Optimization of Gradient Separation for Chromatographic Fingerprint of Herbal Medicine: A Predictive Model Combined with Grid Search

Xiaoying Yin, Weifeng Yao, and **Yuzhu Hu***
Department of Analytical Chemistry, China Pharmaceutical University, Nanjing 210009, China

Abstract

A four-step strategy is described which is aimed at optimizing the separation parameters of the chromatographic fingerprint of herbal medicines. In the optimization procedure, a model is built to predict the retention time of components. The proposed model shows good retention prediction potency with an average relative deviation of 1.78%. The grid search method is used in optimizing gradient parameters. In addition, an iterative procedure is adopted to further improve the prediction accuracy of the gradient retention time and the quality of the chromatogram. The whole optimization process has been validated by real samples.

Introduction

The chromatographic fingerprint technique has been accepted as a universal approach for evaluating the stability and homogeneity of herbal medicines (1–2). A single linear gradient profile is usually preferred to separate complicated compounds in herbal medicine, so optimization of the gradient conditions is a crucial step in fingerprint separation with good sensitivity, feasibility, and reproducibility.

Many methods have been used to optimize parameters of linear gradient elution (3–6). The main optimization method uses predictive models. These models can be generally classified into two categories, theoretical models and chemometric models. Theoretical models are mathematical expressions derived from the retention theory in gradient elution, such as the widely used linear solvent strength (LSS) model proposed by Snyder and co-workers (7–8), and its developed models (9–11). It was defined as:

$$t_R = \frac{t_0}{b} \log (2.3k_0 b + 1) + t_0 + t_D \qquad \text{Eq. 1}$$

where $b = t_0 \Delta\varphi S / t_G$, t_D is the gradient delay time, t_0 is the column dead-time, and k_0 is the capacity factor in the initial mobile phase composition of the gradient; $\Delta\varphi$ is the change in the mobile phase composition, t_G is gradient time, and S is a constant for a given solute and an organic solvent.

In the chemometric models, a regression analysis based on the chromatographic behavior and molecular structure of solutes, namely the quantitative structure retention relationship (QSRR) model, is usually used. QSRR derives relationships between chromatographic parameters and the descriptors characterizing molecular structure of the solutes. Three main types of the QSRR model have been reported in the literature, such as the the linear solvation energy relationship-based model (12–13), the logP-based (*n*-octanol–water partition coefficients) model (14), and models based on quantum chemical indices or other structural descriptors (15). These empirical models show good retention prediction potency, but require structure parameters of solutes that obviously are not all applicable for the possible solutes (e.g., unknown components in herbal medicine). A business-objective-based, constraint-based method for optimizing the operational parameters of high-performance liquid chromatography (HPLC) was presented by Chester (16). It is flexible to optimize isocratic separations, but is difficult to optimize separations of complicated components with retention factor k of a wide range.

In this work, we propose a straightforward but efficient multiple regression model. It can accurately predict retention times under different gradient conditions of experimental domain. It especially suits the retention prediction of complex samples (e.g., herbal medicines). Moreover, a grid search algorithm was applied to optimize parameters of gradient conditions and an iterative procedure was used to further improve the accuracy of prediction. Along with a reasonable experimental design and optimization criteria, the optimal conditions can be obtained. This strategy has been used to optimize the separation of a real herbal medicine.

* Author to whom correspondence should be addressed.

Algorithm

Predictive model based on multiple regressions

Successful application of the multiple regression analysis depends on the selection of reasonable and independent variables. In a gradient system, many factors can affect the chromatographic behavior of a component (e.g., the stationary phase, temperature, the organic solvent, etc.). Some of them can be determined by preliminary experiments in practice. Only a few factors that affect separation markedly need to be considered for further optimization. In a linear gradient elution process, the three gradient parameters, B_0, B_i, and t_G, namely initial, final organic solvent concentration, and gradient time, play the key roles in the retention behavior of solutes. Therefore, the three gradient parameters were selected as the optimization parameters and independent variables.

The best fitting equation between the retention time and the three parameters (B_0, B_i, and t_G) was explored by a least squares regression. A search needs to be performed to find the best fitting model from the simplest linear regression to more complex polynomial regressions. The linear regression has the following form:

$$\log \frac{t_{R_i}}{t_{R_s}} = \alpha_0 + \alpha_1 B_0 + \alpha_2 T_G + \alpha_3 B_i \qquad \text{Eq. 2}$$

where t_{Ri} is the retention time of compound i, i is peak index; t_{RS} is the retention time of the reference peak used to minimize the error resulting from column conditioning effects; α_0 to α_3 are regression coefficients; $T_G = t_G / 100$ is for the dimension consistency of the three parameters in the equation.

The second-order polynomial model is explored as well. The third- and higher-order variable interactions can be omitted because they were shown to be statistically unimportant (17). The second-order polynomial model involving three variables can be described as below:

$$\log \frac{t_{R_i}}{t_{R_s}} = \beta_0 + \beta_1 B_0 + \beta_2 T_G + \beta_3 B_{i+} \beta_4 B_0 T_G + \beta_5 B_0 B_i + \beta_6 T_G B_i + \beta_7 B_0^2 + \beta_8 T_G + \beta_9 B_i^2 \qquad \text{Eq. 3}$$

where β_0 to β_9 are regression coefficients. There are nine variables in Equation 3, so a stepwise regression option should be performed to choose variables entering the regression equation according to a defined criterion or the contribution to the dependent. With the criterion of the probability of F-to-enter ≤ 0.05, the stepwise regression analysis was carried out using the statistical software SPSS. The regression result is that the strong relevant variables in the model are B_0, B_i, and t_G, the same as that of Equation 2. Hence the relationship between logarithm of the relative retention time and the gradient variables is a linearly correlated rather than a second-order relationship. The final fitting model can be expressed by Equation 2.

