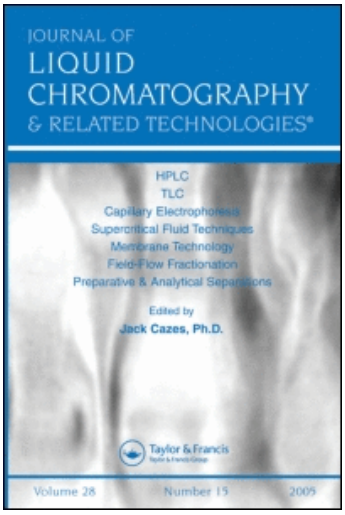

Journal of Liquid Chromatography & Related Technologies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

RETENTION MODELING AND OPTIMIZATION OF pH VALUE AND SOLVENT COMPOSITION IN HPLC USING BACK-PROPAGATION NEURAL NETWORKS AND UNIFORM DESIGN

Yichu Shan[a]; Ruihuan Zhao[a]; Yan Tian[a]; Zhen Liang[a]; Yukui Zhang[a]
[a] National Chromatographic R. & A. Centre, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, P. R. China

Online publication date: 05 August 2002

To cite this Article Shan, Yichu , Zhao, Ruihuan , Tian, Yan , Liang, Zhen and Zhang, Yukui(2002) 'RETENTION MODELING AND OPTIMIZATION OF pH VALUE AND SOLVENT COMPOSITION IN HPLC USING BACK-PROPAGATION NEURAL NETWORKS AND UNIFORM DESIGN', Journal of Liquid Chromatography & Related Technologies, 25: 7, 1033 — 1047
To link to this Article: DOI: 10.1081/JLC-120003422
URL: http://dx.doi.org/10.1081/JLC-120003422

J. LIQ. CHROM. & REL. TECHNOL., 25(7), 1033–1047 (2002)

RETENTION MODELING AND OPTIMIZATION OF pH VALUE AND SOLVENT COMPOSITION IN HPLC USING BACK-PROPAGATION NEURAL NETWORKS AND UNIFORM DESIGN

Yichu Shan, Ruihuan Zhao, Yan Tian, Zhen Liang, and Yukui Zhang*

National Chromatographic R. & A. Centre, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 161 Zhongshan Road, 116011 Dalian, P. R. China

ABSTRACT

A novel method for the optimization of pH value and composition of mobile phase in HPLC using artificial neural networks and uniform design is proposed. As the first step, seven initial experiments were arranged and run according to uniform design. Then the retention behavior of the solutes is modeled using back-propagation neural networks. A trial method is used to ensure the predicting capability of neural networks. Finally, the optimal separation conditions can be found according to a global resolution function. The effectiveness of this method is validated by optimization of separation conditions for both basic and acidic samples.

*Corresponding author. E-mail: ykzhangl@online.ln.cn

INTRODUCTION

The optimization of HPLC separations requires an accurate modeling of the relationship between the retention behavior of solutes and mobile phase parameters. Solvent composition and pH are two important parameters, which have great effects on the resolution of ionogenic compounds.

Great efforts were made on the retention modeling and optimization of pH and solvent composition (1–6) and several methods have been proposed. Generally, these methods can be divided into two categories: empirical and theoretical methods. The chief difference between them lies in how to describe the relationship between retention factor and pH value of mobile phase.

In empirical methods, mathematical models (simple, quadratic, or cubic equations) are derived and evaluated to find the best equation (7–9). Some of them are as follows:

$$\log(k) = a + b\ pH + c\ pH^2 \tag{1}$$

$$k = a + b\ pH + c\ pH^2 \tag{2}$$

$$k = a + b\ pH \tag{3}$$

$$k = a + b\ pH + c\ pH^2 + d\ pH^3 \tag{4}$$

These equations are simple but only suitable over a limited pH range. Moreover, the modeling functions derived may be remarkably insensitive to changes in the values of some of the coefficients, suggesting that they have no physical meaning.

