Calculated Log K_{ow} as a Guide for Key-Set Mobile Phase Selection in Retention Prediction

C. H. Lochmüller* and **Min Hui**
Department of Chemistry, Duke University, Durham, NC 27708

Abstract

Previous reports have demonstrated that the relative solvent strength relationship is transferable between similar columns (e.g., C_{18} packing). Method transfer between columns is realized using a "key-set" solvent projection matrix derived from measurements on one column and measurements for compounds in a key combination mobile phase set on each new column. This paper offers a solution to the problem facing the method developer in selecting the best mobile phases as a "key set" to predict the behavior of solutes in other mobile phases of similar composition. Given the known relation of log K_{ow} (or log P) and ln k' to the free-energy change on transfer, one can estimate the k' value range for new compounds in each previously studied mobile phase. The key-set mobile phases are selected using calculated log K_{ow} as the three to five best-ranked solvents with k' predicted to be between 0.5 and 10. Retention data are collected for 19 new compounds in 20 water–methanol–acetonitrile mobile phases on two different C_{18} columns. The average cross-column prediction error based on k' for a three-factor model is 6%.

Introduction

There have been many attempts to devise a reliable, broadly applicable, and rugged method for the prediction of solute retention behavior in high-performance liquid chromatography (HPLC). Some methods attempt to model the behavior as related to physical properties of the solute, the mobile phase, or the stationary phase. Some of the better models are completely (or at least partially) empirical in origin and have names such as, "solvent strength," and, "linear solvent strength" (1–3). There is little question that the models proposed by Snyder (1) have had a real influence on the direction of thought in this area and that the commercialization of methods based on the known relation of isocratic and gradient-elution retention volumes provides useful tools for the method developer and the analyst facing adaptation of an existing method. In some of the compendial method publications, there has been a change from mandating a specific commercial column to leaving the choice of "a column of like type," (e.g., C_4, C_8, or C_{18}) to the analyst (4). The requirement is only that, for example, a drug substance, be separated from the common impurities and interfering peaks by a specified amount.

In new method development and in compendial method adaptation, the analyst is left with at least two questions: (*a*) "Will this column ever yield the minimum resolution required in a reasonable total elution time?" and (*b*) "What region(s) of the ternary solvent, critical resolution map are the best in which to explore possible method adaptation?" (4–6). If such questions could be answered with only a few well-directed measurements, it would be a time- and cost-saving advantage. Defining such regions need not require prediction accuracy on the order of precision of HPLC retention measurement because once such regions are located, other tools can be used to further refine the prediction. In the compendial example, the goal is to reach *at least* a certain resolution and order of detectability. Greater resolution than the compendial standard's minimum is acceptable from the compendial view, and the analyst must choose.

A prediction method that defined regions of mobile phase composition with acceptable resolution could also aid in determining how steep the rate of resolution growth or degradation is at the location of any given region. Clearly a "flat" region in which the critical resolution is not a strong function of mobile phase composition is the most promising in terms of method stability. What might be desired then is a method for predicting critical resolution maps (6) with enough precision to direct method development for a given problem. A modeling approach shown to be successful in studies of other complex chemical questions (e.g., molecular modeling) is a product of eigenanalysis (7). It is essentially a "soft-model approach," free of exact terms for physicochemical "effects." It is called factor analysis (7), and although it is not a "no-model" approach as is sometimes claimed, it is nearly so.

The application of the mathematical analysis methods called "chemometrics" to HPLC retention was first demonstrated when Lochmüller, Breiner, Reese, and Koel began to examine factor analytical modeling as a possible approach to understanding and even predicting retention behavior (5,6). In the period of 1985 to the present, Lochmüller and co-workers have demonstrated that, in the case of reversed-phase LC, the retention behaviors of a wide variety of neutral and charged species are "factor-analyzable" with respect to mobile phase composition using water, methanol, and acetonitrile or tetrahy-

* Author to whom correspondence should be addressed.

drofuran mixtures. An important consequence is that, using target factor methods, one can predict the behavior of new solutes in many mobile phases from a few measurements of retention in "key" solvent mixtures. It has been shown that, for columns of the same bonded-phase type, prediction using a three-factor model can be made with a 2–5% error and that it is possible to predict from one manufacturer's column type to another and with reversed-phase column types as well (4).

Lochmüller et al. proposed a factor analytical model for retention prediction that suggested that three factors were sufficient to explain the variation in the retention behavior of neutral compounds in binary and ternary solvent systems. In a demonstration of "worst-case" prediction demand (5), prediction in more than 30 solvents was performed using new retention measurements in only the three solvents that appeared to best span "solvent space" in the factor analysis. The fact that all subsequent measurements confirm the three-factor nature of the retention behavior does not suggest that prediction does not improve when more than three different mobile phase retention measurements are made; it does. There is no escaping the fact that if one wants to completely reproduce a data structure such as a matrix of reversed-phase LC measurements, the number of needed measurements is $n - 1$ where n is the total number of measurements. Perfect reproduction means perfectly reproducing not just real retention but white noise associated with the measurement as well.