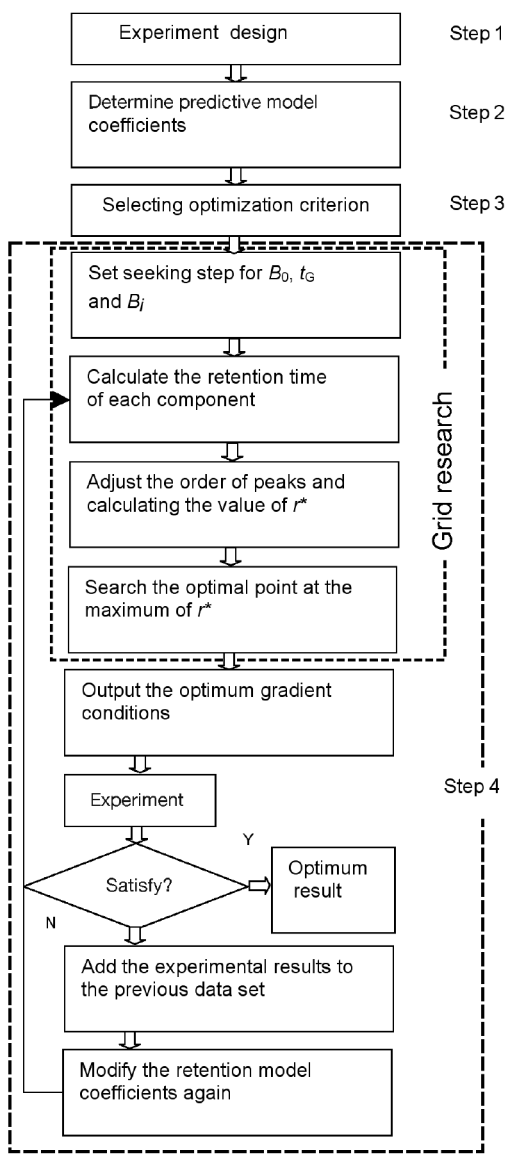

Figure 1. Outline of the optimization procedure.

Optimization strategy

A four-step optimization strategy was used to achieve optimized gradient conditions. Figure 1 shows the flow chart of the optimization procedure and the details are described below.

Experimental design

In this work, the Chinese traditional medicine *Pueraria thomsonii* was used as the sample. It contains a high content of flavonoid derivatives, of which puerarin is most abundant. Reversed-phase LC methods with ODS columns and methanol–water mobile phases were recommended to separate this kind of substance (18). All chromatographic parameters, including column temperature, detection wavelength, injection volume, and flow rate were selected by preliminary experiments. According to the preliminary experiments, the selected factors (B_0, B_i, and t_G) and their optimization ranges were set to be 0.16–0.28% (v/v), 0.52–0.76% (v/v), and 34–58 min, respectively.

In order to test the rule between the solute retention and the gradient parameters, a great deal of experiments are necessary. Therefore, selecting the typical experiments, which are enough to represent the whole range of parameters, is required by statistical methods. The technique for experimental design, named the uniform design, proposed by K.T. Fang in the 1980s (19–20)

based on a quasi-Monte Carlo method or number-theoretic method, was selected to design the starting experiments. The uniform design allocates the experimental points regularly and uniformly, so it is a good choice.

Determining predictive model coefficients

A predictive model that accurately describes the effect of variables on solute retention is necessary. In this work, a predictive model was built based on multiple regressions. The retention equation of each component can be expressed by Equation 2. Coefficients of the retention model (α_0 to α_3) were obtained by multiple regression analysis based on the data of starting experiments. Then, the retention time of solutes can be predicted under any gradient conditions in the optimization domain.

Selecting optimization criterion

Before the optimization begins, a quality criterion must be defined to reflect the desirability of separation. Many quality criterion functions have been used in HPLC optimization, such as hierarchical chromatographic response function (HCRF) (21), chromatographic response function (CRF) (22), and global resolution function (r*) (23–24). HCRF was defined as:

$$\mathrm{HCRF} = 1{,}000{,}000n + 100{,}000R_{min} + (100 - t_L) \qquad \text{Eq. 4}$$

where n is the number of peaks in the chromatogram, R_{min} is the resolution of the least separated pair of peaks, and t_L is the retention time of the last peak. CRF can be expressed as following:

$$\mathrm{CRF} = \sum_{i=1}^{L} R_i + L^{W_1 - W_2}\left|T_A - T_L\right| - W_3(T_0 - T_1) \qquad \text{Eq. 5}$$

where R_i is the resolution between the ith and the $(i + 1)$th peaks, L is the number of peaks appearing in the chromatogram, T_A, T_L, T_1, and T_0 are the maximum acceptable time, retention time of the final peak, retention time of the first peak, and the minimum retention time of the first peak, respectively, and w_1 to w_3 are weighting factors selected by the analyst. Global resolution function r* (23,24) is calculated based on:

$$r^* = \prod_{i=0}^{n-1}\left(\frac{S_{i,i+1}}{\overline{S}_{i,i+1}}\right) \qquad \text{Eq. 6}$$

where separation factor $s_{i,i+1} = (t_{i+1} - t_i) / (t_{i+1} + t_i)$, $S_{i,i+1} = 1 / n \times \sum (S_{i,i+1})$, t_{i+1}, t_i are the retention time of adjacent peaks, n is the total number of peaks, and $i = 0$ to n-1 is peak index. In the fingerprint of an herbal medicine, the overlapping of peaks often appears and accurate determination of peak widths becomes difficult. However, the separation factor $S_{i,i+1}$, is independent of the column efficiency and its calculation requires only retention time. So the basic quality criterion based on $S_{i,i+1}$ seems more appropriate.