In theoretical methods, the following equation (10) was widely used:

$$k' = \frac{k'_{HA} + k'_{A^-}(K_a/a_{H_m^+}y_{A_m^-})}{1 + (K_a/a_{H_m^+}y_{A_m^-})} \tag{5}$$

This model is more reliable, but some constants of compounds such as Ka must be known or calculated first. Moreover, the fact that both the pH of mobile phase and dissociation constant of solute vary with the proportion of organic solvent enhances the difficulty of modeling this relationship.

As a competitive nonlinear fitting tool, artificial neural network (ANN) can reduce this difficulty. It is a "soft model" that can help to solve a great variety of real problems such as modeling, association, mapping, calibration, and pattern recognition (11–14). Recently, more and more applications in the chromatographic area have been reported (15–17). The disadvantage of ANN is that a lot of experimental data is needed to train the neural networks. A combination of experimental design and ANN can be used to model the relationship between retention behavior of solutes and pH as well as solvent composition using limited

experiments, which was proved by its application to the separation condition optimization of anilines and derivatized amino acids.

THEORY

A layered feed forward neural network is composed of layers, which consist of simple computational nodes (see Figure 1). Nodes in each layer are fully connected to nodes of the succeeding layer. A number called weight represents each connection. Each layer of nodes receives its input from the previous layer. The output of each node feeds the next layer. Hidden layers provide the interconnection between input and output.

There are a lot of learning algorithms used to train neural networks. The most commonly used one is called error back propagation algorithm that was first described by Rumelhart and McClelland (18). During the learning procedure, a series of input patterns with their corresponding output values are presented to the network. Then the network is trained by performing optimization of all the

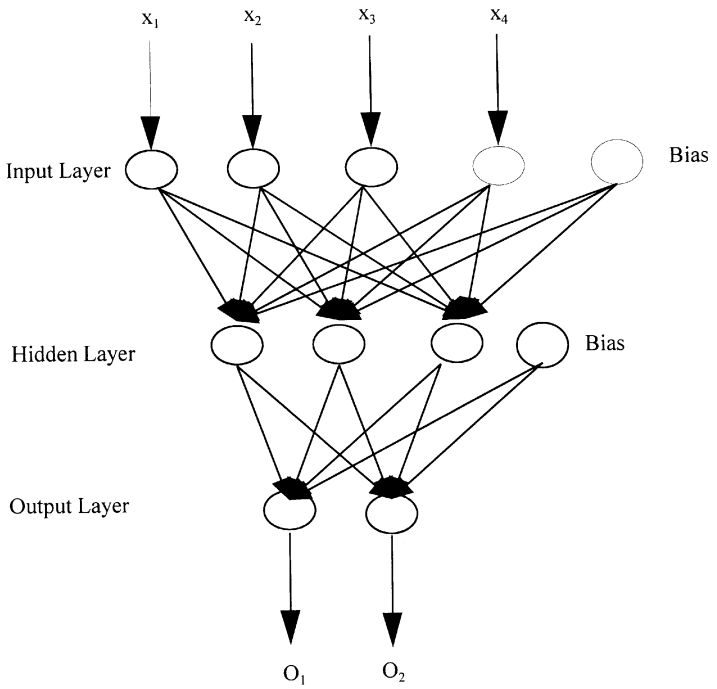

Figure 1. A diagram of the architecture of a three-layer artificial neural network.

weights and bias until the error between values output at the output layer neurons and actual outputs achieves a minimum. Generally the error function can be expressed as:

$$E = 1/2 \cdot \sum_{k=1}^{N} E_k = 1/2 \cdot \sum_{k=1}^{N} \sum_{i=1}^{P} (y_{ki} - \hat{y}_{ki})^2 \quad (6)$$

where y_{ki} is the output of nodes i in output layer for the input pattern k and $\hat{y}_{ki}$ is the corresponding actual output of the network. The weights are adjusted by a gradient descent method according to following equation:

$$w_{ij}(n+1) = w_{ij}(n) - \eta \cdot \frac{\partial E}{\partial w_{ij}(n)} + \alpha \cdot (w_{ij}(n) - w_{ij}(n-1)) \quad (7)$$

where η is learning rate, α is the momentum term.

