table 4. These results indicate the superiority of the generated networks over the regression models.

Table 5 compares the MLR and ANN calculated values of the sulfonamide mobilities. The correlation coefficients and standard error values of these models show the superiority of the ANN over that of the MLR models for the prediction of the electrophoretic mobility of these compounds. The standard errors (SE) of 6.05 and 4.51% for the MLR models of anionic sulfonamides should be compared with the values of 1.91 and 1.07%, respectively, for the ANN model. For cationic sulfonamides, standard error has improved considerably from 13.49% for the MLR model to 1.96% for the ANN model. Fig. 4(a–b) shows the plot of the ANN calculated against the experimental values of the electrophoretic mobility of sulfonamides as cationic and anionic forms. The residuals of the ANN predicted values of electrophoretic mobilities are plotted against the experimen-

Table 5
Comparison between the results obtained using the ANN and MLR models

Model	ANN		MLR	
	R	SE (%)	*R*	SE (%)
Cationic	0.999	1.96	0.934	13.49
Anionic (1)	0.991	1.91	0.909	6.05
Anionic (2)	0.997	1.07	0.937	4.51

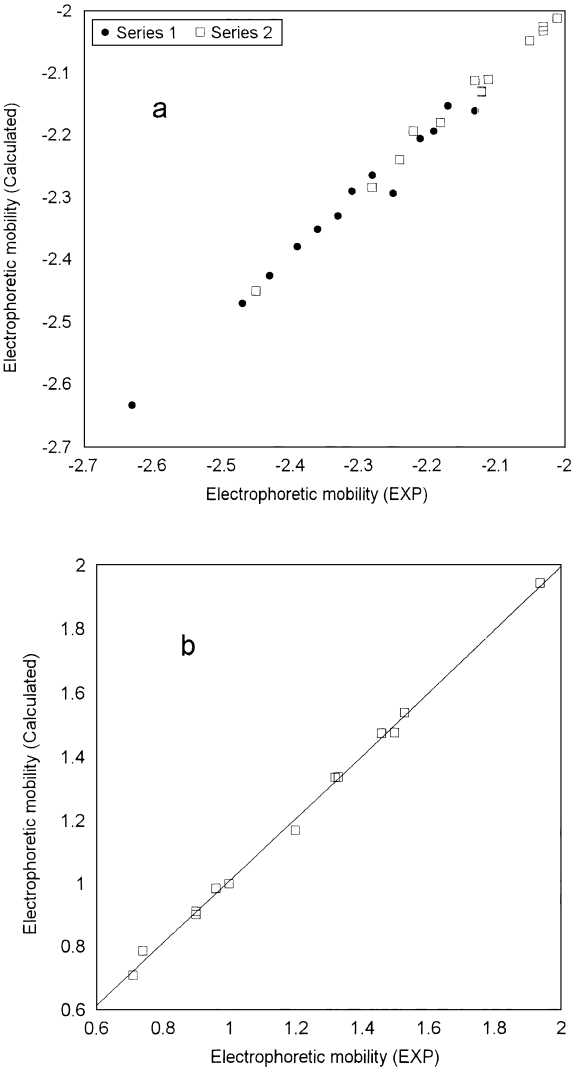

Fig. 4. Experimental versus calculated values of electrophoretic mobilities. (a) Anionic sulfonamides; (b) cationic sulfonamides.

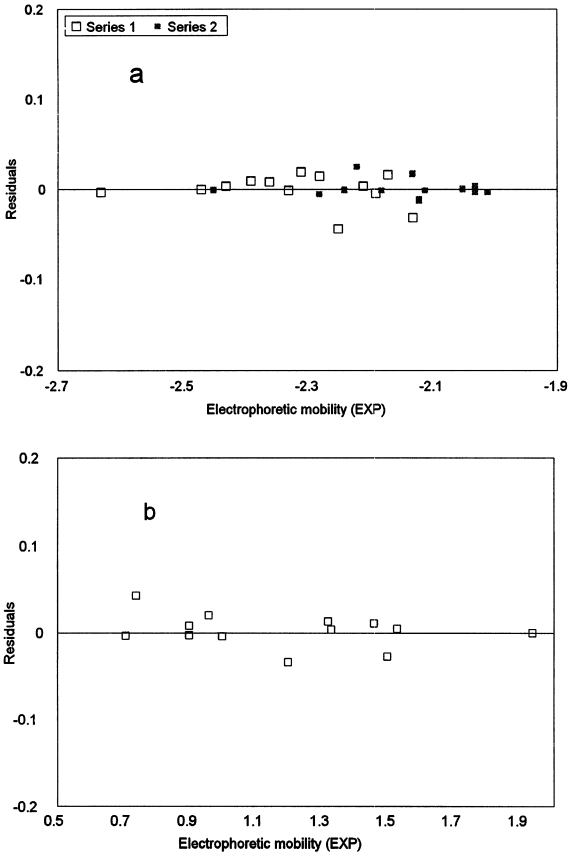

Fig. 5. Plot of residuals versus experimental values of electrophoretic mobilities. (a) Anionic sulfonamides; (b) cationic sulfonamides.

tal values and are shown in Fig. 5(a–b). The propagation of residuals in both sides of zero indicates that no systematic error exists in the development of the neural networks.

Acknowledgements

The authors wish to acknowledge the vice-presidency of research, Sharif University of Technology, for financial support of this work.

References

[1] S. Fu, C.A. Lucy, Anal. Chem. 70 (1998) 173.
[2] S. Fu, C.A. Lucy, Analyst 123 (1998) 1487.
[3] H. Liang, H. Vuorela, M. Riekkola, R.J. Hiltunen, J. Chromatogr. A 798 (1998) 233.
[4] C.E. Lin, C.C. Chang, W.C. Lin, E.C. Lin, J. Chromatogr. A 755 (1996) 261.
[5] C.E. Lin, C.C. Chang, W.C. Lin, E.C. Lin, J. Chromatogr. A 792 (1997) 37.
[6] C.E. Lin, C.C. Chang, W.C. Lin, J. Chromatogr. A 768 (1997) 105.
[7] S.R. Junson, J.M. Sutter, H.L. Engelhart, P.C. Jurs, J. White, J.S. Kauer, Anal. Chem. 69 (1997) 4641.
[8] H. Liu, X.W. Cao, R.J. Xu, N.Y. Chen, Anal. Chim. Acta 342 (1997) 223.
[9] P.K. Hopke, X. Song, Anal. Chim. Acta 348 (1997) 375.
[10] J. Zupan, J. Gasteiger, Anal. Chim. Acta 248 (1991) 1.
[11] M. Jalali-Heravi, M.H. Fatemi, Anal. Chim. Acta 415 (2000) 95.
[12] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 825 (1998) 161.
[13] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 897 (2000) 227.
[14] H. Chan, A. Butler, D.M. Falck, M.S. Freund, Anal. Chem. 69 (1997) 2373.
[15] J.J.P. Stewart, MOPAC Package, Version 6, U.S. Air Force Academy, Colorado Springs, CO, 80840.
[16] Hyperchem, Molecular Modeling System, Hyper Cube, Inc. and Auto Desk, Inc., 1993.
[17] T.R. Stouch, P.C. Jurs, J. Chem. Inf. Comput. Sci. 26 (1986) 26.
[18] SPSS/PC, the Statistical Package for IBMPC, Quiad Software, Ontario, 1986.
[19] R.E. Offord, Nature (London) 211 (1966) 591.
[20] P.D. Grossman, J.C. Colbum, H.H. Lauer, Anal. Biochem. 179 (1989) 28.
[21] R.D. Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5960.