For chromatographic fingerprints of herbal medicines, we pay more attention to three aspects: a large number of peaks of pharmacologically active components, a good separation of complex mixtures, and a short analysis time. HCRF excessively focuses on the effect of the resolution for the least separated pair of peaks while ignoring distribution of other peaks. For CRF, the quality of the chromatogram is determined by well-resolved peaks while the poorly resolved peaks do not influence much of the function value. The global resolution function r* aims at a uniform distribution of the detected peaks in a short analysis time. r* seems not to reflect the effect of peak number. In fact, the peak number is the largest when r* reaches the maximum. Obviously, r* is more suitable in herbal medicine fingerprint separations, and was chosen as the optimization criterion.

Searching for the optimization point

An appropriate optimization algorithm can find the desired value of the selected criterion function within the experimental domain as well as corresponding optimal experimental parameters. In this work, the principles for selection of an optimization algorithm include a short run time, a simple algorithm, and obtaining a global optimum. The grid search algorithm is an appropriate alternative to be employed to search for optimal separation conditions. The search procedure consists of: (*i*) setting step-length to each parameter. Namely, each parameter is divided into different grid numbers in the optimization domain. Thus the simulative experimental points can be obtained. (*ii*) Calculating the retention time of each solute in different experimental points. (*iii*) Adjusting the order of peaks according to the sequence of time and calculating the value of r* under each experimental point. (*iv*) Screening the whole experimental points in the optimization domain and selecting the maximum value of r*; its corresponding gradient parameters are the optimum conditions.

Once the optimization points are obtained by the previously mentioned optimization steps, experiments can then be performed. If the experiment results obtained are satisfied, the optimization process is stopped. Otherwise, an iterative operation needs to be sequentially performed to improve the separation. The iterative procedure includes two steps. In the first step, the experiment data presently obtained will be added into the previous data set and the regression coefficients of the retention model can then be modified. In the second step, the grid search procedure as stated previously will be carried out again, and then a new optimal point can be obtained. The iterative process should be repeated until the experiment result is satisfied. The end requirements are: (*i*) $| r_{n+1}^* - r_n^* | < 0.01$, where r_{n+1}^* is the value of r* from the new point; r_n^* is the value of r* from the previous optimization point; (*ii*) $|B0_{n+1} - B0_n| < 0.5\%$, $|TG_{n+1} - TG_n| < 0.005\%$ and $|Bi_{n+1} - Bi_n| < 0.5\%$, where $B0_{n+1}$, TG_{n+1}, and Bi_{n+1} are the gradient parameters of the new optimal point; $B0_n$, TG_n, and Bi_n are that of the previous ones.

Experimental

HPLC instrumentation

The gradient elution of samples was performed with a Shimadzu LC-10AD HPLC system (Shimadzu, Kyoto, Japan). It is equipped with two low-pressure pumps (LC-10ADvp), a DGU-12A degasser, and a SPD-M10Avp diode-array UV detector, and operated with a CLASS-vp workstation (Shimadzu). A Hypersil C_{18} column (150 mm × 4.6 mm i.d., 5 µm particle size, Dalian, China) was used. The dead time (t_0) of the column was determined with uracil as the marker. The system's dwell-time (t_D)

Table I. Conditions of Gradient Experiments and Corresponding Values of HCRF, CRF, and *r

Test no.*	Factor B_0 (%)	Factor t_G (min)	Factor B_i (%)	HCRF	CRF†	*r**
1	20	58	64	14000962.9	1.837	0.7207
2	26	54	72	14001335.1	0.626	0.7866
3	18	34	68	13000889.9	0.159	0.6433
4	16	50	56	13000798.9	2.204	0.5426
5	22	38	76	14001012.8	0.172	0.7302
6	24	46	52	13001071.2	1.589	0.6811
7	28	42	60	13001508.4	0.693	0.5312
8	24	45	65	14001190.7	0.713	0.7919
9	25	48	62	14001312.5	0.518	0.8037
10	25	48	62	14001312.5	0.518	0.8037
11	17	44	70			
12	23	40	66			
13	25	45	65			

* Experiments 1 to 7 are starting experiments by uniform design. 8 to 10 show the number of optimization experiments; 11 to 13 list the experiment conditions to validate the robustness of model.
† T_A = 60 min, T_O = 1 min, $W_1 = W_2 = W_3 = 0.1$.