The transfer of a method between columns may be realized by using a "key-set" solvent projection matrix derived from measurements on one column and measurements for compounds in a "key combination solvent set" on each new column (4,5). One of the challenges is to select the three or more "best" solvents to be used as a key set to predict the behavior of solutes in other solvents of similar chemical composition. Factor analysis reveals those solvents that best span solvent space (best characterize the mobile phase effect on retention) and makes it possible to rank some (e.g., 40) mobile phase compositions that have already been studied. Ordinarily, if one wants to predict the behavior of new solutes on the column from which the data was derived and wishes to make five measurements, the choice of mobile phases is the first five in the ranked list of 40. The situation can be more complicated if one changes columns (another C_{18} column but more or less phase mass or greater or lower surface coverage). The goal is to direct the analyst to a region of, say, ternary solvent space where adequate resolution in the shortest assay time can be obtained. A brute-force attempt at use of the highest ranked solvents can lead to capacity factors (k') greater than 100 in some cases. What the analyst needs to do is use a chemically sound selection of the mobile phases to be used for prediction purposes. One must also keep in mind that errors in prediction are in proportion to the number of measurements available and the correspondence between the composition of the basis-set mobile phase compositions and those used in a new method effort. This latter constraint is very important. Requesting that a gradient mixer produce a water–methanol–acetonitrile mixture in the ratio of 50:12.5:37.5 is no guarantee that one will get a mobile phase of that composition. Not all chromatographic pump–mixing–blending systems produce exactly the same mixture of solvents requested, even if they are internally very consistent and reproducible, and not all are truly linear over all ranges of volume percentage.

How can one get estimated but useful k' values for known structure compounds in each of the ranked key solvents? This paper describes an attempt based on the long-held assumption that some relation should exist between the familiar octanol–water partition coefficient (P or K_{ow}) and reversed-phase LC retention. The specific goal was to determine if calculation of K_{ow} or $\log K_{ow}$ would permit one to guess at the retention volume of a new solute based on the behavior of compounds already studied and their calculated values. Simple correlation of $\log P$ and $\ln k'$ gives a scattered diagram, not a simple linear trend. If one creates subsets of solutes based on chemical class, quite reasonable linear relations are obtained, and it is possible to interpolate using calculated $\log K_{ow}$ to estimate a new $\ln k'$.

It is reported here that one can make measurements in which three to five of the ranked "best" solvents were selected based on $\log P$ and predicted $\ln k'$. The goal was to keep measured retention between 0.5 and 10. The test set consists of two commercial but different C_{18} reversed-phase columns and 19 new compounds previously unstudied in the development of the factor analytical prediction "library" used. Retention was first predicted from the selected best solvents and then measured in a total of 20 mobile phases. This paper presents simulated four-component mixture critical resolution maps for comparison. The maps were based on either predicted data or the experimentally measured actual results. The agreement of the predicted and observed resolution maps was remarkably good.

Experimental

Chromatographic measurements

The retention behavior of 19 randomly selected new compounds was measured over the water–methanol–acetonitrile system. The experiments were carried out using a Perkin-Elmer (Norwalk, CT) series 4 LC pump, a PE ISS-100 autosampler, a PE LC-235 detector, and a PE LCI-100 computing integrator. All retention measurements were collected at 25.0°C at a wavelength of 255 nm. The chromatography was performed on two different commercial 3-cm C_{18} reversed-phase columns. Void volumes were determined by the elution time of ammonium nitrate in each mobile phase used.

Computational process

All calculations were performed using MATLAB (Mathworks, Natick, MA). The MATLAB routine "svd( )" was used to decompose the data matrices, and locally written routines were used to make all predictions. The solvent projection matrix (P) is the matrix that contains the solvent projections of all solvents in the data onto the key solvents. The procedure for obtaining the solvent projection matrix was as follows:

Singular value decomposition

The data, an *rxc* matrix in which the row and column cofactors are tested as indices for retention properties of the compounds and solvents, respectively, are the first singular values decomposed into three matrices, U, S, and V^T. The solvent projection matrix is as follows:

$$D = USV^T \quad \text{Eq 1}$$

where U is an *rxr* orthonormal matrix containing the eigenvectors that span the compound retention space, V is a *cxc* orthonormal matrix containing the eigenvectors that span the mobile phase retention space, and S is an *rxc* diagonal matrix that contains the singular values. The singular values are the square roots of the eigenvalues of the matrix D^TD, and they are a measure of the amount of variance associated with the corresponding eigenvectors. Three factors (4,6) were found to be significant in the retention data; the other factors appeared to be associated with noise. Therefore, Equation 1 could be reduced to:

$$\bar{D} = \bar{U}\bar{S}\bar{V}^T \quad \text{Eq 2}$$

where $\bar{D}$ is the reproduced data matrix using only three factors, $\bar{U}$ contains the first three columns of matrix U, $\bar{S}$ contains the first three rows and columns of matrix S, and $\bar{V}$ contains the first three columns of matrix V.