EXPERIMENTAL

Separations were made on a C-18 column (250×4.6 mm i.d. Elite Scientific Instruments Co., Ltd., Dalian, China). The column was kept at 25°C. The elution was performed at a flow rate of 1.0 ml min^{-1}. The absorption was monitored at 254 nm. Combination of acetonitrile:water adjusted to an acid pH by the adding of acetic acid was used as the mobile phase.

Anilines and amino acids studied are listed in Table 1. They are all analytical grade. The derivatizing reagent for amino acids is a new fluorescent

Table 1. List of Compounds Studied

Anilines		Amino acids	
No.	Name	No.	Name
1	1,4-phenylenediamine	1	Glutamic acid
2	2,4-diaminotoluene	2	Aspartic acid
3	Biphenylamine	3	Arginine
4	4,4′-methylene-dianiline	4	Serine
5	4-nitroaniline	5	Glycine
6	3-nitroaniline	6	Threonine
7	o-tolidine		
8	2-nitroaniline		
9	2,6-dinitroaniline		
10	2,4-dinitroaniline		
11	3,5-dinitroaniline		

reagent-2-(9-carbazole)-ethyl chloroformate (CEOC). The derivatization process is as follows: 20–30 μL of aqueous amino acids was added to a vial; then, 200 μL of 0.2 M borate buffer (pH 9.5) and 300 μL of CEOC acetonitrile were successively added. After 1 min, the reaction mixture was extracted with pentene to remove excess reagent. All the solvents used for the preparations of the mobile phase were HPLC grade and the mixtures were filtered and degassed before use.

RESULTS AND DISCUSSION

Separation of Eleven Anilines

Experimental Design

Experimental design has been used with the advantage of reduction of the number of experiments. It provides a mathematical framework for changing all pertinent factors, simultaneously, using the smallest number of experiments. Uniform experimental design method was used because a minimum number of experiments are required for a certain number of levels and factors (19). In this case, the factors studied are pH value and percentage of acetonitrile in the mobile phase. The percentage of acetonitrile was varied from 30% to 60% v/v and the pH value was taken between 3.5 and 6.5. Seven initial experiments were arranged according to the uniform experimental design (see Figure 2).

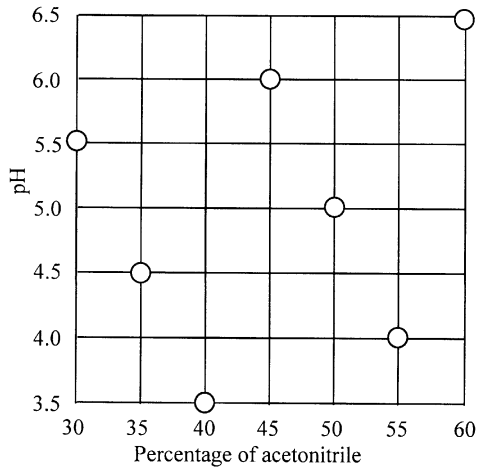

Figure 2. Experiments arranged according to uniform experimental design.

Modeling of Retention Behavior Using Artificial Neural Networks

A self-developed artificial neural network simulation software AnnLab was used. Calculations were performed on a 586 personal computer. For each solute, the relationship between retention factor and pH and solvent composition was emulated by a network with two inputs (pH and fraction of acetonitrile): one hidden layer, and one output (retention factor of solute).

To obtain best predicting results, several ANN models with different numbers of hidden nodes (1–6) were tested. At the start of training, all weights were initialized with random values. It was found that best results could be obtained using an ANN with 2-1-1 structure.

Overfitting is a common problem of ANN. A testing set is used to supervise the training process. In this case, retention data of one experiment (pH 5.0, 50% acetonitrile) was taken as a testing set; others were used as training set. Usually, the training process stopped when the error of the testing set began to increase. However, it was not applicable in this case. When the training process stopped, the performances of ANN were not sufficient enough to model the retention behavior of solutes. This may be caused by the small size of training set. Alternatively, a trial method was adopted. The testing set was still used in the training process, but it was not tested until the error of training set was lower than a desired small value. If the error of the testing set was in accordance with that of the training set, training process was stopped; otherwise, the ANN model was initialized and trained again. The training and testing results using finally obtained neural networks are shown in Table 2. It can be seen that the performances of neural networks were sufficient enough, and can be used to predict the retention behavior of solutes.

