AIChE

A Sequence Optimization Strategy for Chromatographic Separation in Reversed-Phase High-Performance Liquid Chromatography

Xueling Du, Ye Li, and Qipeng Yuan
State Key Laboratory of Chemical Resource Engineering, Beijing University of Chemical Technology, Beijing 100029, China

DOI 10.1002/aic.12006
Published online August 31, 2009 in Wiley InterScience (www.interscience.wiley.com).

A sequence optimization strategy combining an artificial neural network (ANN) and a chromatographic response function (CRF) for chromatographic separation in reversed-phase high-performance liquid chromatography has been proposed. Experiments were appropriately designed to obtain unbiased data concerning the effects of varying the mobile phase composition, flow-rate, and temperature. The ANN was then used to simultaneously predict the resolution and analysis time, which are the two most important features of chromatographic separation. Subsequently, a CRF consisting of resolution and analysis time was used to predict the optimum operating conditions for different specialized purposes. The experimental chromatograms were consistent with those predicted for given conditions, which verified the applicability of the method. Furthermore, the proposed optimization strategy was applied to literature data and very good agreement was obtained. The results show that a strategy of sequential combination of ANN and CRF can provide a more flexible and efficient optimization method for chromatographic separation. © 2009 American Institute of Chemical Engineers *AIChE J,* 56: 371–380, 2010
Keywords: sequence optimization, artificial neural network, chromatographic response function, solanesol, chromatography

Introduction

High-performance liquid chromatography (HPLC) based on a reversed stationary phase, coupled with suitable mobile phase composition, flow-rate, and temperature, is one of the most powerful techniques in research, quality control, and many other areas in analytical chemistry. To obtain more intelligent and efficient chromatographic separation, optimization of the HPLC method has become an important objective.[1–10]

Correspondence concerning this article should be addressed to Q. Yuan at yuanqp@mail.buct.edu.cn

The first attempt to solve this problem was based on the theoretical advances in chromatography, which allowed the experimental response to be predicted by controlling and modifying the chromatographic parameters.[11] Subsequently, computer-assisted methodologies were suggested as a more sophisticated method for optimization purposes.[12,13]

Multicriteria decision making (MCDM), a branch of operations research, is a useful method that can be employed when more than one optimization criterion has to be taken into account.[14–17] The essence of MCDM involves judging the different quality aspects of a chromatogram individually and quantitatively.[18,19] In the original MCDM method, a combined criterion called a utility function-by which a compromise between very different chromatographic goals may

be achieved-is used for simultaneous optimization of different chromatographic aspects. As a result, the multicriteria problem is reduced to a single criterion problem, namely optimizing the utility function.[19] Some utility functions consisting of the factors related to different chromatographic aspects have been developed, such as the chromatographic exponential function and the chromatographic optimization function.[20–24] Although utility functions have been extensively used in chromatography, the application procedure has some disadvantages. Firstly, it is difficult to consider the prior weights for all the criteria. Furthermore, it is possible that the multicriteria optimum result leads to an unacceptable value of one or more of the criteria. It can happen that very good solutions are found for one of the criteria with high weight, so that the bad results for some of the other criteria are compensated.[19] Subsequently, another improved MCDM method for simultaneously optimizing different criteria was presented by Harrington.[25] He proposed that one can multiply the criteria instead of summing them. This mathematical model was put into a more general form by Derringer and Suich.[26] Derringer's desirability function was introduced into chromatography by Bourguignon and Massart.[16] It is based on the transformation of the measured properties to a dimensionless desirability scale for each criterion, so that values of several properties obtained from different scales of measurements may be combined. The advantage of Derringer's desirability function is that if one of the criteria has an unacceptable value, then the overall product will also be unacceptable. Therefore, the appearance of an unacceptable value for one or more criteria can be avoided. However, the difficulty of prior selection of weighting factors still exists in the Derringer's desirability function.[27–29] Moreover, it is noteworthy that there is another obvious limitation of all the existing optimization procedures. Nowadays, in most applications of the combined criterion functions (utility functions or Derringer's desirability functions), the determined response functions are directly used as the targets in some optimization methods such as artificial neural network (ANN)[24,30] and multivariate techniques[10,24] including multiple linear regression, partial least squares regression, and response surface analysis. In these cases, the authors had to construct a special chromatographic response function (CRF), according to the chromatographic purpose, before developing the model. In this kind of procedure, the flexibility of the CRF is limited and the optimum operating parameters obtained are only suitable for specialized chromatographic purpose. Once the optimization purpose changes, the CRF has to be constructed again and the whole optimization process has to be repeated. Therefore, the development of a more general and flexible strategy is necessary.

In this study, a sequence optimization strategy involving the combination of ANN with CRF is proposed. The strategy will be applied to the chromatographic separation of solanesol from tobacco leaf extracts. An appropriate experimental design including three important operating parameters (mobile phase composition, flow-rate, and temperature) is carried out to obtain the basic data for optimization. The ANN is used to simultaneously predict resolution and analysis time, which are well accepted as two crucial indicators of separation efficiency in HPLC analysis. The CRF including both resolution and analysis time is then used to locate the optimum conditions for different chromatographic purposes. Experiments under the conditions used in predictions are performed in order to verify the proposed method. Finally, experimental data from the literature is utilized to test the general applicability of the proposed strategy.

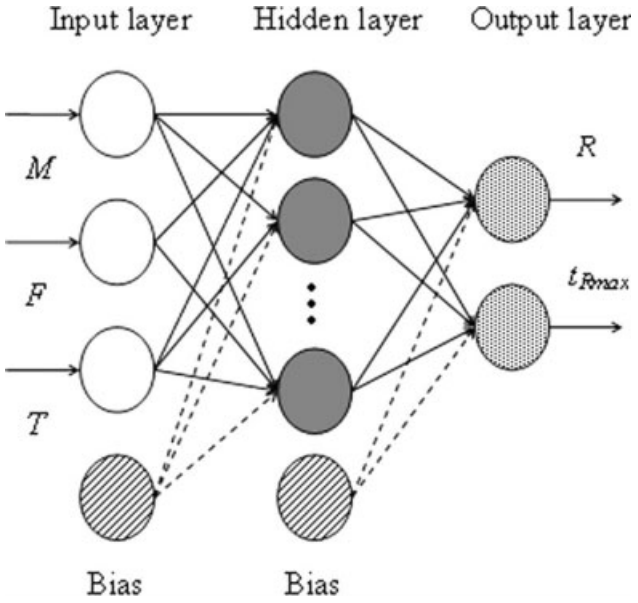

Figure 1. Scheme of the three-layer network used in this study.