Combination test and target transformation factor analysis

A combination test (7) was used to find the key combination set. A matrix $D^{\neq}$ containing three column vectors (solvents) of the data matrix served as a target test matrix, and then the eigenvectors of the data matrix were transformed into new axes by using the transformation matrix T, so that they were best aligned with the target test vectors. The transformation matrix was obtained from:

$$T = \lambda^{-1} (\bar{U}\bar{S})^T D^{\neq} \quad \text{Eq 3}$$

where λ contains the three largest eigenvalues of the matrix D^TD. The solvent projection matrix P is the matrix that contains the solvent projections of all the solvents in the data onto the key solvents. The solvent projection matrix P associated with $D^{\neq}$ was obtained by:

$$P = T^{-1}\bar{V}^T \quad \text{Eq 4}$$

and the data was reproduced using:

$$\tilde{D} = D^{\neq}P \quad \text{Eq 5}$$

Once the solvent projection matrix associated with the key combination set was obtained, it served as a solvent projection library. Only three retention measurements were required for each compound to predict the retention behavior in all mobile phase compositions in the library when changing columns. To make the retention predictions in the library mobile phases, a matrix ($\ddot{D}$) containing the three required retention measurements for each new solute was generated. The predictions were made using:

$$\hat{D} = \ddot{D}P \quad \text{Eq 6}$$

where $\hat{D}$ is the predicted data matrix, $\ddot{D}$ is the measured retention data matrix, and P is the solvent projection matrix obtained from the library.

By definition, the combination set that reproduces the data

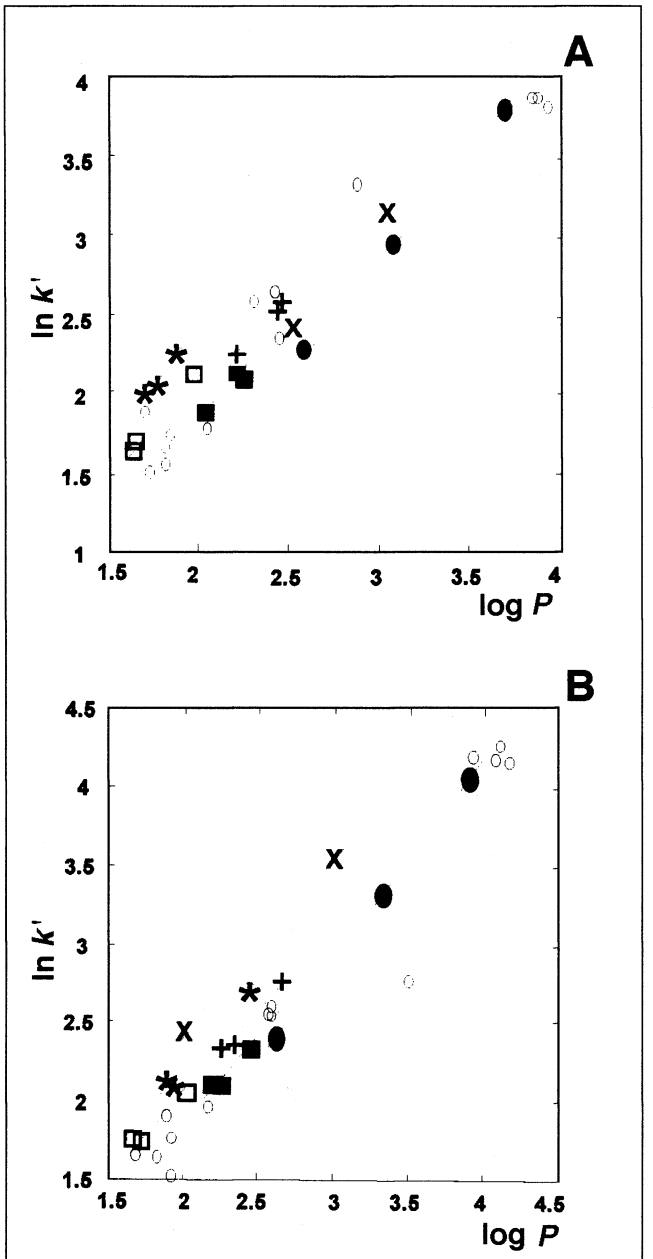

Figure 1. (A) Ln k' versus log P with the 60:30:10 water–methanol–acetonitrile mobile phase. (B) Ln k' versus log P with the 60:40:00 water–methanol–tetrahydrofuran mobile phase. *o*-Fluoronitrobenzene, *m*-fluoronitrobenzene, and *p*-fluronitrobenzene (■); sec-phenylethylalcohol, 2-phenylethylalcohol, and 3-phenyl-propanol (□); dimethylphthalate and diethylphthalate (X); benzene, toluene, and ethylbenzene (●); nitrobenzene, *m*-dinitrobenzene, and *p*-dinitrobenzene (✱); and 2,4-dinitrotoluene, 2,6-dinitrotoluene, and 3,4-dintrotoluene (+).

with minimum error is selected as the key combination set, and the solvents in that set are called the "key" solvents. When changing to a new column, it is possible that the old best solvents give k' values that are practically too large or so small as to produce unexpected errors in prediction arising from errors in measurement. A more convenient approach to prediction might be to select the key solvents so that all results fall in the range $0.5 < k' < 10$.

Log P *calculations*

The partition coefficient can be written as:

$$P = C_s / C_w \qquad \text{Eq 7}$$

where C_s and C_w are the equilibrium concentrations of the solute in the organic and aqueous phases, respectively (8). The common two-phase system used is octanol–water. A solute with a high P value is regarded as lipophilic, and a solute with a low P value is regarded as hydrophilic. Because the P scale usually covers a range of more than 10^n, logarithmic P values are preferred. This transforms Equation 7 into:

$$\log P = \log C_s - \log C_w \qquad \text{Eq 8}$$

R. F. Rekker et al. (8) used the following formula to calculate log P:

$$\log P = \Sigma a_n f_n + c \qquad \text{Eq 9}$$

where f represents the hydrophobic fragmental constant (the lipophilicity constant of a part of the structure to the total lipophilicity), a is a numerical factor indicating the number of a given fragment in the structure, and c is the constant term (intercept), which should be zero, but in fact is not, depending on the groups attached to the adjacent carbon atoms.