Optimization of pH and Solvent Composition

After being trained, ANNs were used to predict the retention factor of the solutes under elution conditions with certain pH values and solvent composition. The resolution of samples can be calculated according to the following global resolution function (20):

$$R = \mathrm{Min}(R_{ij}) \qquad \text{if } \mathrm{Min}(R_{ij}) \leq R_1 \tag{8}$$

$$R = R_1 + 1/t_a \qquad \text{if not} \tag{9}$$

where $\mathrm{Min}(R_{ij})$ is the resolution for the worst separated pair of peaks on the chromatogram, R_1 is the limit resolution accepted, t_a is the analysis time.

A mapping method was used to find the global optimal separation conditions. The retention time of each solute under all possible combinations of

Table 2. Relative Error Between Experimental and Predicted Retention Time of Anilines Using ANN

Condition	Training set						Testing set
RE (%)	pH 3.5	pH 4.0	pH 4.5	pH 5.5	pH 6.0	pH 6.5	pH 5.0
Compounds	40% ACN	55% ACN	35% ACN	30% ACN	45% ACN	60% ACN	50% ACN
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	−0.04	0.14	0.06	−0.70	−0.03	0.63	−0.07
3	0.49	0.30	−0.70	−0.97	−0.27	1.05	0.12
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.35	−0.36	−0.15	0.11	0.22	−0.12	−0.04
6	0.57	−0.34	−0.61	0.34	0.16	−0.33	0.21
7	5.31	−4.12	−3.90	−1.42	3.00	0.54	0.89
8	0.00	−0.05	−0.01	0.29	0.00	−0.26	0.02
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.35	−0.19	−0.40	0.22	0.08	−0.19	0.15
11	0.99	−1.12	−0.24	0.08	0.70	−0.24	−0.17

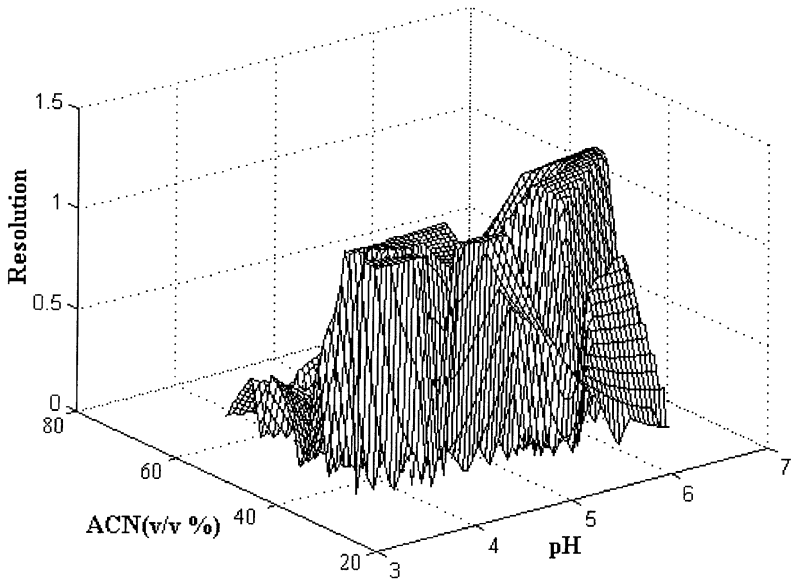

Figure 3. Response surface of resolution function versus pH and percentage of acetonitrile for anilines.

pH and solvent composition was predicted using the trained neural networks. Then, the corresponding values of resolution function were calculated and the response surface of resolution function was plotted (see Figure 3). According to the response surface, best separation performance can be obtained at a point at which the pH value is 6.3 and the percentage of acetonitrile is 44.5%. The experimental chromatograms of the samples under the optimal separation conditions are shown in Figure 4. The separation was satisfactory. Moreover, the average relative error between experimental and predicted retention time of all the solutes is 1.98% (see Table 3). The experimental results were in good accordance with the predicted results. In order to compare with ANN, the retention times of some compounds were also predicted according to equation (5) and listed in Table 3. It can be seen, that the prediction results obtained by theoretical method were slightly better than those obtained by ANN. However, it was difficult to predict the retention time of the other compounds using equation (5) because their pKa was unknown or they have more than one pKa. It is an advantage of ANN to be able to predict the retention time of any kind of ionogenic compounds without any information of pKa.