Table II. The Multiple Regression Equation Coefficients and Goodness of Fit Assessment from the Initial Data Sets and Enlarged Data Sets Adding Optimal Points

Peak no.	Coefficients α_0	α_1	α_2	α_3	Statistical parameters R	SE	F	P
1	–0.6231	0.6598	–0.3001	0.2084	0.9394	0.0232	7.513	0.0659
2	–0.0686	–1.1750	–0.2249	0.1829	0.9797	0.0166	23.94	0.0135
3	–0.0098	–1.1570	–0.1405	0.1689	0.9979	0.0049	236.3	0.0005
4	–0.1270	–0.1700	–0.1268	0.1252	0.9263	0.0105	6.046	0.0868
5	–0.2750	0.8457	–0.1662	0.1507	0.9613	0.0167	4.060	0.3460
6	1.7056	–2.5017	0.2986	–0.2801	0.9971	0.0124	173.6	0.0007
7	0.0471	0.1068	0.0818	–0.0815	0.9572	0.0051	10.94	0.0401
8	0.0230	0.5099	0.0666	–0.0965	0.9995	0.0011	978.7	0.0000
9	0.0358	0.6954	0.0864	–0.1401	0.9996	0.0012	1423	0.0000
10	0.0646	0.7210	0.1353	–0.1956	0.9985	0.0029	334.4	0.0003
11	0.0621	1.0951	0.1391	–0.2239	0.9996	0.0020	1384	0.0000
12	0.0637	1.1473	0.1388	-0.2292	0.9996	0.0023	1145	0.0000
13	0.1046	1.9508	0.2414	-0.4045	0.9995	0.0044	923.9	0.0000
14	0.1687	2.1815	0.2763	-0.4473	0.9993	0.0081	252.4	0.0462
	The modified coefficients				Statistical parameters after modification			
1	–0.6500	0.6158	–0.3467	0.3039	0.9601	0.0184	11.78	0.0363
2	–0.0789	–1.2133	–0.2440	0.2258	0.9765	0.0179	20.52	0.0168
3	–0.0098	–1.1667	–0.1411	0.1718	0.9969	0.0058	162.8	0.0008
4	–0.1335	–0.1829	–0.1382	0.1489	0.9168	0.0101	5.268	0.0903
5	–0.2744	0.8337	–0.1640	0.1517	0.9624	0.0119	8.370	0.1085
6	1.7260	–2.4827	0.3331	–0.3483	0.9996	0.0048	1178	0.0000
7	0.0509	0.1143	0.0886	–0.0955	0.9529	0.0047	9.871	0.0000
8	0.0224	0.5101	0.0657	–0.0948	0.9995	0.0011	921.9	0.0000
9	0.0345	0.6931	0.0842	–0.1356	0.9997	0.0011	1702	0.0000
10	0.0660	0.7219	0.1375	–0.1999	0.9984	0.0028	303.9	0.0003
11	0.0601	1.0912	0.1356	–0.2167	0.9997	0.0018	1573	0.0000
12	0.0614	1.1425	0.1346	–0.2205	0.9996	0.0020	1335	0.0000
13	0.1416	1.9427	0.2337	–0.3886	0.9996	0.0038	1113	0.0000
14	0.1659	2.1786	0.2771	–0.4434	0.9993	0.0060	470.3	0.0021

was determined by running a blank gradient where 0.3% acetone was increased from 0% to 1% in 20 min as previously reported (25). The average values of t_0 and t_D were 2.688 ± 0.015 min (average ± SD, n = 6) and 2.530 ± 0.066 min (average ± SD, n = 6), respectively.

Chemicals and reagents

The herbal sample of *Pueraria thomsonii* (Fen-Ge in Chinese) was purchased from the Beijing Tongrentang herbal shop. Standard puerarin was purchased from Chinese National Institute for Control of Pharmaceutical and Biological Products (Beijing, China), lot number: 752-200108. Chromatographic-grade methanol was purchased from Hanbang Science & Technology Company. Other reagents were of analytical grade or better and of commercial availability. Double distilled water was used throughout.

Chromatographic parameters

The optimized chromatographic procedure consists of a linear gradient elution using a mixture of 0.1% glacial acetic acid aqueous solution and methanol as the mobile phase. The flow rate was 1.0 mL/min, and the injection volume was 20 µL. The column temperature was maintained at 30ºC. The detection wavelength was set at 250 nm.

Sample preparation

Approximately 0.8 g *Pueraria thomsonii* powder were weighed and then transferred to a 100-mL conical beaker with padding. A 50.0 mL 30% methanol solution was then added to the beaker, followed by a 30 min sonication. After cooling to 25ºC, the solution was filtered firstly through a filter paper and then a 0.45 µm filter. The filtrate was kept for HPLC separation.

Software

The grid search algorithm was written in Matlab7.0 (The Mathworks Inc.) by the authors and run on a PC with 600MHz, 256M Memory Lenovo P_4 processor PC operated with Windows XP/2000. Regression analysis was performed using SPSS11.5 software (SPSS Inc.) and graphical outputs were produced by Origin7.5 (Originlab Corp.).

Results and Discussion

Fitting of experimental data to regression model

According to the uniform design method (18–19), the uniform table U_7 (7^3) (26) is selected. Because B_0, B_i, and t_G are continuous quantitative factors and homogeneously fixed at seven levels, seven experiments were subsequently designed. The sequence numbers from 1 to 7 are the starting experiments in Table I. By

performing the designed experiments, 14 components were separated from the chromatographic fingerprint of *Pueraria thomsonii* extracts. Among them, the 6th peak was identified as puerarin by comparing with standard puerarin, and taken as the reference peak. Experimental data of each component were then fitted using Equation 2 by least squares approximation. Subsequently, 14 total regression models can be obtained and their coefficients and statistical results are listed in Table II. The modified coefficients (obtained after iterative optimization, see below for details) and the corresponding statistical results are also listed in Table II.