The fragmental constants have been shown to be suitable for building up the total lipophilicity of a structure with surprising accuracy. Many f values have been calculated by statistical analysis (9) as well. These same f values can be used to calculate an estimate of log P for all the organic compounds in the retention time library. In practice, two methods were used to obtain the log P values: (*a*) using Equation 9 as described by Rekker and the hydrophobic fragmental constants obtained from the published lists, and (*b*) using the commercial program "ACD/LogP" (10). In the latter approach, the chemical structure for a known compound was entered in a "molecule editor window," and the log P value was then calculated.

Though it is relatively simple to get the estimated log P values by these two methods, the first method does not differentiate the log P values of meta-substituted benzene compounds from the para-substituted isomers. The ACD/LogP program resolves this and produces different log P values for the ortho-, meta- and para- benzene compounds.

Results and Discussion

Log P (or log K_{ow}) and HPLC retention

In comparing the capacity factor (k') in HPLC with ($1/R_F - 1$), a free energy-based constant in paper and thin-layer chromatography (11), R. F. Rekker (8) assumed that the following equation applies in HPLC:

$$\log P = \log K + \log k' \qquad \text{Eq 10}$$

This implies that the capacity factor (k') is linearly related with the octanol–water partition coefficient log P, where K is a constant for the HPLC system.

To test the above equation, work was carried out to calculate the octanol–water partition coefficients (log P) and to correlate them with the capacity factors (k') of those compounds present in our retention library for water–methanol–acetonitrile and water–methanol–tetrahydrofuran. Linear-model regression analysis (12) was performed for log P against ln k' for 31 organic compounds. If a linear relationship exists, the retention behavior of a new solute can be estimated from its calculated log P value and the retention behavior of compounds already studied. Then key solvents for the prediction set would be selected according to the estimated values of k' ($0.5 < k < 10$) for all the new compounds.

Based on the known root mean square (RMS) measurement error values (two biggest, two median, two smallest), six solvent systems were chosen from the water–methanol–acetonitrile and water–methanol–tetrahydrofuran retention

Table I. Linear Regression Analysis of Log P versus Ln k' of Homologues in Water–Methanol (v/v)

Homologues		20:80	25:75	30:70	35:65	40:60	45:55	50:50	55:45	60:40
Acetones	m	0.303	0.352	0.401	0.457	0.510	0.565	0.626	0.684	0.754
	r^2	0.994	0.994	0.993	0.993	0.994	0.993	0.993	0.993	0.993
Alcohols	m	0.295	0.332	0.389	0.436	0.483	0.537	0.593	0.650	0.734
	r^2	0.959	0.943	0.944	0.952	0.946	0.947	0.947	0.948	0.946
Benzenes	m	0.313	0.359	0.405	0.455	0.504	0.554	0.608	0.667	0.720
	r^2	0.998	0.998	0.998	0.999	0.999	0.999	0.999	0.999	0.999
Ethers	m	0.325	0.369	0.426	0.447	0.524	0.562	0.623	0.680	0.733
	r^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Phenols	m	0.296	0.332	0.398	0.430	0.490	0.530	0.594	0.656	0.722
	r^2	0.986	0.986	0.986	0.986	0.986	0.986	0.986	0.985	0.985
Phenones	m	0.319	0.357	0.408	0.452	0.535	0.570	0.626	0.671	0.756
	r^2	0.993	0.993	0.993	0.993	0.993	0.992	0.992	0.992	0.992

Table II. Data Sets Used in This Study

Data set	Column Type	Phase	Mobile phase
1	Whatman ODS-3 (Whatman, Clifton, NJ)	C_{18} bulk	unbuffered
2	Perkin-Elmer 3 x 3 cartridge	C_{18} brush	unbuffered
3	Perkin-Elmer 3 x 3 cartridge	C_{18} brush	unbuffered