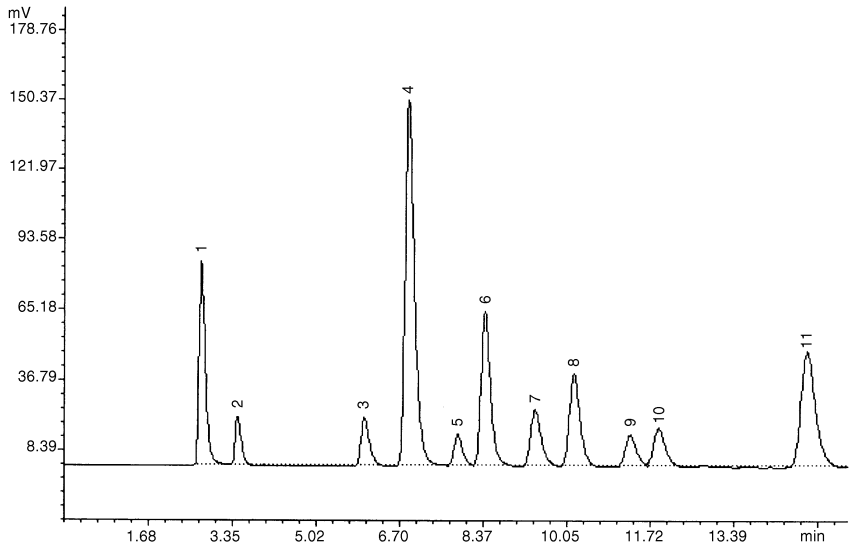

Figure 4. Experimental chromatogram of eleven anilines under optimal separation conditions (pH value = 6.3, percentage of acetonitrile = 44.5%).

Table 3. Experimental and Predicted Retention Time of Eleven Kinds of Anilines Under the Optimal Separation Conditions

Compounds	tr (exp.)	tr (pre.)	RE (%)	tr (predicted using eq. 5)	RE (%)
1	2.71	2.7	−0.37		
2	3.43	3.44	0.29		
3	5.97	5.86	−1.84		
4	6.86	7.11	3.64		
5	7.84	8.01	2.17	8.00	2.04
6	8.39	8.56	2.03	8.44	0.6
7	9.39	9.61	2.34		
8	10.17	10.37	1.97	10.23	0.59
9	11.3	11.41	0.97	11.37	0.62
10	11.87	12.21	2.86	12.03	1.35
11	14.84	15.33	3.30	15.11	1.82
Average relative error			1.98		1.17

Separation of Six Derivatized Amino Acids

Experimental Design and Retention Modeling of Derivatized Amino Acids

For the separation of six kinds of derivatized amino acids, the range of pH is from 3.0 to 6.0 and the percentage of acetonitrile varies from 22% to 40%. The conditions of seven experiments arranged are listed in Table 4. These experiments were run and the retention data obtained were used to train the neural networks. Thus, the retention time of solute under arbitrary pH value and mobile phase composition can be predicted. Figure 5 shows the response surface of the retention time of serine versus pH and percentage of acetonitrile. It is shown that the retention time increases along with the decrease of pH value and percentage of acetonitrile. Figure 6 shows the variation of retention time of six kinds of derivatized amino acids along with the pH value of mobile phase. It can be seen that the sequences of peaks change accompanying the variation of pH value, which indicates that the selectivity can be greatly affected by varying the pH value of mobile phase.

Optimization of Separation Conditions: pH Value and Percentage of Acetonitrile

As described before, because the retention time of each solute can be modeled using trained neural networks, the response surface of the total resolution of derivatized amino acids can be plotted according to global resolution function (see Figure 7). It can be seen that the resolution of samples is affected by both pH value and acetonitrile percentage. The resolution always increases along with the decrease of acetonitrile percentage. However, the effect of pH value on resolution is more complex; this can be shown more clearly by fixing the acetonitrile percentage (see Figure 8). Although there are several maximum points on the curve of resolution versus pH value, the best pH value for the separation is about 3.0.