From the coefficients of these equations in Table II, we found that $|\alpha_1|$ is much larger than $|\alpha_2|$ and $|\alpha_3|$. This indicates that the contribution of B_0 is the largest in the equation, therefore B_0 is the key factor. These statistical results show that multiple correlation coefficients (R) of most compounds are near to 1 and their corresponding F-test value is highly significant. Take the number 9 peak as an example (see Table II), multiple correlation coefficients (R = 0.9996), standard error of estimate (SE = 0.0012), the significance level of the whole equation (P = 0.000), and the value of F-test of significance (F = 1423) are all very good. Therefore, we can conclude that the regression model is significant and its data fitting is satisfied. It also indicates that the logarithm of the relative retention time and the gradient variables (B_0, B_i, and t_G) are highly linearly correlated. A few bad data can also be found in Table II; for example, peak 5, with F = 4.060, P = 0.3460 (P > 0.05) shows statistically insignificant. The reason is that when some peaks appear crossing or overlapping, their fitting results may appear a statistically quality descending.

Comparison of our regression model with the LSS model

A model with a solid goodness-of-fit measure may not perform as well as expected. In order to estimate the robustness of the model, its predictive accuracy should be further determined. Thus another three gradient experiments were carried out to validate the accuracy of retention predicting in the optimization region. Table I lists the experimental conditions of experiments 11 to 13. The LSS model (Equation 1) is also employed to predict the retention time of each component under the three gradient conditions. Parameters k0 and S and in the LSS model of each solute are calculated by a non-linear regression using SPSS11.5 or the Microsoft Solver (27) in Microsoft Excel with the data of experiments.

The predictive results by these different models are shown in Table III. We evaluate the predictive power of a model by comparing the relative deviations. It can be seen that the maximal

Table III. Comparison of Predictive Power by the Multiple Regression and LSS Models Under Three Gradient Conditions

	The 11th experiment				
Peak no.	Exp.* (min)	MLR (min)	LSS (min)	EM (%)	EL (%)
1	5.387	5.231	4.351	2.90	–8.80
2	9.355	9.469	6.869	–1.22	26.6
3	11.211	11.628	8.569	–3.72	23.6
4	12.213	12.557	9.698	–2.81	20.6
5	12.981	13.084	11.442	–0.79	11.9
6	16.968	16.429	13.054	3.18	23.1
7	17.653	18.189	14.551	–3.03	17.6
8	19.008	19.366	16.044	–1.88	15.6
9	20.117	20.395	17.170	–1.38	14.6
10	20.853	21.155	17.829	–1.45	14.5
11	23.243	23.349	20.342	–0.46	12.5
12	23.637	23.714	20.756	–0.33	12.2
13	33.333	32.835	30.878	1.49	7.40
14	37.429	36.611	35.327	2.18	5.60
E% (max)†				3.72	26.6
E% (ave)				1.92	15.3
	The 12th experiment				
Peak no.					
1	4.000	4.083	4.274	2.09	–13.8
2	5.717	5.710	6.724	0.12	–17.6
3	7.232	6.984	8.543	3.43	–18.1
4	8.715	8.719	9.672	–0.05	–11.0
5	10.613	10.406	11.532	1.95	–8.7
6	12.053	11.608	13.225	3.69	–9.7
7	13.571	13.043	14.878	3.89	–9.6
8	15.189	14.722	16.442	3.07	–8.2
9	16.384	15.943	17.672	2.69	–7.9
10	17.077	16.605	18.446	2.76	–8.0
11	19.712	19.343	21.125	1.87	–7.2
12	20.149	19.798	21.568	1.74	–7.0
13	30.859	30.837	32.670	0.07	–5.9
14	35.435	35.523	37.667	–0.25	–5.5
E% (max)				3.89	18.0
E% (ave)				1.98	9.90
	The 13th experiment				
Peak no.					
1	3.755	3.757	4.416	–0.04	18.0
2	4.885	4.873	6.545	0.25	–34.0
3	6.165	6.025	8.506	2.27	–38.0
4	7.861	7.908	9.634	–0.59	–22.6
5	10.069	9.819	11.668	2.48	–15.9
6	11.083	10.776	13.488	2.77	–21.7
7	12.683	12.306	15.403	2.97	–21.4
8	14.443	14.131	17.083	2.16	–18.3
9	15.797	15.484	18.493	1.98	–17.1
10	16.587	16.259	19.477	1.98	–17.4
11	19.552	19.280	22.446	1.34	–14.8
12	20.043	19.792	22.939	1.25	–14.1
13	32.544	32.502	35.839	0.13	–10.1
14	38.005	38.029	40.953	–0.06	–7.8
E% (max)				2.97	38.0
E% (ave)				1.45	19.4

* Exp., MLR, and LSS represent experimental value, prediction value by the proposed model, and LSS model of the retention time, respectively.
† E% is the relative deviation.

relative deviation and average deviation of the retention time by our regression model are much lower than the ones by the LSS model for the three experimental data, and the predicted value of each peak by our regression model is closer to the actual value. The predictive accuracy by our regression model is satisfactory with the maximal relative deviation of 3.89% and an average relative deviation of 1.78%. Ninety-five percent confidence intervals of the relative deviation by our regression model and LSS model are 1.78% ± 0.37% (n = 42) and 14.9% ± 2.3% (n = 42), respectively.