Table III. Compounds and Mobile Phases Used in Each Data Set

No.	Compound	Data set	No.	Mobile phase (water–methanol–acetonitrile)	Data set
1	acetophenone	1	1	70:30:00	1
2	anisole	1	2	60:40:00	1
3	benzaldehyde	1	3	50:50:00	1, 2, 3
4	benzene	1	4	40:60:00	1, 2, 3
5	benzonitrile	1	5	30:70:00	1, 2, 3
6	benzophenone	1	6	20:80:00	1, 2, 3
7	*p*-chlorobenzaldehyde	1	7	10:90:00	1
8	chlorobenzene	1	8	00:100:00	1
9	*p*-chlorotoluene	1	9	70:22.5:7.5	1
10	*o*-dichlorobenzene	1	10	60:30:10	1, 2, 3
11	diethyl phthalate	1	11	50:37.5:12.5	1, 2, 3
12	dimethyl phthalate	1	12	40:45:15	1, 2, 3
13	*m*-dinitrobenzene	1	13	30:52.5:17.5	1, 2, 3
14	*p*-dinitrobenzene	1	14	20:60:20	1
15	2,4-dinitrotoluene	1	15	10:67.5:22.5	1
16	2,6-dinitrotoluene	1	16	0:75:25	1
17	3,4-dinitrotoluene	1	17	50:25:25	1, 2, 3
18	ethylbenzene	1	18	40:30:30	1, 2, 3
19	*m*-fluoronitrobenzene	1	19	30:35:35	1, 2, 3
20	*o*-fluoronitrobenzene	1	20	20:40:40	1, 2, 3
21	*p*-fluoronitrobenzene	1	21	10:45:45	1
22	*p*-methoxybenzaldehyde	1	22	0:50:50	1
23	methyl benzoate	1	23	50:12.5:37.5	1, 2, 3
24	naphthalene	1	24	40:15:45	1, 2, 3
25	*p*-nitroacetophenone	1	25	30:17.5:52.5	1, 2, 3
26	*p*-nitrobenzaldehyde	1	26	70:00:30	1
27	nitrobenzene	1	27	60:00:40	1, 2, 3
28	*sec*-phenylethyl alcohol	1	28	50:00:50	1, 2, 3
29	2-phenylethyl alcohol	1	29	40:00:60	1, 2, 3
30	3-phenyl-1-propanol	1	30	30:00:70	1, 2, 3
31	toluene	1	31	20:00:80	1, 2, 3
32	1,3-dinitronaphthalene	2, 3	32	10:00:90	1
33	*p*-anilinophenol	2, 3	33	00:00:100	1
34	azobenzene	2, 3			
35	diethyl phenylmalonate	2, 3			
36	2′,5′-dimethoxy-acetophenone	2, 3			
37	benzil	2, 3			
38	benzyl chloride	2, 3			
39	*o*-anisidine	2, 3			
40	1-chloro-3-nitrobenzene	2, 3			
41	*m*-nitrobenzoic acid	2, 3			
42	2-bromo-4-phenyl phenol	2, 3			
43	2-amino-4-nitrophenol	2, 3			
44	*m*-chloroaniline	2, 3			
45	2,4-dinitroaniline	2, 3			
46	4-chloro-2-nitrophenol	2, 3			
47	2-naphthol	2, 3			
48	*p*-chlorophenol	2, 3			
49	*p*-*n*-butoxyphenol	2, 3			
50	2-anilinoethanol	2, 3			

value libraries, respectively, to predict ln k' using the estimated log P values. Plots based on the smallest RMS errors are shown in Figure 1. Upon cursory examination, one might conclude that there is no clear, linear-model correlation between ln k' and log P at all. Closer inspection suggests that Rekker's assumption may still be reasonable if chemical compound class or related chemical isomers are considered separately. In the plots of Figure 1, data for compounds with similar structures are given the same symbol and are connected by a dotted line. Compounds of the same "type" fall on one line, but compounds with different chemical classes or isomeric relation fall on other straight lines that are often almost parallel.

Linear regression analysis was also performed between the ln k' and log P for the homologous compounds in the water–methanol solvent system in the library. Results for the different kinds of compounds in the same solvent conditions were compared, and the results are shown in Table I. The results indicate that, for all the solvent conditions, linear-model regression produced correlation coefficients (r^2) greater than 0.94. Also, the slopes of these correlation lines increased as the water content increased. Homologues occupy their own lines, distinct from non-homologues. The parallel slope trend persisted, for example, in the 20:80 (water–methanol) mobile phase, in which the slopes found for six kinds of organic compounds were all in the range 0.310 with different intercepts.

Based on this conclusion, the authors calculated the log P values for all the compounds studied in the existing library and sorted all the retention data based on log P values. The likely k' values in a given solvent for other, new solutes were then estimated using the known k' values of their similarly structured compounds. A new log P–retention library was constructed.

Retention prediction

Three data sets (see Tables II and III) were used in this study. Data set 1 was collected by Lochmüller et al. (6) and served as the "library data set" to produce the solvent projection matrix. Data sets 2 and 3 were collected for the first time using the same compounds but two different C_{18} columns, both different than the column used for the library data set.

For each of the 19 arbitrarily selected new compounds, the authors estimated the likely k' range in all the library