Table 4. Initial Experiment Arranged for the Separation of Derivatized Amino Acids

No. of Experiment	1	2	3	4	5	6	7
pH value	3.0	3.5	4.0	4.5	5.0	5.5	6.0
Percentage of ACN	28	37	25	34	22	31	40

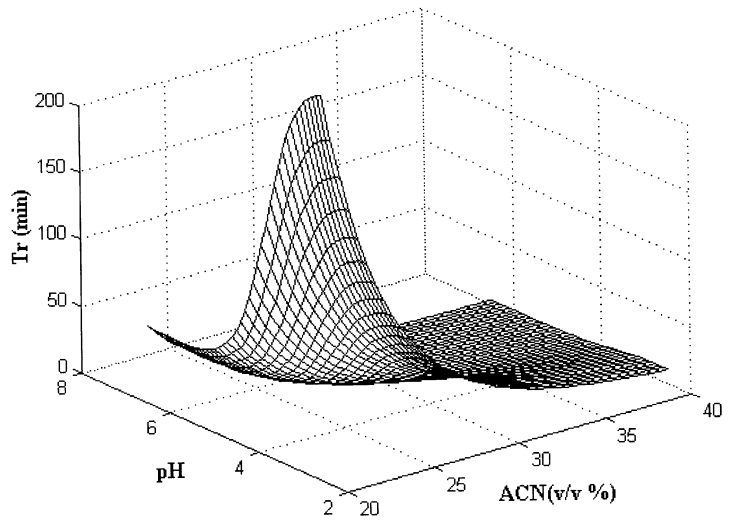

Figure 5. Response surface of retention time of derivatized serine versus pH value and acetonitrile percentage of mobile phase.

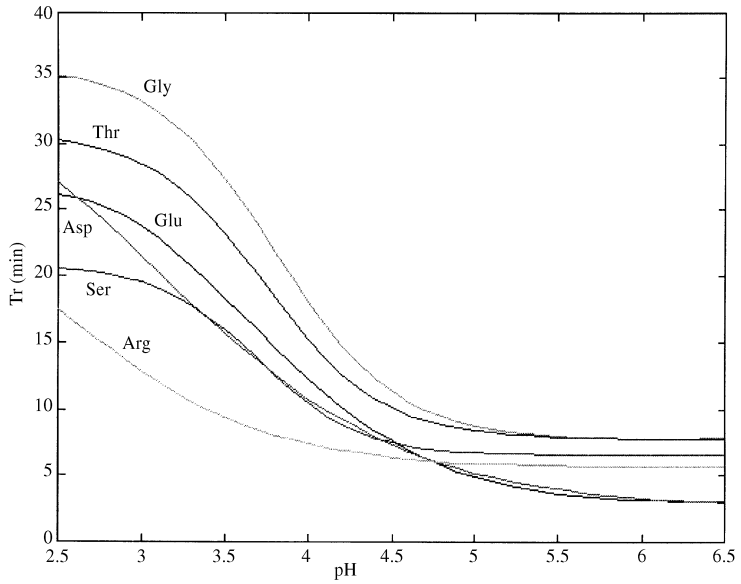

Figure 6. Variation of retention times of derivatized amino acids along with the pH value of mobile phase (ACN percentage fixed at 27%).