M, mobile phase composition; *F*, flow rate (mL/min); *T*, temperature (K); *R*, resolution; t_{Rmax}, analysis time (min).

Theory

Artificial neural network

ANNs are computational models designed to simulate the way in which the human brain processes information. They consist of simple processing units (or neurons) linked with weighted modifiable interconnections. The neurons are generally organized into a layered structure, formed by one input layer, one output layer and at least one hidden layer. In this article, a three-layer feed-forward neural network trained with an error back-propagation algorithm is adopted. In this network, signals propagate from the input layer through the hidden layer to the output layer. A node thus receives signals via connections from other nodes, or in the case of the input layer, the outside world. The net input for a node j is given by:

$$\mathrm{net}_j = \sum_j w_{ji} o_i \tag{1}$$

where i represents nodes in the previous layer, w_{ji} is the weight associated with the connection from node i to node j, and o_i is the output of node i. The output of a node is determined by the transfer function and the net input of the node. The following sigmoid transfer function in the hidden layer was used:

$$f(\mathrm{net}_j) = \frac{1}{1 + e^{-(\mathrm{net}_j + \theta_j)}} \tag{2}$$

Table 1. Experimental Parameters and Levels Used in Star Design

Experimental Parameters	Level (−1)	Level (0)	Level (+1)
Mobile phase composition	0.7	0.8	0.9
Temperature (°C)	20	30	40
Flow rate (mL/min)	0.5	1	1.5

where θ_j is a bias term or threshold value of node j, which is responsible for accommodation of nonzero offsets in the data.

The weights are updated after each epoch as follows:

$$\Delta w_{ij} = -\eta \frac{\partial E(t)}{\partial w(t)} + \alpha \Delta w_{ij}(t-1) \tag{3}$$

where η is the learning rate, α is the momentum, and $\delta\ (t) = \partial E/\partial w$ is the actual error at time t. The learning rate, η, controls the rate at which the network learns and the momentum, α, controls the influence of the last weight change on the current weight.

The ANN used in this work is represented schematically in Figure 1. The input layer consists of three nodes representing the mobile phase composition (M), its flow-rate (F), and its temperature (T). The output layer consists of two nodes representing resolution (R) and analysis time (t_{Rmax}). In addition, there is a bias (threshold of neuron activation) in the hidden and output layers, which connects to the nodes via modifiable weights.

Chromatographic response function

The simultaneous optimization of two independent parameters, resolution and analysis time, was required for overall optimization of solanesol separation. For example, operating conditions that require long analysis times may resolve two peaks, but peak broadening, the excessive consumption of solvents and increased probability of mechanical failures exacerbated by lengthy analysis times would reduce the overall desirability of such operating conditions. Derringer's desirability function was used to combine the two target responses of peak resolution and analysis time. Resolution between the two adjacent peaks was first calculated as:

$$R = \frac{2(t_2 - t_1)}{w_1 + w_2} \tag{4}$$

where t and w denote the retention time and the corresponding baseline peak widths of the two adjacent peaks 1 and 2.

The merging of the two parameters (resolution and analysis time) was accomplished by the use of Derringer's one-sided or two-sided transformations,[26] where each measured response was transformed into a dimensionless desirability (d). In the one-sided transformation, the response variables Yi ($i = 1, 2, \ldots, n$, where n is the number of response variables) are transformed into the desirability functions, d_i, according to the following equations:

$$\begin{aligned} d_i &= 0 && \text{if } Y_i \le Y_i^{(-)} \\ d_i &= \left(\frac{Y_i - Y_i^{(-)}}{Y_i^{(+)} - Y_i^{(-)}}\right)^r && \text{if } Y_i^{(-)} \le Y_i \le Y_i^{(+)} \\ d_i &= 1 && \text{if } Y_i \ge Y_i^{(+)} \end{aligned} \tag{5}$$

where $Y_i^{(-)}$ is the minimum acceptable value of the criterion Y_i and $Y_i^{(+)}$ is the value beyond which improvements would serve no further benefit. Both the values and the parameter,

Table 2. Star Design Matrix with Experimental Data of Resolution and Analysis Time

Experiment	Mobile Phase Composition	Temperature (°C)	Flow Rate (mL/min)	Resolution*	Analysis Time (min)
1	−1	−1	−1	1.17	17.87
2	−1	−1	0	0.87	9.02
3	−1	−1	+1	0	6.03
4	−1	0	−1	0.94	16.63
5	−1	0	0	0	8.34
6	−1	0	+1	0	5.54
7	−1	+1	−1	0	15.37
8	−1	+1	0	0	7.82
9	−1	+1	+1	0	5.19
10	0	−1	−1	1.55	27.43
11	0	−1	0	1.29	13.74
12	0	−1	+1	1.10	9.14
13	0	0	−1	1.41	25.11
14	0	0	0	1.18	12.64
15	0	0	+1	1.00	8.40
16	0	+1	−1	1.27	22.96
17	0	+1	0	1.05	11.50
18	0	+1	+1	0	7.72
19	+1	−1	−1	1.57	55.00
20	+1	−1	0	1.40	28.09
21	+1	−1	+1	1.27	18.66
22	+1	0	−1	1.84	50.77
23	+1	0	0	1.60	25.41
24	+1	0	+1	1.46	16.71
25	+1	+1	−1	1.87	46.03
26	+1	+1	0	1.63	23.28
27	+1	+1	+1	1.52	13.78

*When the peak can not be recognized, the value of resolution is set as 0.