Table III shows that the relative deviation of the retention time predicted by the LSS model is too high, with an average relative deviation of 14.9%. To find these possible sources of error, the LSS model was further investigated. The LSS model was built based on the theory of gradient elution separation. The LSS model assumes a linear relationship between the logarithm of the retention factor k and the organic solvent concentration, and an ideal gradient condition of the gradient system (7). For example, no time delay between the gradient mixer and sample injector and solutes not moving along the column during tD. However, this assumption is not true for an actual gradient elution system. In fact, the deviation can come from a non-linear elution of components and a non-ideal gradient system. The non-ideality of a gradient mainly includes two effects: the instrument error and in-column factors (28). The former effect includes (i) the solvent misproportion and flow-rate errors due to pump design; and (ii) the gradient delay due to the volume of the gradient mixer and the connecting tubing, etc. The latter includes (i) the change of dead-time during a gradient elution; (ii) uptake of organic modifier on the stationary phase and inducing a kind of extra dwell time; and (iii) solute molecules exhibiting size-exclusion effects. In summary, the accuracy of the LSS model is related to the precision of an instrument, the character of column, and the property of solute molecules. Table III also indicates that the stronger the retention of solutes, the higher the predictive accuracy of the LSS model. It implies that delay in systems and in columns largely affects solutes of weak retention and leads to a large deviation. Although the LSS model has been applied successfully to predict the retention of many solutes, in this work, the results also demonstrate that gradient retention time cannot be accurately calculated with the LSS model.

Besides, compared to the LSS model, this proposed model also has many other advantages. It requires no prior knowledge of properties of the mixture or the retention theory of gradient elution, no consideration of the effects from the instrument error and in-column factors, and no determination of the system parameters, such as the dead-time and the dwell time. Only a little statistical knowledge is necessary to predict the retention. Also, the model is simpler and the regression coefficients are easier to be obtained than the LSS model. These advantages are especially fit for the retention prediction of complex mixtures like fingerprints of herbal medicines.

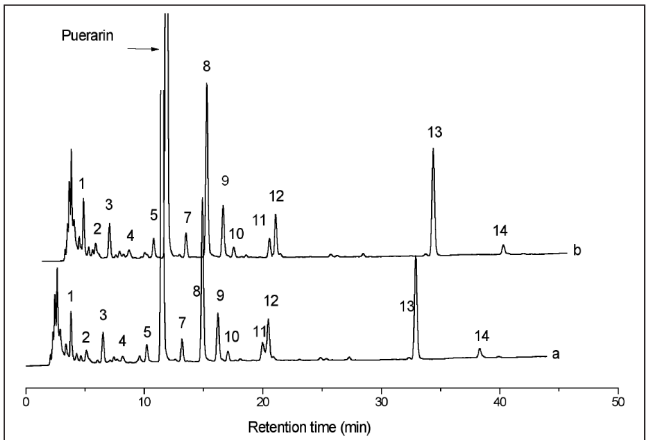

Figure 2. Chromatograms of two optimal points. Gradient condition at the first optimal point (the 8th experiment): initial organic solvent concentration B_0 is 24%, gradient time t_G is 45 min, and final organic solvent concentration B_i is 65% (A). Experimental conditions at the second optimum point (the 9th experiment): B_0 is 25%, t_G is 48 min, and B_i is 62% (B). The sixth peak is Puerarin and is taken as the reference peak.

The evaluation of optimization criteria

The values of HCRF, CRF, and the r^* of experiments are listed in Table I. The column of CRF shows that the 4th experiment has a maximum value of 2.204. Obviously, the evaluation of the separation quality is not reasonable because only 13 peaks were separated in that experiment. This problem may be caused by the fact, as mentioned previously, that well-resolved peaks determine the function value and cover the effects of the poorly resolved peaks, including the overlapping peaks. Although the values of HCRF and the r^* are very different, their evaluation results are similar. Due to the fact that many subjective coefficients of different levels are imposed, the surface of HCRF function is not smooth and easily leads to local optima during the optimization process. From these results, it can be seen that the r^* is a reasonable criterion for evaluating the separation quality of fingerprint chromatograms.

Table IV. Experimental and Predicted Retention Times for 14 Components at the Optimal Point and Relative Deviation Between Them

	The 8th experiment			The 9th experiment		
Peak no.	Exp. (min)	Pre. (min)	E%rel (%)	Exp. (min)	Pre. (min)	E%rel (%)
1	3.808	3.919	2.93	3.723	3.765	1.14
2	5.109	5.303	3.80	4.811	4.903	1.91
3	6.507	6.555	0.73	6.069	6.125	0.93
4	8.160	8.401	2.95	7.829	7.882	0.67
5	10.208	10.200	0.07	10.048	9.982	0.66
6	11.488	11.415	0.63	11.179	11.202	0.21
7	13.173	13.003	1.29	12.949	12.961	0.09
8	14.901	14.794	0.72	14.816	14.877	0.41
9	16.213	16.142	0.44	16.288	16.367	0.49
10	17.045	16.939	0.62	17.237	17.310	0.43
11	19.979	19.925	0.27	20.459	20.579	0.59
12	20.459	20.419	0.20	21.003	21.121	0.56
13	32.898	32.916	0.06	35.189	35.350	0.46
14	38.304	38.311	0.02	41.493	41.539	0.11
E% (max)			2.93			1.91
E% (ave)			1.05			0.62

The search of optimization parameters

In the light of the previously mentioned grid search procedure, the seeking steps of B_0, B_i, and t_G were set to 1%, 1%, and 1 min, respectively. The first optimal point was obtained as B_0, 24%; B_i, 65%; and t_G, 45 min. Figure 2A shows a chromatogram run at the searched optimal point (the 8th experiment). Its corresponding r^* is 0.7919, higher than that of the seven initial experiments (see Table I).