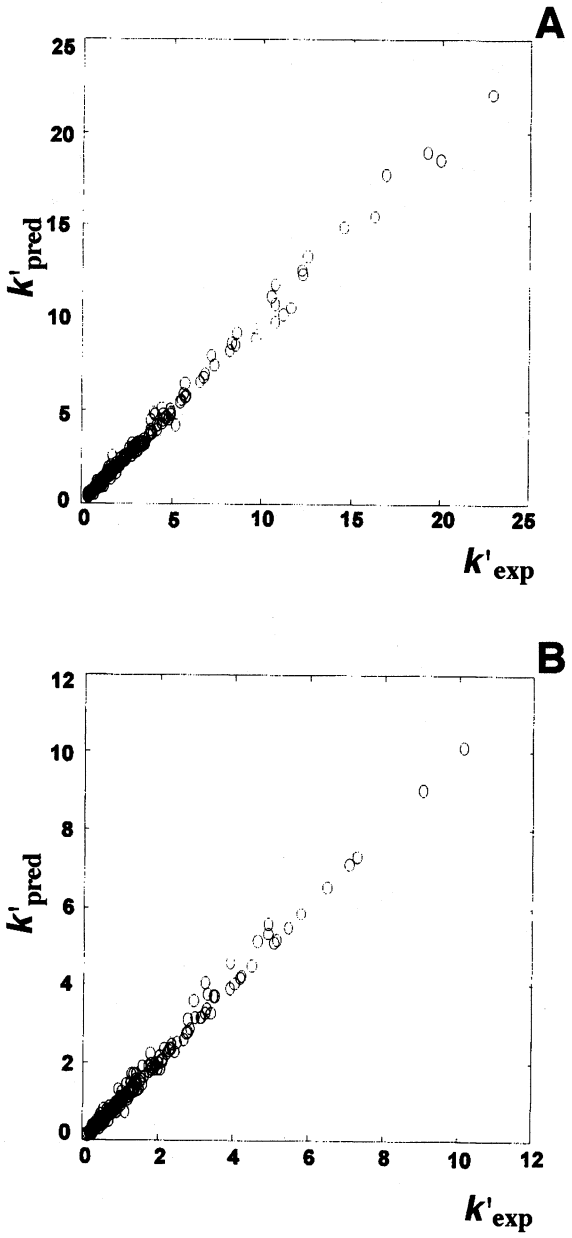

Figure 2. Predicted capacity factors (k'_{pred}) versus experimental capacity factors (k'_{exp}) based on the three key solvents 40:45:15, 50:12.5:37.5, and 50:25:25 (water–methanol–acetonitrile). Prediction was achieved by using the solvent projections obtained from the library data set. (A) PE-3 × 3 CR column, 216 data points. (B) PEB-6 column, 225 data points.

mobile phases by comparing their log *P* values with the log *P* values and *k'* values of a similar compound already in the library retention data set as: $P_1/P_2 = k_1'/k_2'$. It was found that for all these 19 new compounds, only six mobile phases should give $0.5 < k' < 10$. These mobile phases for water–methanol–acetonitrile were 50:12.5:37.5, 50:25:25, 40:30:30, 50:00:50, 40:45:15, and 40:60:00.

A combination test was performed on the these six solvents in the library data set (data set 1). The combination set of 40:45:15, 50:12.5:37.5, and 50:25:25 gave the best results (4.19% reproduction error of capacity factors) and was therefore selected as the key combination set for a three-factor analytical model. The combination set of 40:30:30, 40:45:15, 40:60:00, and 50:25:25 gave the best results (6.45% reproduction error of capacity factors), whereas 40:30:30, 40:45:15, 40:60:00, 50:00:50, and 50:12.5:37.5 gave the best results (7.02% reproduction error of capacity factors) as the key members were increased.

A solvent projection matrix associated with the three-, four-, and five-member key combination set was obtained using Equation 4. The retention measurements of each compound in data sets 2 and 3 in the key solvents were stored in matrix $\ddot{D}$. $\ddot{d}_{ij}$, an element of $\ddot{D}$, is the natural logarithm of the capacity factor of compound *i* in the key solvent *j*. The retention prediction of data sets 2 and 3 was done using Equation 6. For the two different PE C_{18} reversed-phased columns, the retention prediction results for a three-factor analytical model are shown in Figure 2. The average prediction errors for the two different columns and three different factor analytical models are shown in Table IV. From Table IV, it can be seen that a three-factor analytical model with three key mobile phase measurements gave the lowest prediction error for data sets 2 and 3. Compared with the overall 4.19% reproduction error of data set 1 itself using the log *P* selected solvents, overall cross-column prediction errors of 6.13 and 6.72% for data sets 2 and 3 are reasonable.

Critical resolution maps (critical resolution as a function of solvent composition) were constructed for both columns. In each case, the authors selected four new, previously unstudied compounds: 1,3-dinitronaphthalene, *p*-anilinophenol, diethyl phenylmalonate, and 2'5'-dimethoxy-acetophenone. The horizontal axis represents the water content, and the vertical axis represents the methanol content (% v/v). The amount of acetonitrile in the mobile phase is the difference between 100% and the sum of the other two components. Resolution surfaces were generated for simulated mixtures of the four studied compounds. Surfaces were generated for both the predicted and the experimental resolutions, and an error surface ($R_{experimental} - R_{predicted}$) was generated for comparison purposes. These maps show graphically the results of these calculations in the form of contour plots. Visual comparison of the experimental (Figures 3A and 4A) and predicted surfaces (Figures 3B and 4B) showed the very good agreement between the general shapes of the surfaces. Although the predicted surface showed a higher absolute

Table IV. Average Prediction Error for Two New Retention Data Sets

Key combination set (water–methanol–acetonitrile)	Reproduction error (%) Library data set	Prediction error (%) Data set 2	Data set 3
40:45:15, 50:12.5:37.5, 50:25:25	4.19	6.13	6.72
40:30:30, 40:45:15, 40:60:00, 50:25:25	6.45	11.15	8.31
40:30:30, 40:45:15, 40:60:00, 50:00:50, 50:12.5:37.5	7.02	12.40	10.37