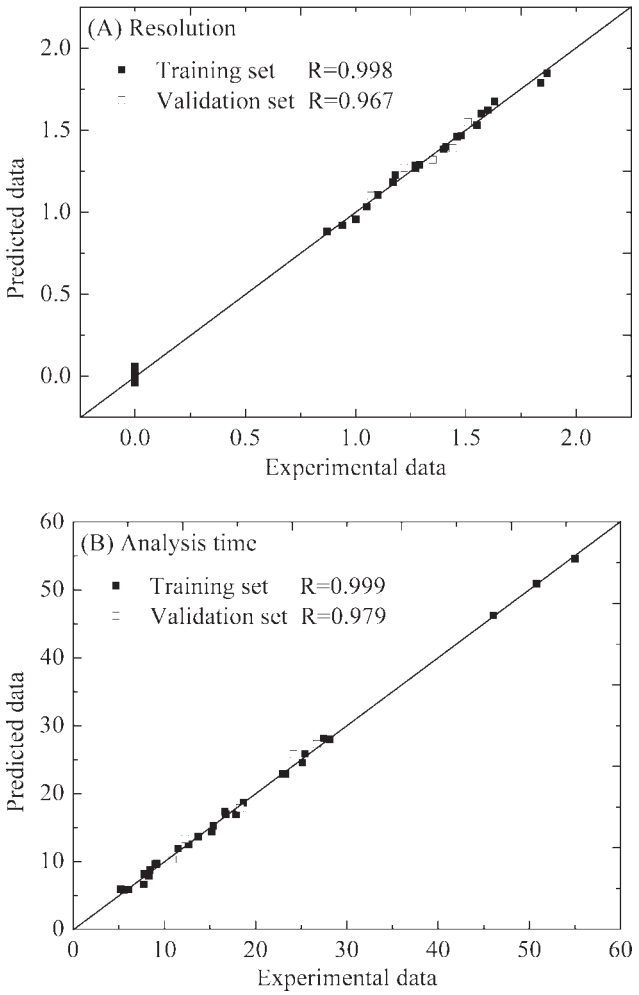

Figure 2. Correlation plots for both training set and validation set for the optimal architecture of the network.

which is in fact a kind of weighting factor, should be selected by the user. Such a transformation is valid for the separation parameter. However, this is not the case for the criteria used for evaluation of the analysis time where the d-value needs to be minimized. Therefore, $d = 0$ for $Y_i \geq Y_i^{(+)}$, $d = 1$ for $Y_i \leq Y_i^{(-)}$, and a value somewhere in between for $Y_i^{(-)} < Y_i < Y_i^{(+)}$ should be considered in such cases.

The overall quality D is calculated by multiplying the desirability values obtained for the different criteria[10] or by using their geometric mean.[1] It is noteworthy that the outcome of the overall quality D depends on the r-value and the selection of a suitable r-value for specialized chromatographic purposes offers the user flexibility in the definition of desirability functions. In this article, the individual desirabilities were combined to give a single assessment value of D using the equation:

$$D = d_r \times d_a \tag{6}$$

where d_r and d_a represent the transformations of resolution and analysis time, respectively.

Experimental

Materials and reagents

Solanesol standard was purchased from Sigma Chemical (St. Louis, MO). Ethanol, methanol, and acetone of analytical grade were purchased from Beijing Chemical Company (Beijing, China). Methanol and hexane of HPLC grade were purchased from Beijing Dikma Technologies (Beijing, China). Crude solanesol was purchased from Henan Shangqiu Tobacco Fine Chemical Plant (Henan province, China). All solutions prepared for HPLC were filtered through 0.45 μm nylon membranes before use.

Apparatus and chromatographic conditions

All experiments were performed on a Hitachi HPLC system consisting of Hitachi model L-7100 pumps, a L-7420 tunable absorbance detector, and a reversed-phase C_{18} column (250 $\times$ 4.6 mm, 5 μm, Diamonsil^{TM}). The mobile phase was a methanol–hexane system, the UV detector was set at a wavelength of 220 nm, the injection volume was 10 μL, and the temperature was controlled by a thermostat.

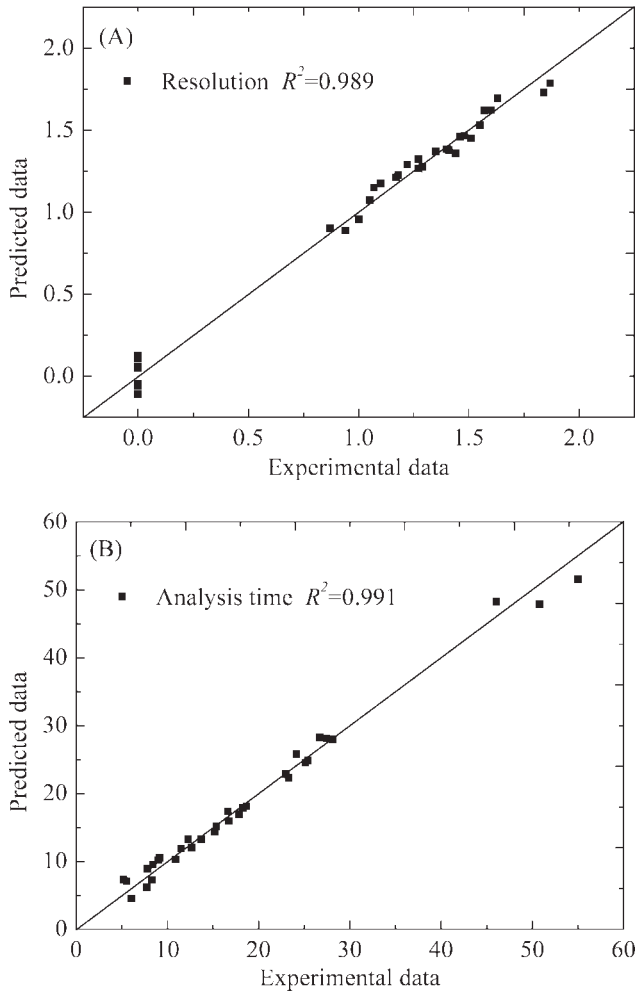

Figure 3. Plot of crossvalidated predicted values by ANN vs. experimental values.

Table 3. The Linear Regression Results at 95% Confidence Level for Figure 3

Regression Lines	Terms	LCI	UCI
Resolution			
Slope	0.981	0.943	1.019
Intercept	0.025	−0.019	0.069
Analysis time			
Slope	0.966	0.928	1.004
Intercept	0.658	−0.208	1.525

LCI, lower confidence interval; UCI, upper confidence interval.