Though the first optimal point was superior to other experiments, the separation between the 11th peak and the 12th peak were not satisfactory, and an iterative option needed to be carried out. Thus, the retention data of the 8th experiment were added to the previous retention data set of the seven experiments and a regression analysis was performed with the new data set. The modified coefficients and new statistical results for the 14 components are listed in Table II. These results show that the new statistical parameters, including R, R2, the values of F-test, and P for most of the peaks are slightly superior to the previous values. A grid search was carried out again with the modified regression model. A new optimal point was found as shown in Figure 3, with the initial organic solvent concentration B_0 25%, the final organic solvent concentration B_i 62%, and the gradient time t_G 48 min. Figure 3 describes the relationship between the global chromatographic function r^* and the three gradient parameters. It can be seen that the surface of function is very smooth, thus the optimal point is easy to be found. The chromatogram run at the new optimal point (taken as the 9th experiment) is shown in Figure 2B.

From Table I, it can be seen that the r^* value at the 9th experiment is 0.8037, which is higher than that at the 8th experiment. Because none of requirements conforms to end the optimization procedure, the iterative option needs to be further performed. The 10th experiment was then carried out and the same optimal point as that of the 9th experiment was obtained. Thus, one of requirements to end optimization procedure was fulfilled, the optimization procedure was then stopped, and the corresponding conditions of the 10th experiment were taken as the final optimal parameters.

In Table I, the sequence numbers from 8 to 10 are the optimization experiments carried out by the grid search. The experimental values and predictive values of the retention time for the first two optimal points (the 8th and 9th experiments) are compared and shown in Table IV. It can be seen that the maximal deviation and average deviation of the retention time of the 8th experiment are much higher than that of the 9th experiment, indicating that the second optimal point is superior to the first one; and the higher the accuracy of the model prediction, the closer the searched optimization condition is to the actual optimal condition. Figure 2 shows that the baseline separation between the 11th and the 12th peak was obtained by the iterative optimization procedure.

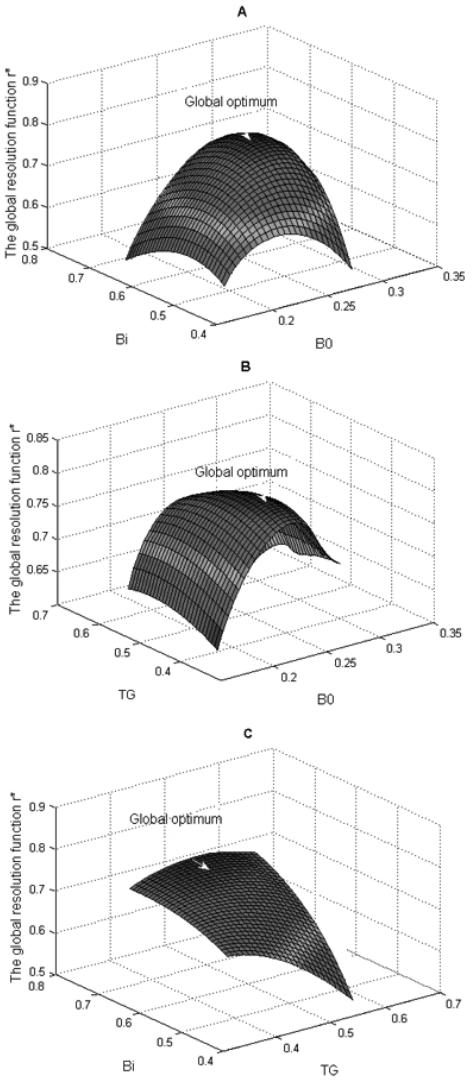

Figure 3. Response surfaces for the global resolution function r^* with three gradient parameters. The 3-D plot A, B, and C describe the relationship between r^* and T_G, B_0 for B_i of 62%, the relationship between r^* and B_0, B_i for T_G of 0.48, and the relationship between r^* and T_G, B_i for B_0 of 25%, respectively. The white arrows indicate the optimum point.

Conclusions

The proposed predictive model based on multiple regressions is simple and able to accurately predict the retention times of herbal medicines under the gradient conditions in the optimization regions. With the aid of the grid search algorithm, the optimal parameters can be found through an iterative optimization procedure. The optimal results are satisfactory, with an average deviation between the predicted and experimental values of only 0.62%. The results indicate that the four-step optimization strategy is efficient and feasible.

References

1. State Food Drug Administration of China. *Chin. Trad. Pat. Med.* **22:** 671–675 (2000).
2. World Health Organization. Guidelines for the Assessment of Herbal Medicines. Munich, 28 June 1991, WHO, Geneva, 1991.
3. G. Vivi-Truyols, J.R. Torres-Lapasio, and M.C. Garcia-Alvares-Coque. Enhanced calculation of optimal gradient programs in reversed-phase liquid chromatography. *J. Chromatogr. A* **1018:** 183–196 (2003).