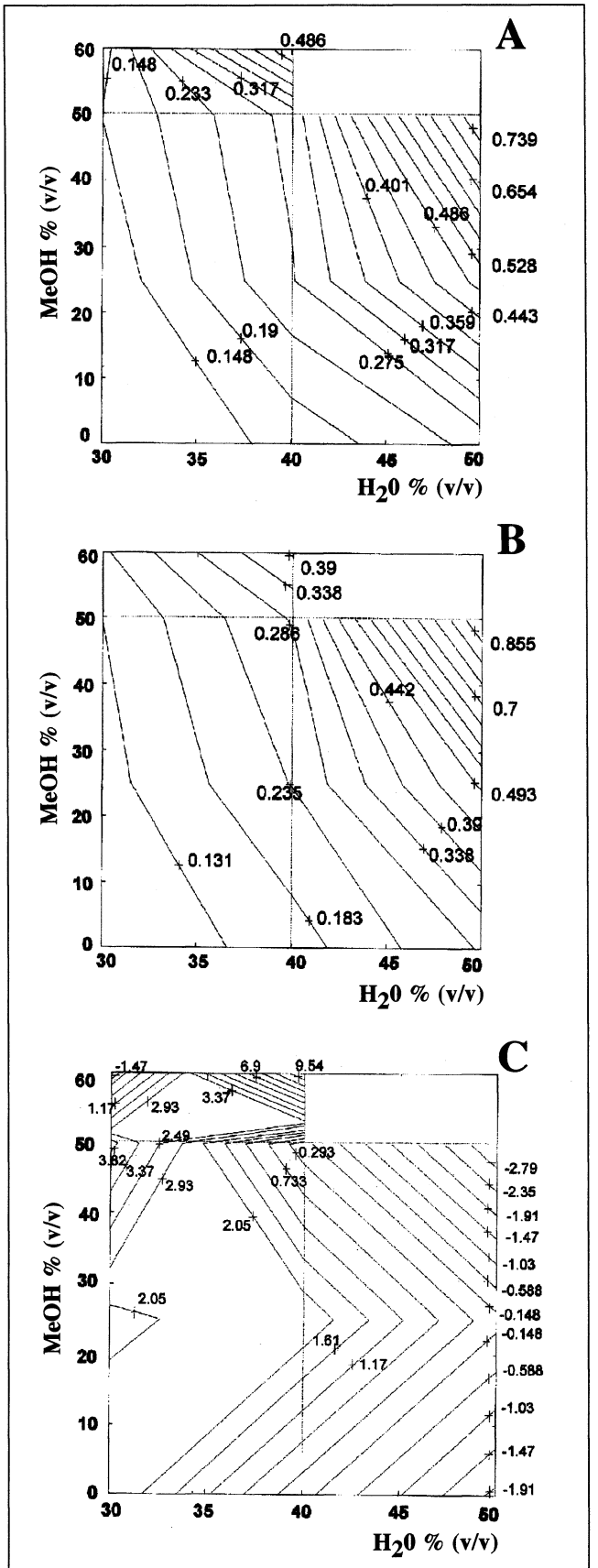

Figure 3. Contour maps for a simulated four-component mixture showing resolution based on (A) experimental capacity factors and (B) predicted capacity factors (PE-3 x 3 CR column). (C) Error surface for predicted versus observed resolution obtained as the difference between Figures 3A and 3B and expressed in percent error (PE-3 x 3 CR column).