Experimental design

A variety of experimental designs have been used for the purpose of ANN modeling. Havlis and coworkers[31] found that the experimental data from a star design was most suitable for ANN modeling when optimizing the separation of a simple mixture. The aim of this study was to optimize the separation of solanesol from tobacco leaf extracts. According to some preliminary tests, the main difficulty in this process is the separation of two adjacent peaks representing solanesol and an unknown component with lower polarity [with longer retention time in reversed-phase high-performance liquid chromatography (RP-HPLC)]. Therefore, a star design was adopted in this study to obtain suitable experimental data for ANN modeling. Table 1 shows the three chosen experimental parameters and their selected levels. Mobile phase composition (defined here as the ratio of methanol volume to total volume of solution) and column temperature have obvious effects on the chromatographic selectivity (α) and were thus chosen as parameters.[32–36] Although flow rate does not affect the selectivity factor, it plays a definite role in determining the analysis time and thus the CRF, and therefore, it was also chosen as a parameter to be optimized. The experimental level for each parameter was determined based on univariate studies.

Actually, this experimental design is a derivative of an ordinary star design. Some additional experiments were carried out at boundary points of the experimental space in order to obtain more information for the ANN training. The total number of experiments was 27 and the design matrix of operating conditions and experimental results is shown in Table 2.

Optimization procedure

Experimental data obtained from the modified star design were used to optimize the ANN. Then, the values of analysis time and resolution were predicted by the optimized ANN, and a mapping database within the experimental space was formed, which can show how the analysis time and resolution change as a function of operating parameters. Subsequently, different CRFs with different purposes were constructed by modulating the relative importance of the separation factor and the time factor. Based on the predicted database, the optimum operating conditions for different chromatographic purposes were obtained by maximizing the CRFs. Finally, the experiments were conducted under the predicted optimum chromatographic conditions in order to verify the strategy.

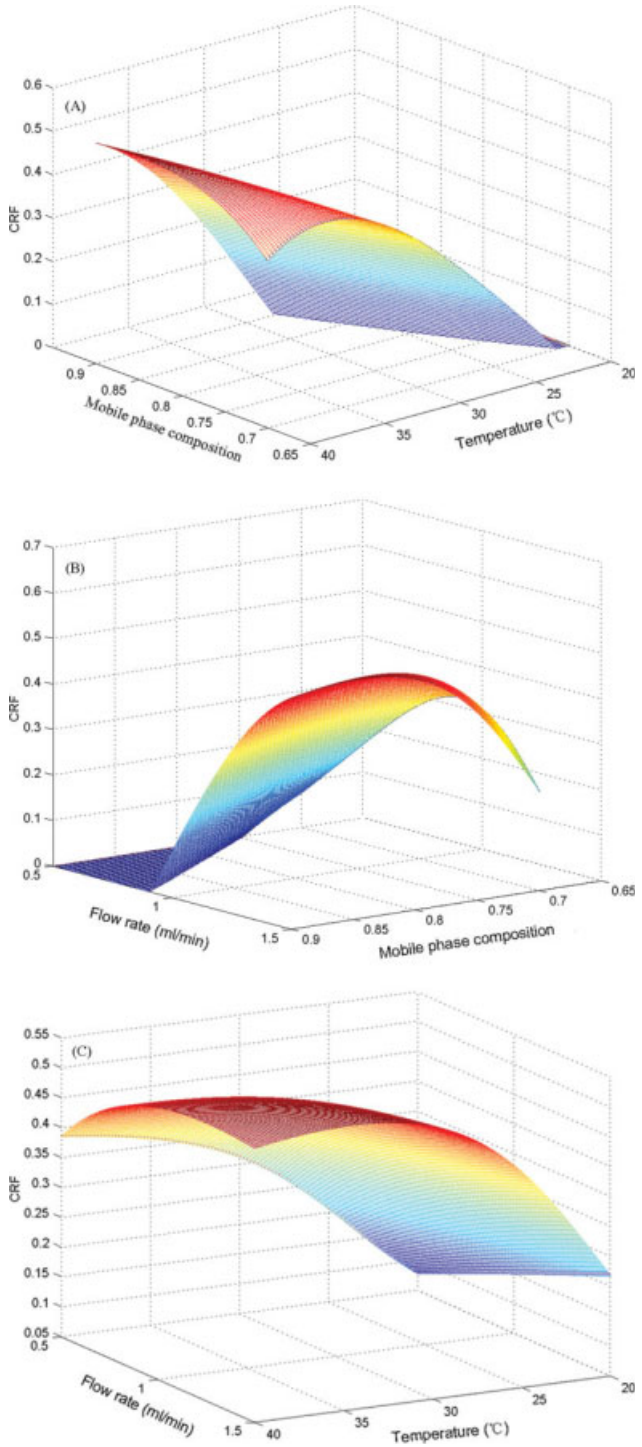

Figure 4. Response surfaces for CRF2: (A) CRF as a function of mobile phase composition and temperature; (B) CRF as a function of mobile phase composition and flow rate; (C) CRF as a function of temperature and flow rate.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 4. The CRFs for Different Chromatographic Purposes and Corresponding Optimal Operation Conditions

CRF	$Y^{(-)}$	$Y^{(+)}$	r	Mobile Phase Composition	Flow Rate (mL/min)	Temperature (°C)	CRF_p	CRF_e
CRF1								
Resolution	0.5	2.0	1.0					
Analysis time (min)	5	200	0.5	0.9	0.5	40	0.793	0.811
CRF2								
Resolution	0.5	2.0	1.0					
Analysis time (min)	5	60	1.0	0.9	1.5	34	0.525	0.533
CRF3								
Resolution	0.5	5.0	0.1					
Analysis time (min)	5	60	1	0.8	1.5	20	0.763	0.756

CRF_p, predicted values of CRF; CRF_e, experimental values of CRF.

Software

In this study, the ANN was constructed using the ANN toolbox in MATLAB simulation software version 6.1.

Results and Discussion

Artificial neural network

There are several approaches to the optimization of chromatographic separation with regard to specific goals (minimal analysis time, maximal resolution, etc.). Usually, optimum separation means that all the components of the sample are separated in a reasonable time. The analysis time of the whole separation process is limited by the analysis time of the most strongly retained component. Another important aspect of separation is resolution which should, in the case of two components, be at least 1.5. Hence, as the output data of ANN, we chose two criteria in this article: the resolution of the two adjacent peaks (R) representing solanesol and an unknown component with lower polarity, which should have as large a value as possible, and the analysis time of the more strongly retained component of these two species (t_{Rmax}), which should be as low as possible.