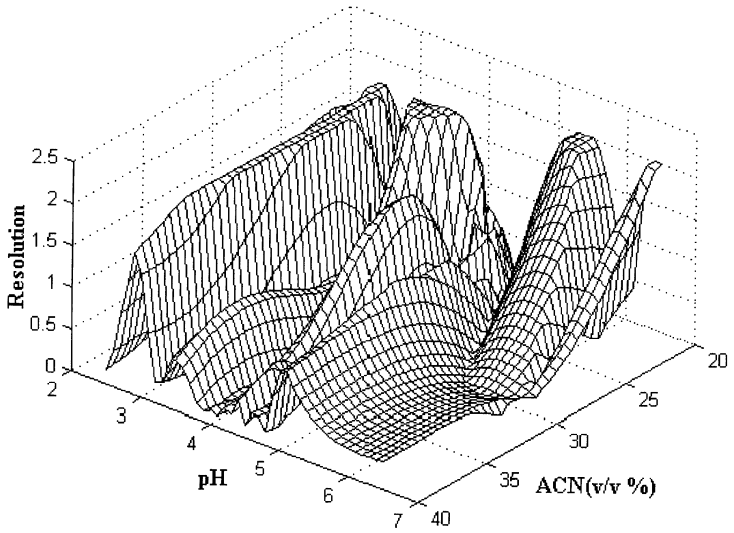

Figure 7. Response surface of resolution function versus pH and percentage of acetonitrile for derivatized amino acids.

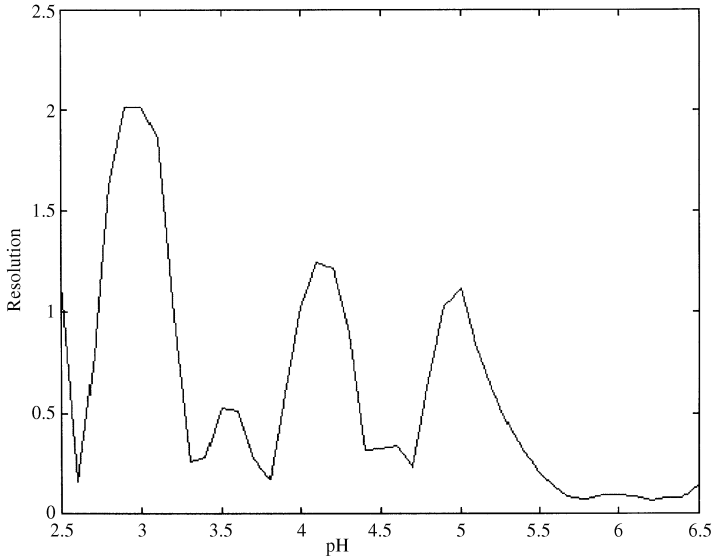

Figure 8. The effect of pH value on the resolution of derivatized amino acids (ACN percentage fixed at 27%).

Table 5. Experimental and Predicted Retention Time of Derivatized Amino Acids Under the Optimal Separation Conditions

Compounds	tr (exp.)	tr (pre.)	RE (%)
1	10.77	11.1	2.97
2	10.12	10.31	1.84
3	6.99	6.67	−4.80
4	9.55	9.4	−1.60
5	15.23	15.22	−0.07
6	12.74	12.68	−0.47
Average relative error			1.96

Examining all possible combinations of pH value and acetonitrile percentage, the best separation conditions were pH value of 2.8 and acetonitrile percentage of 36%. The predicted and experimental retention time of solutes under this optimal condition is listed in Table 5. The average relative error between predicted and experimental time is 1.96%. Also, the chromatogram (see Figure 9) indicates that the separation was successful.

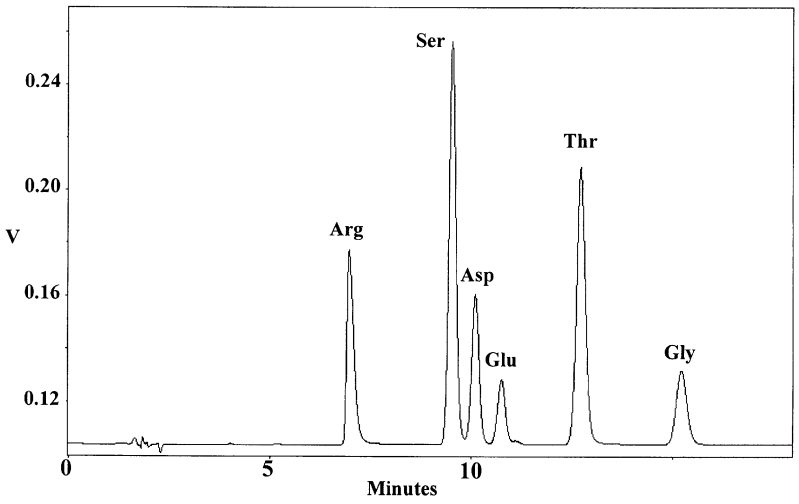

Figure 9. Experimental chromatogram of derivatized amino acids under optimal separation conditions (pH value = 2.8, percentage of acetonitrile = 36%).