4. N. Kuppithayanant, M. Rayanakorn, S. Wongpornchai, T. Prapamontol, and R.L. Deming. Enhanced sensitivity and selectivity in the detection of polycyclic aromatic hydrocarbons using sequential simplex optimization, the addition of an organic modifier and wavelength programming. *Talanta* **61:** 879–888 (2003).
5. C.L. Liu, P.L. Zhu, and M.C. Liu, Computer-aided development of a high-performance liquid chromatographic method for the determination of hydroxyanthraquinone derivatives in Chinese herb medicine rhubarb. *J. Chromatogr. A* **857:** 167–174 (1999).
6. J. Havel, M. Breadmore, M. Macka, and P.R. Haddad. Artificial neural networks for computer-aided modelling and optimisation in micellar electrokinetic chromatography. *J. Chromatogr. A* **850:** 345–353 (1999).
7. L.R. Snyder, J.W. Dolan, and J.R. Gant. Gradient elution in high-performance liquid chromatography: I. Theoretical basis for reversed-phase systems. *J. Chromatogr.* **165:** 3–30 (1979).
8. M.A. Stadalius, H.S. Gold, and L.R. Snyder. Optimization model for the gradient elution separation of peptide mixtures by reversed-phase high-performance liquid chromatography: Verification of retention relationships. *J. Chromatogr.* **296:** 31–59 (1984).
9. L.R. Snyder and J.W. Dolan. Initial experiments in high-performance liquid chromatographic method development I. Use of a starting gradient run. *J. Chromatogr. A* **721:** 3–14 (1996).
10. L.R. Snyder and J.W. Dolan. The linear-solvent-strength model of gradient elution. *Adv. Chromatogr.* **38:** 115–188 (1998).
11. Y. Mengerink, R. Peters, Sj. van der Wal, H.A. Claessens, and C.A. Cramers. Analysis of linear and cyclic oligomers in polyamide-6 without sample preparation by liquid chromatography using the sandwich injection method: III. Separation mechanism and gradient optimization. *J. Chromatogr. A* **949:** 307–326 (2002).
12. L. Szepsy. Effect of molecular interactions on retention and selectivity in reversed-phase liquid chromatography. *J. Chromatogr. A* **960:** 69–83 (2002).
13. A. Wang and P.W. Carr. Comparative study of the linear solvation energy relationship, linear solvent strength theory, and typical-conditions model for retention prediction in reversed-phase liquid chromatography. *J. Chromatogr. A* **965:** 3–23 (2002).
14. T. Baczek and R. Kaliszan. Predictive approaches to gradient retention based on analyte structural descriptors from calculation chemistry. *J. Chromatogr. A* **987:** 29–37 (2003).
15. T. Baczek and R. Kaliszan. Combination of linear solvent strength model and quantitative structure-retention relationships as a comprehensive procedure of approximate prediction of retention in gradient liquid chromatography. *J. Chromatogr. A* **962:** 41–55(2002).
16. T.L. Chester. Business-objective-directed, constraint-based multivariate optimization of high-performance liquid chromatography operational parameters. *J. Chromatogr. A* **1016:** 181–193 (2003).
17. M.L. Cotton and G.R.B. Down. Reversed-phase high-performance liquid chromatography of sulindac and related compounds using a computer simulation. *J. Chromatogr. A* **259:** 17–36 (1983).
18. Z. Guo, Q. Jin, G. Fan, Y. Duan, C. Qin, and M. Wen. Microwave-assisted extraction of effective constituents from a Chinese herbal medicine *Radix puerariae*. *Anal. Chim. Acta* **436:** 41–47 (2001).
19. K.T. Fang. Experiment design by uniform distribution. *Acta Math. Appl. Sin.* **3:** 363–372 (1980).
20. Y. Wang and K.T. Fang. A note on uniform distribution and experimental design. *Ke Xue TongBao* (*Sci. Bull. China*) **26:** 485–489 (1981).
21. P.C. Lu and H.X. Huang. An intelligent search method for HPLC optimization. *J. Chromatogr. Sci.* **27:** 690–697 (1989).
22. J.C. Berridge. Unattended optimization of reversed-phase high-performance liquid chromatographic separations using the modified simplex algorithm. *J. Chromatogr.* **244:** 1–14 (1982).
23. P.F. Vanbel, B.L. Tilquin, and P.J. Schoenmakers. Criteria for optimizing the separation of target analytes in complex chromatogram. *Chemom. Intell. Lab. Syst.* **35:** 67–86 (1996).
24. A.C.J.H. Drouen, H.A.H. Billiet, P.J. Schoenmakers, and L. de. Galan. An improved optimization procedure for the selection of mixed mobile phase in reversed phase liquid chromatography. *Chromatographia* **16:** 48–52 (1982).
25. J.W. Dolan, D.C. Lommen, and L.R. Snyder. Drylab computer simulation for high-performance liquid chromatographic method development: II. Gradient Elution. *J. Chromatogr. A* **485:** 91–112 (1989).
26. http://www.math.hkbu.edu.hk/UniformDesign/
27. L. Stokes and J. Plummer. Using spreadsheet solvers in sample design. *Comput. Stat. Data An.* **44:** 527–546 (2004).
28. M.A. Quarry, R.L. Grob, and L.R. Snyder. Measurement and use of retention data from high-performance gradient elution: Correction for "non-ideal" processes originating within the column. *J. Chromatogr. A* **285:** 19–51(1984).

Manuscript received October 20, 2006;
Revision received March 11, 2007.