One hidden layer back-propagation ANN with three inputs and two outputs was developed. When optimizing the ANN, two data sets were used (a training set and a validation set). The training set was used to update the weights of the network. Then, the weights optimized by running error back-propagation on the training data were simply used to model the validation data, that is, the validation set used here was used to check the network response for untrained data. The optimum network architecture is that giving the minimum error for the validation set. The experimental data obtained by experimental design were used as the training set (Table 2). The validation set consisted of five sets of input and output data. Inputs were randomly chosen inside the space given by the input variable limits and corresponding outputs were obtained experimentally. Before fitting the data, input and output variables were normalized to a range from 0 to 1 using the linear normalization method. According to trial-and-error calculations, the value of the momentum factor was set to 0.9, for which the best results can be obtained. Furthermore, because the learning rate η has an apparent impact on the error, the adaptive learning rule was applied with a maximum value of 1.05 η and a minimum value of 0.7 η with the original value of η set to 0.02, which decreased the iteration number significantly. After the learning rate and the momentum factor were determined, the optimum hidden neuron number and training cycle number were easily found. Because the minimum error of the

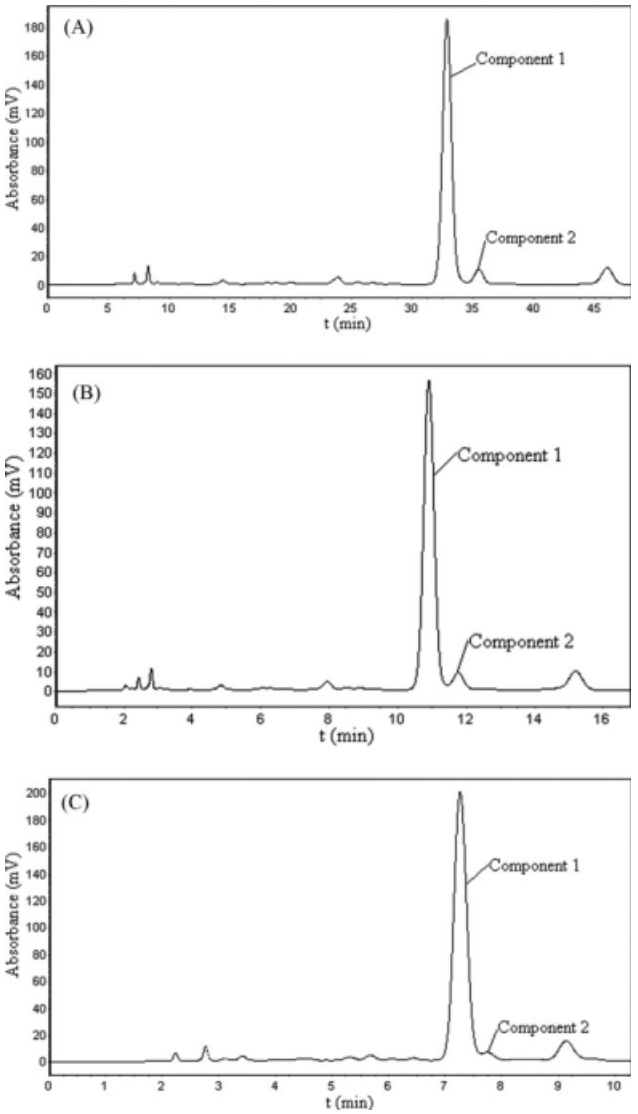

Figure 5. Chromatograms obtained under the predicted optimal conditions for (A) CRF1, (B) CRF2, and (C) CRF3.

Component 1 and 2 represent solanesol and an unknown component with lower polarity, respectively.

Table 5. Literature Data for ANN Modeling

Experiments	[SDS] (M)	N	V_m (v/v %)	pH	T (°C)	t_{Rmax} (min)	R_g
1	0.03	2	3.0	3	40	72.03	1.36
2	0.10	2	3.0	3	30	25.73	1.22
3	0.03	4	3.0	3	30	29.99	1.71
4	0.10	4	3.0	3	40	12.99	1.45
5	0.03	2	10.0	3	30	48.99	1.90
6	0.10	2	10.0	3	40	17.58	1.42
7	0.03	4	10.0	3	40	17.97	2.16
8	0.10	4	10.0	3	30	8.65	1.57
9	0.03	2	3.0	7	30	80.1	1.69
10	0.10	2	3.0	7	40	24.51	1.26
11	0.03	4	3.0	7	40	27.8	1.68
12	0.10	4	3.0	7	30	13.56	1.46
13	0.03	2	10.0	7	40	43.79	1.55
14	0.10	2	10.0	7	30	18.56	1.36
15	0.03	4	10.0	7	30	16	1.94
16	0.10	4	10.0	7	40	7.03	1.45
17	0.065	3	6.5	5	35	17.89	1.53
18	0.065	3	6.5	5	35	17.84	1.55
19	0.065	3	6.5	5	35	17.47	1.56
20	0.065	3	6.5	5	35	17.38	1.56
21	0.065	3	6.5	5	35	16.6	1.56
22	0.065	3	6.5	5	35	17.49	1.56
23	0.03	3	6.5	5	35	27.5	1.77
24	0.10	3	6.5	5	35	13.45	1.47
25	0.065	2	6.5	5	35	26.07	1.44
26	0.065	4	6.5	5	35	11.05	1.60
27	0.065	3	3.0	5	35	24.58	1.48
28	0.065	3	10.0	5	35	13.4	1.51
29	0.065	3	6.5	3	35	16.85	1.55
30	0.065	3	6.5	7	35	16.64	1.56
31	0.065	3	6.5	5	30	18.75	1.54
32	0.065	3	6.5	5	40	17.09	1.53

[SDS], sodium dodecyl sulfate concentration (M); N, alkyl chain length of the organic modifier; V_m, organic modifier content (v/v %); pH, mobile phase pH; T, temperature (°C); t_{Rmax}, the analysis time of the most retentive component; R_g, the geometrical mean value of all resolution between two adjacent peaks.

validation set was obtained for five hidden neurons and 2000 training cycles, this architecture was taken as the optimum one for the present study.