CONCLUSION

Relationship between chromatographic retention and pH and solvent composition can be modeled by a combination of ANN and experimental design using limited initial experiments. Optimal separation conditions were found using a global resolution function. Separation of anilines and derivatized amino acids under the optimal separation conditions were successful. The average relative error between predicted and experimental retention time under the optimal separation conditions is 1.98% for anilines and 1.96% for derivatized amino acids, which shows the good predicting ability of ANN. Owning the advantage of modeling retention behavior of solutes without the use of any mathematical model, ANN can be a powerful and simple tool for the method development of HPLC.

ACKNOWLEDGMENTS

Financial support from the National Natural Science Foundation of China (Project 20175029) is greatly acknowledged.

REFERENCES

1. Lopes-Marques, R.M.; Schoenmakers, P.J. J. Chromatogr. **1992**, *157*, 592.
2. Berges, R.; Sanz-Nebot, V.; Barbosa, J. J. Chromatogr. A **2000**, *869*, 27.
3. Hernández-Arteseros, J.A.; Barbosa, J.; Compañó, R.; Prat, M.D. Chromatographia **1998**, *48*, 251.
4. Bosch, E.; Espinosa, S.; Rosés, M. J. Chromatogr. A **1998**, *824*, 137.
5. Bergés, R.; Sanz-Nebot, V.; Barbosa, J. J. Chromatogr. A **2000**, *869*, 27.
6. Goga, S.; Heinisch, S.; Rocca, J.L. Chromatographia **1998**, 48, 237.
7. Schoenmakers, P.J.; Tijssen, R. J. Chromatogr. A **1993**, *656*, 577.
8. Haddad, P.R.; Drouen, A.C.J.H.; Billiet, H.A.H.; De Galan, L. J. Chromatogr. **1983**, *282*, 71.
9. Kiel, J.S.; Morgan, S.L.; Abramson, R.K. J. Chromatogr. **1989**, *485*, 585.
10. Horvath, C.; Melander, W.; Molnar, I. Anal. Chem. **1977**, *49*, 142.
11. Gasteiger, J.; Zupan, J. Angew. Chem. Int. Eng. **1993**, *32*, 503.
12. Mittermayr, C.R.; Drouen, A.C.J.H.; Otto, M.; Grasserbauer, M. Anal. Chim. Acta **1994**, *294*, 227.
13. Zhao, R.H.; Yue, B.F.; Ni, J.Y.; Zou, H.F.; Zhang, Y.K. Chemometrics Intell. Lab. Syst. **1999**, *45*, 163.
14. Zhao, R.H.; Xu, G.W.; Yue, B.F.; Liebich, H.M.; Zhang, Y.K. J. Chromatogr. A **1998**, *828*, 489.

15. Bruchmann, A.; Zinn, P. Anal. Chim. Acta **1993**, *283*, 869.
16. Jiménez, O.; Benito, I.; Marina, M.L. Anal. Chim. Acta **1997**, *353*, 367.
17. Agatonovic-Kustrin, S.; Zecevic, M.; Zivanovic, Lj.; Tucker, I.G. Anal. Chim. Acta **1998**, *364*, 265.
18. McCelland, J.L.; Rumelhart, D.E. *Explorations in Parallel Distributed Processing*; MIT Press: Massachusetts, 1988.
19. Fang, K.T. *Generalized Multivariate Analysis*; Springer-Verlag, Inc.: New York, 1990.
20. Guillaume, Y.; Cavalli, E.J.; Peyrin, E.; Guinchard, C. J. Liq. Chrom. & Rel. Technol. **1997**, *20*, 1741.

Received December 14, 2001
Accepted January 8, 2002
Manuscript 5701