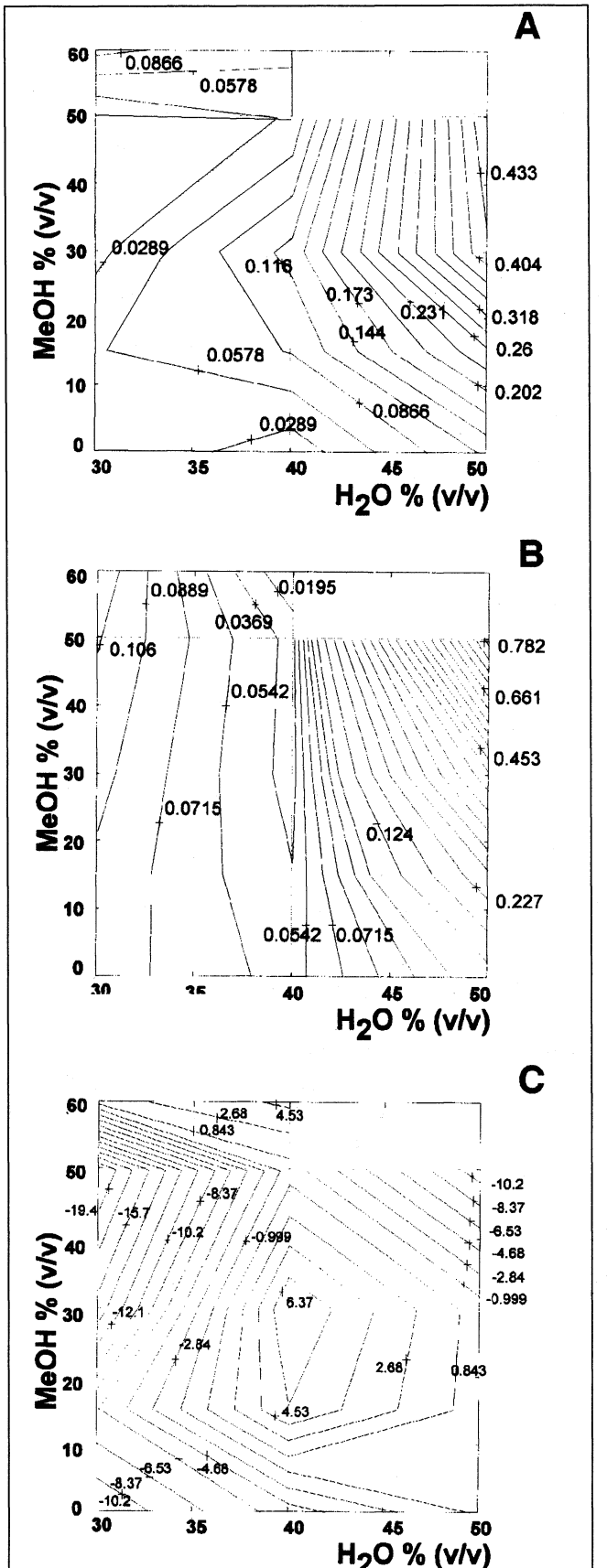

Figure 4. Contour maps for a simulated four-component mixture showing resolution based on (A) experimental capacity factors and (B) predicted capacity factors (PEB-6 column). (C) Error surface for predicted versus observed resolution obtained as the difference between Figures 4A and 4B and expressed in percent error (PEB-6 column).

resolution near the maximum than the experimental surface, the location of the optimum, in terms of solvent axes, was virtually identical with that of the experimental result. In each case with these two different columns, the global optimum area was near 50:50:00 (water–methanol–acetonitrile). Through the analysis of Figures 3 and 4 (the contour maps based on experimental and predicted data), it was found that the simulated four-compound mixture showed the highest resolution in 50:50:00 but relatively smaller error surface, expressed as the difference between the A and B sections of each figure (% error). Furthermore, under different mobile phase conditions, one can sort the capacity factor value for the mixture according to its decreasing or increasing order by comparing the A and B sections of each figure with the C section. Clearly the prediction would be improved as more measurements were made at library mobile phase compositions in the region of the global optimum.

Conclusion

The preliminary study of the relation between the octanol–water partition coefficient and the capacity factor has shown that for a given solvent, the magnitude of log P is a good indicator of the k' value range. It is therefore possible to select the key mobile phases for new measurement so as to have reasonable retention using a calculated estimate of log P. It is also clear that this is true for classes of compounds within their classes but not as true for interclass comparisons. The solvent projection matrix associated with three key solvents is still useful in cross-column retention prediction, even if one selects the best solvents from the ranked list based on the likely range of k' values estimated by their partition coefficients. The prediction would be better, of course, if the same column type was used for both library and new compound separations. Cross-column prediction is a more demanding test of the ruggedness of the method but seems to give quite good estimates of the location of the global resolution maximum. Further work is underway to extend this selection method for key solvents to other column or solvent systems with the inclusion of more complex and varied solute classes such as amino acids and peptides.

References

1. L.R. Snyder and M.A. Quarry. Computer simulation in HPLC method development: Reducing the error of predicted retention times. *J. Liq. Chromatogr.* **10:** 1789–1820 (1987).
2. P.J. Schoenmakers, H.A.H. Billiet, and L.D. Galan. Systematic study of ternary solvent behavior in reversed-phase liquid chromatography. *J. Chromatogr.* **218:** 261–84 (1981).
3. J.W. Weyland, C.H.P. Bruins, and D.A. Doornbos. Use of three-dimensional minimum α-plots for optimization of mobile phase composition for RP-HPLC separation of sulfonamides. *J. Chromatogr. Sci.* **22:** 31–39 (1984).
4. C.H. Lochmüller, C.E. Reese, and S.H. Hsu. Cross-column retention prediction in reversed-phase liquid chromatography using factor analytical modeling. *Anal. Chem.* **66:** 3806–13 (1994).
5. C.H. Lochmüller, S.J. Breiner, C.E. Reese, and M.N. Koel. Characterization and prediction of retention behavior in reversed-phase chromatography using factor analytical modeling. *Anal. Chem.* **61:** 367–75 (1989).
6. C.H. Lochmüller, M.A. Hamzavi-Abedi, and C.-X. Ou. Retention anomaly in ternary solvent systems in reversed-phase liquid chromatography. *J. Chromatogr.* **387:** 105–20 (1987).
7. E.R. Malinowski and D.G. Howery. *Factor Analysis in Chemistry*, 2nd ed. John Wiley & Sons, New York, NY, 1980, pp. 59–100.
8. R.F. Rekker and W.T. Nauta. *The Hydrophobic Fragmental Constant*, Vol. 1. Elsevier Scientific, New York, NY, 1977, pp. 5–55.
9. C. Hansch and S.M. Anderson. The effect of intramolecular hydrophobic bonding on partition coefficient. *J. Org. Chem.* **32:** 2583–86 (1967).
10. ACD/Labs Programs, Advanced Chemistry Development, Inc., 1995.
11. G.D. Christian. *Analytical Chemistry*, 5th ed. John Wiley & Sons, New York, NY, 1994, pp. 552–55.
12. M.R. Spiegel. *Theory and Problems of Statistics.* Schaum Publishing, New York, NY, 1961, pp. 241–68.

Manuscript accepted August 18, 1997.