Figure 2 shows the correlation plot for both training set and validation set at the optimum architecture of the network. The high correlation coefficients between the calculated and experimental data indicated that the network was well trained. To further test the predictive power and robustness of the model obtained, a crossvalidation approach using the leave-one-out method[37,38] was used. In the leave-one-out method, a model is constructed after deleting one observation from the data set, and then this observation is predicted by the model based on the remaining data. This procedure is repeated for the entire data set and the crossvalidated coefficient is calculated. Figure 3 shows a comparison between the values predicted by ANN and the experimental values. The high R^2 values indicate the good stability and predictive ability of the ANN model developed. As shown in Figure 3, a lower prediction error occurred within the experimental space and a relatively high prediction error occurred near the boundary of the experimental space. This indicates that the ANN has better predictive ability for interpolation. Table 3 shows the linear regression results at the 95% confidence level for Figure 3. Small deviations between the regression lines and ideal ones are apparent (an ideal regression line has a slope of 1.0 and an intercept of 0.0). The significance of small deviations from the ideal slope and the ideal intercept can be assessed using the confidence intervals of the observed slope and intercept. As shown in Table 3, the intervals overlapped the ideal values, which indicated that the differences from ideal were not of any practical importance, that is, they were not statistically significant. These statistical results showed that the ANN model developed here was reliable. Therefore, the predicted values of resolution and analysis time can be used for calculating the CRF in order to locate the optimum operation conditions.

Chromatographic response function

In general, the judgment of the acceptable degree of separation is a compromise between resolution and analysis time. Moreover, the acceptable separation quality is often governed by the purpose of the chromatographic experiment. Because of these circumstances, it is very hard to develop a totally general response function describing separations with different chromatographic aims. The CRF often consists of a time factor and a separation factor. Both of these factors can be modulated by the user to give a performance, in terms of resolution and analysis time, that is, optimized for each specific separation problem. The user can also describe the relative importance of analysis time and resolution by weighting one of the factors in favor of the other, a property that makes the new CRF suitable for chromatography with other purposes.

In this article, the CRF was based on the database generated by ANN, which reflects how the resolution and analysis time changed with operation parameters. By using this

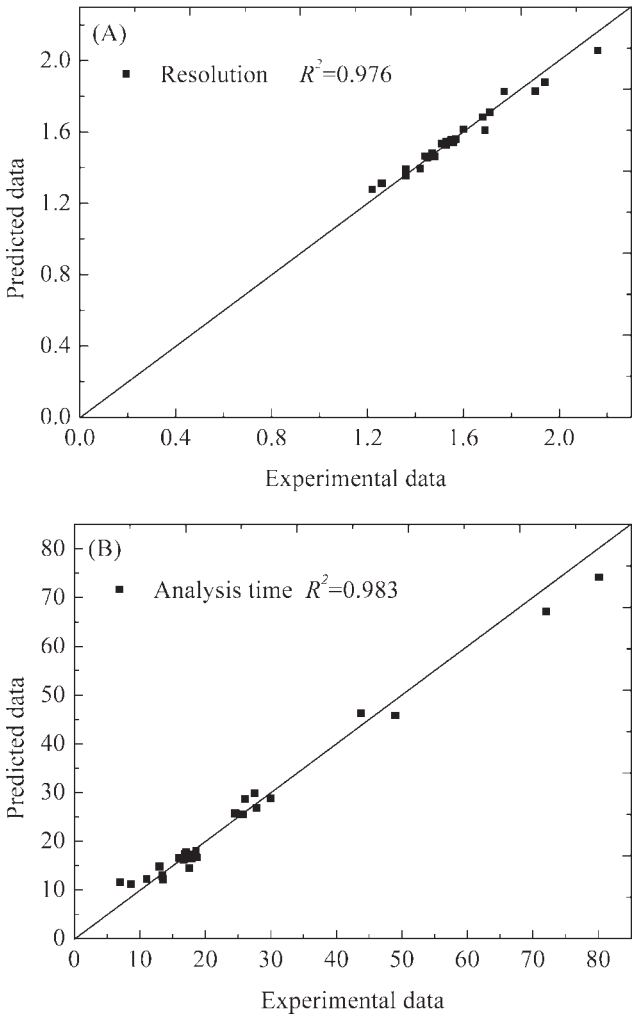

Figure 6. Correlation plot of crossvalidated predicted data by ANN with the literature data.

database, the values of different CRFs for different chromatographic purposes were calculated. Then, the optimum conditions for the different purposes were easily located at the maximum values of the different CRFs. The flexibility and generality were markedly improved by following this strategy.

To study the application of this strategy, it was applied to identify the optimum operating conditions for separation of solanesol from tobacco leaf extracts. Two extreme situations (CRF1 and CRF3) giving preference to resolution and analysis time, respectively, and one intermediate situation (CRF2) with regard to resolution as well as analysis time were investigated. The corresponding CRFs were constructed by modulating the time and separation parameters ($Y_i^{(-)}$ and $Y_i^{(+)}$) and the weighting factor (factor r in Eq. 4). Calculation of the values of CRFs on the basis of predicted data by ANN allowed us to generate three response surfaces for each CRF. As a representative, Figure 4 shows a series of response surfaces for CRF2. From the figure, we can assess how the predicted response changed when varying two operating parameters simultaneously while keeping the third one constant. The optimum operating conditions can then be

Table 6. The Linear Regression Results at 95% Confidence Level for Figure 6

Regression Lines	Terms	LCI	UCI
Resolution			
Slope	0.959	0.902	1.017
Intercept	0.063	−0.027	0.153
Analysis time			
Slope	0.969	0.930	1.008
Intercept	1.062	−0.085	2.209

LCI, lower confidence interval; UCI, upper confidence interval.

located by analysis of these response surfaces. Table 4 shows the CRFs for different chromatographic purposes and corresponding optimum operating conditions.

Method performance

Experiments were conducted under each of the predicted optimum operating conditions in Table 4. The chromatograms obtained under the predicted conditions are shown in Figure 5. Subfigures (A), (B), and (C) were produced under the optimum predicted conditions for CRF1, CRF2, and CRF3, respectively. From the figure, we can see that subfigure (A) possessed the best resolution while the shortest analysis time was obtained in subfigure (C); subfigure (B) is a compromise between analysis time and resolution. These results are in accordance with the initial purposes of the three selected CRFs, which confirms that the proposed strategy is effective and has good flexibility. The predicted and experimental values of the CRFs are also presented in Table 4. The high consistency also indicates the good stability and predictive capability of the strategy proposed.

Further applications

To further apply the strategy and investigate its generality, it was used to predict the optimum chromatographic conditions for the separation of phenylthiohydantoin amino acids

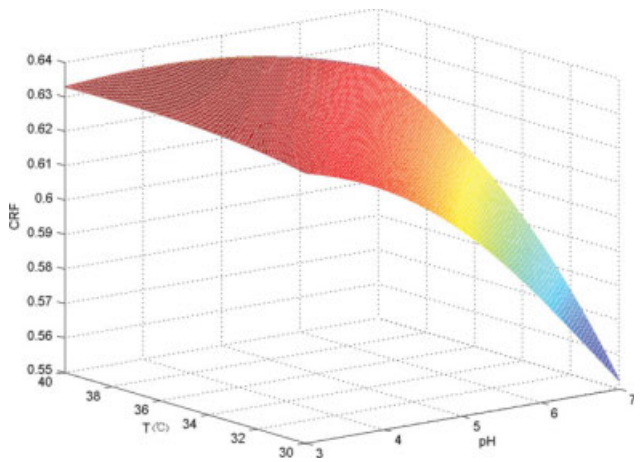

Figure 7. Plot of CRF as a function of pH and *T* in a further application (the separation of phenylthiohydantoin amino acids in micellar liquid chromatography).

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 7. Literature Data and Prediction Results

CRF	$Y^{(-)}$	$Y^{(+)}$	Data	[SDS] (M)	N	V_m (v/v, %)	pH	T (°C)
Parameters								
Resolution	0.5	2.5	Literature data	0.065	4	9.3	3.0	40
Analysis time (min)	10	45	Predicted data	0.071	4	9.57	3.0	40

in micellar liquid chromatography.[10] The experimental data taken from the literature are listed in Table 5. The factors studied in the literature were the concentration of sodium dodecyl sulfate ([SDS], 0.03−0.1 M), alkyl chain length of the alcohol (N) used as the organic modifier, organic modifier content (V_m, 3.0−10.0, v/v %), mobile phase pH, and temperature (T, 30–40°C), which were used as inputs to the ANN model. The outputs we used were the values of analysis time of the most strongly retained component (t_{Rmax}) and the geometrical mean value of all the resolutions between two adjacent peaks (R_g). The optimization of ANN was conducted using the same method described above. Finally, the optimum architecture of the ANN was found to have five neurons in the input layer, three neurons in the hidden layer, and two neurons in the output layer. Crossvalidation using the leave-one-out method was also used to test the robustness and predictive ability of the optimized ANN. The data predicted by ANN, together with the literature data, are shown in Figure 6. The correlation coefficient indicates that the ANN can predict the literature data with a high degree of accuracy. Table 6 shows the linear regression results at 95% confidence level for Figure 6. Although there were small deviations between the regression lines and the ideal ones, the ideal values of the slope and intercept fell within the confidence intervals, which indicates that the deviations were not significant. These statistical results confirm that the ANN model developed was reliable. Based on the predicted values of t_{Rmax} and R_g, the values of Derringer's desirability function (D) were calculated and furthermore the changes in D as a function of chromatographic conditions ([SDS], N, V_m, pH, T) were investigated. Figure 7 shows the plot of CRF as a function of pH and T when the other parameters were kept constant. As shown in the figure, the maximum CRF value was obtained at a pH value of 3 and temperature of 40°C, and the CRF changed but slightly when the values of pH and T were close to 3 and 40°C. However, with increasing pH and decreasing T, the CRF changed dramatically, which indicates the existence of a significant two-factor interaction between pH and T. These results are in good agreement with the observations by Safa and Hadjmohammadi in the literature.[10]

The optimum chromatographic conditions were located at the maximum value of the desirability function. A comparison of the literature data with the data predicted by the proposed strategy is shown in Table 7. The results showed that the predicted data are consistent with the literature data, which confirms that the optimization strategy proposed in this article has good reliability and generality.

Conclusions

The sequential combination of an ANN with a CRF has been found to be a general and flexible tool for the optimization of chromatographic separation in RP-HPLC. A modified star design was carried out to provide unbiased experimental data for the ANN modeling. The results showed that the optimized ANN can simultaneously predict analysis times and resolutions with good accuracy. Subsequently, the predicted values of the analysis times and resolutions were used to calculate the CRF. The optimum operating conditions for different chromatographic purposes were obtained at the maximum values of the different CRFs. The corresponding experimental chromatograms were recorded under the predicted optimum conditions and the good agreement between experimental and predicted results confirmed the reliability of the proposed strategy. To further apply the strategy and investigate its generality, it was used to predict the optimum chromatographic conditions for the separation of phenylthiohydantoin amino acids in micellar liquid chromatography using the literature data. The results showed that the predicted data were consistent with the literature data, which indicates the good generality of the proposed strategy.

Acknowledgments

The authors would like to acknowledge financial support from the Foundation of Key Laboratory of Bioprocess of Beijing (Grant No. SYS 100100421), the National Natural Science Foundation of China (Grant No. 20576010), and the Program for New Century Excellent Talents (NCET-05-0117). They also like to express the gratitude to Professor David G. Evans for his kindly help in language improvement.

Literature Cited

1. Gonzalez A, Foster KL, Hanrahan G. Method development and validation for optimized separation of benzo-alpha-pyrene-quinone isomers using liquid chromatography-mass spectrometry and chemometric response surface methodology. *J Chromatogr A*. 2007;1167:135–142.
2. Fraga CG. Chemometric approach for the resolution and quantification of unresolved peaks in gas chromatography-selected-ion mass spectrometry data. *J Chromatogr A*. 2003;1019:31–42.
3. Fraga CG, Bruckner CA, Synovec RE. Increasing the number of analyzable peaks in comprehensive two-dimensional separations through chemometrics. *Anal Chem*. 2001;73:675–683.
4. Jimidar M, Bourguignon B, Massart DL. Application of Derringer's desirability function for the selection of optimum separation conditions in capillary zone electrophoresis. *J Chromatogr A*. 1996;740:109–117.
5. Li H, Hou J, Wang K, Zhang F. Resolution of multicomponent overlapped peaks: a comparison of several curve resolution methods. *Talanta*. 2006;70:336–343.
6. Marengo E, Gennaro MC, Gianotti V. Chemometrically assisted simultaneous separation of 21 aromatic sulfonates in ion-interaction RP-HPLC. *Chemom Intell Lab Syst*. 2000;53:57–67.
7. Pace G, Berton A, Bergamaschi L. Bi-quadratic surface response for quantitative determination of analytes leading to partially overlapped chromatographic peaks. *J Chromatogr A*. 2001;907:81–88.
8. Pierce KM, Wood LF, Wright BW, Synovec RE. A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data. *Anal Chem*. 2005;77:7735–7743.

9. Prazen BJ, Synovec RE, Kowalski BR. Standardization of second-order chromatographic/spectroscopic data for optimum chemical analysis. *Anal Chem*. 1998;70:218–225.
10. Safa F, Hadjmohammadi MR. Simultaneous optimization of the resolution and analysis time in micellar liquid chromatography of phenyl thiohydantoin amino acids using Derringer's desirability function. *J Chromatogr A*. 2005;1078:42–50.
11. Snyder LR, Kirkland JJ, Glach JL. *Practical HPLC Method Development*, 2nd ed. New York: Wiley-Interscience, 1997.
12. Glajch L, Snyder LR. *Computer-Assisted Method Development for High-Performance Liquid Chromatography*. Amsterdam: Elsevier, 1990.
13. Cela R, Martinez JA, González-Barreiro C, Lores M. Multi-objective optimisation using evolutionary algorithms: its application to HPLC separations. *Chemom Intell Lab Syst*. 2003;69:137–156.
14. Smilde AK, Knevelman A, Coenegracht PMJ. Introduction of multi-criteria decision making in optimization procedures for high-performance liquid chromatographic separations. *J Chromatogr*. 1986; 369:1–10.
15. Nsengiyumva C, De Beer JO, Van de Wauw W, De Swaef S, Parmentier F. Resolution optimization in micellar electrokinetic chromatography: use of overlapping resolution mapping scheme in the analysis of dinitrophenyl derivatized amino acids. *Chromatographia*. 1993;35:560–566.
16. Bourguignon B, Massart DL. Simultaneous optimization of several chromatographic performance goals using Derringer's desirability function. *J Chromatogr A*. 1991;586:11–20.
17. Glajch JL, Kirkland JJ. Method development in high-performance liquid chromatography using retention mapping and experimental design techniques. *J Chromatogr*. 1989;485:51–63.
18. Coenegracht PMJ, Smilde AK, Metting HJ, Doornbos DA. Comparison of optimization methods in reversed-phase high-performance liquid chromatography using mixture designs and multi-criteria decision making. *J Chromatogr*. 1989;485:195–217.
19. Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi PJ. *Handbook of Chemometrics and Qualimetrics, Part A*. Amsterdam: Elsevier Science BV, 1997.
20. Morris VM, Hughes JG, Marriott PJ. Examination of a new chromatographic function, based on an exponential resolution term, for use in optimization strategies: application to capillary gas chromatography separation of phenols. *J Chromatogr A*. 1996;755: 235–243.
21. Glajch JL, Kirkland JJ, Squire KM, Minor JM. Optimization of solvent strength and selectivity for reversed-phase liquid chromatography using an interactive mixture-design statistical technique. *J Chromatogr*. 1980;199:57–79.
22. Berridge JC. Unattended optimization of reversed-phase high-performance liquid chromatographic separations using the modified simplex algorithm. *J Chromatogr*. 1982;244:1–14.
23. Dose EN. Off-line optimization of gas chromatographic temperature programs. *Anal Chem*. 1987;59:2420–2423.
24. Bylund D, Bergens A, Jacobsson SP. Optimisation of chromatographic separations by use of a chromatographic response function, empirical modelling and multivariate analysis. *Chromatographia*. 1997;44:74–80.
25. Harrington EC. The desirability function. *Ind Qual Control*. 1965;21:494–498.
26. Derringer G, Suich R. Simultaneous optimization of several response variables. *J Qual Technol*. 1980;12:214–219.
27. Sáiz-Abajo MJ, González-Sáiz JM, Pizarro C. Multi-objective optimisation strategy based on desirability functions used for chromatographic separation and quantification of L-proline and organic acids in vinegar. *Anal Chim Acta*. 2005;528:63–76.
28. Dewe W, Marini RD, Chiap P, Hubert P, Crommen J, Boulanger B. Development of response models for optimising HPLC methods. *Chemom Intell Lab Syst*. 2004;74:263–268.
29. Carda-Broch S, Torres-Lapasió JR, Garcia-Alvarez-Coque MC. Evaluation of several global resolution functions for liquid chromatography. *Anal Chim Acta*. 1999;396:61–74.
30. Bolanca T, Cerjan-Stefanovic S, Novic M. Application of artificial neural network and multiple linear regression retention models for optimization of separation in ion chromatography by using several criteria functions. *Chromatographia*. 2005;61:181–187.
31. Novotna K, Havlis J, Havel J. Optimisation of high performance liquid chromatography separation of neuroprotective peptides: fractional experimental designs combined with artificial neural networks. *J Chromatogr A*. 2005;1096:50–57.
32. Giddings JC. *Dynamics of Chromatography. I. Principles and Theory*. New York: Marcel Dekker, 1965.
33. Snyder LR. *Principles of Adsorption Chromatography*. New York: Marcel Dekker, 1968.
34. Sander LC, Wise SA. Subambient temperature modification of selectivity in reversed-phase liquid chromatography. *Anal Chem*. 1989; 61:1749–1754.
35. Sander LC, Pursch M, Wise SA. Shape selectivity for constrained solutes in reversed-phase liquid chromatography. *Anal Chem*. 1999; 71:4821–4830.
36. Dolan JW. Temperature selectivity in reversed-phase high performance liquid chromatography. *J Chromatogr A*. 2002;965:195–205.
37. Osten DW. Selection of optimal regression models via cross-validation. *J Chemom*. 1988;2:39–48.
38. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA) 1 effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc*. 1988;110:5959–5967.

Manuscript received Mar. 24, 2008, and revision received Jun. 8, 2009.