Journal of Chromatographic Science, Vol. 34, February 1996

Prediction of the Retention Behavior of Ionizable Compounds in Reversed-Phase LC Using Factor-Analytical Modeling

C.H. Lochmüller*, Su-Hsiu Hsu, and **Charles Reese**
Department of Chemistry, Duke University, Durham, North Carolina 27708

Abstract

Principal component factor analysis (PCFA) and target transformation factor analysis (TTFA) are used to examine the reversed-phase high-performance liquid chromatographic retention behavior of some neutral, basic, and acidic organic compounds in water–methanol–acetonitrile solvent systems at pH 7, 5, and 3. The ratios of the pairs of HA/A^- (acid) or B/BH^+ (base) are controlled by the buffered mobile phases, and this permits the factor analytical model to be extended to ionic compounds. Four factors are sufficient to reproduce the data within an error margin of 3%.

Introduction

Accurate prediction of retention behavior in high-performance liquid chromatography (HPLC) would allow the rapid development of optimal HPLC separation methods; this is a major reason for the interest in predicting retention behavior in reversed-phase HPLC. Differential migration of solutes is the basis of separation in reversed-phase HPLC; however, many of the fundamental variables that determine this differential migration are not well understood. This is the principal reason for the current relative lack of success in making reliable retention predictions. A factor analytical model for predicting retention that does not require knowledge of the mechanism of HPLC retention was developed (1); it was reported that three factors were required to predict the reversed-phase HPLC retention behavior of neutral compounds in ternary solvent systems. The practical consequence of this approach is a dramatic reduction in the number of mobile phases that must be used to completely characterize the retention behavior of new, previously unstudied solutes. In a worst case demonstration, the paper shows that the study of retention in three "key" mobile phases permitted the prediction of retention in 35+ other combinations of water–methanol–acetonitrile with an RMS error of approximately 5%. Precision improved dramatically with the addition of only a few more mobile phases and reached measurement precision when five mobile phases were studied.

* Author to whom correspondence should be addressed.

It is important that the factor analytical methods described here be understood to reduce rather than increase the amount of work needed to reproduce results obtained on a column of the same manufacture or even between columns of different types. As previously shown, the method can permit intercolumn transfer of methods with a few measurements on the new column or batches of the same nominal column materials (2).

General trends in retention change can be predicted for ionizable species as the mobile phase buffer pH changes and ionization increases or decreases. The goal of this work was to extend the previously reported success in the prediction of the reversed-phase liquid chromatographic retention of neutral species to the prediction of the behavior of ionizable species. The prediction of reversed-phase LC retention using factor-analytical methods is reported for solutes that are ions in a pH range from 3 to 7 units. The fundamental question involved the continued factor analyzability of the retention behavior and the broadening of the power of prediction. For that reason, the number of careful retention measurements made is in excess of those necessary to make predictions on systems previously studied or known to be factor analyzable (2).

The retention behavior of ionizable compounds at each pH level was investigated using principal component factor analysis (PCFA) and target transformation factor analysis (TTFA). PCFA is a least-squares method used for eigenanalysis of a data matrix. The first eigenvector is extracted from the data matrix to remove as much of the total variance (the sum of the magnitudes of the projections of all data points on the eigenvector) as possible. The second eigenvector is orthogonal to the first one and is computed such that the maximum of the remaining variation is extracted. This procedure is repeated until all the variation in the data matrix is described in terms of the extracted eigenvectors. Each eigenvector is associated with an eigenvalue. The larger the eigenvalue, the more important the eigenvector is. If there are no experimental errors in the data, PCFA produces the exact number of eigenvectors (n) that define the data space. When real data are examined by PCFA, a larger number of eigenvectors, m (the smaller of the number of the rows or columns in the data matrix), is produced because of experimental errors. One main task of PCFA is to determine how many factors are significant

and how many factors are to be considered as noise. Once the number of significant factors has been determined, the data space can be described using the smaller dimensions. The result of PCFA at this stage is an abstract solution that describes retention behavior using a minimum of variables without any reference to a physical mechanism.

TTFA is a least-squares method of finding a 'valid' vector giving the best fit to a target vector hypothesized to be a real factor; in this context 'valid' refers to any vector that is located in the space defined by the significant eigenvectors. Real vectors, that is, vectors defined by the values of the real functions which are responsible for retention, are valid vectors, as are the actual columns of the data itself. The data columns are often called 'typical' factors. All valid vectors are linear combinations of the significant eigenvectors. If there is a valid vector reasonably similar to the target vector, then the target vector may be a real factor, otherwise it is not. The objective of TTFA is to present the data in terms of physically meaningful vectors instead of abstract eigenvectors. However, finding a complete set of physically meaningful vectors that span the data space is not a simple task.

Even if the complete set of real factors was unknown, predictions can still be made in a abstract manner by using the data columns. The results of target testing on the data columns (typical factors) should be always successful because they always lie in the data space if the effects of noise on the data are neglected. If ***n*** factors are required to describe the data space, then the data points can be located by an appropriate set of ***n*** typical factors which span the data space. Some sets of typical factors may span the data space better than others because not all of the columns may require n dimensions, that is, not all of the real factors (mechanisms) controlling retention may be required to explain retention in some of the mobile phases or for some of the compounds. A combination test can be used to find the best ***n*** typical factors to define the data space. The remaining data columns can be expressed in terms of the data columns chosen as the representative typical factors. For example, the xth data column factor, $\mathbf{D_x}$, can be expressed by a linear equation:

$$\mathbf{D_x} = \mathbf{P_1 D_1} + \mathbf{P_2 D_2} + \cdots + \mathbf{P_n D_n} \qquad \text{Eq 1}$$

where $\mathbf{D_1}$, $\mathbf{D_2}$, ... $\mathbf{D_n}$ are the chosen representative typical factors, and $\mathbf{P_1}$, $\mathbf{P_2}$, ... $\mathbf{P_n}$ are projections of $\mathbf{D_x}$ onto these representative typical factors. Therefore, prediction can be made from Equation 1 without knowing any of the physically meaningful factors that control retention.

Table I. Compounds Used at pH 7, pH 5, and pH 3

No.	Compound	Mobile phase pH 7	pH 5	pH 3
1	Acetopheonone	X	X	X
2	Benzene	X	X	X
3	Benzonitrile	X	X	X
4	*o*-Dichlorobenzene	X	X	X
5	*p*-Dinitrobenzene	X	X	X
6	Methylbenzoate	X	X	X
7	2-Phenylethyl alcohol	X	X	X
8	Toluene	X	X	X
9	Caffeine	X	X	X
10	4-Chloroaniline	X	X	X
11	*N,N*-Dimethylaniline	X	X	X
12	2,6-Dimethylaniline	X	X	X
13	3,4-Dimethylaniline	X	X	X
14	Pyridine	X	X	X
15	4-*tert*-Butylpyridine	X	X	X
16	2-Aminopyridine	X	X	X
17	Quinoline	X	X	X
18	3-Aminoquinoline	X	X	X
19	Benzoic acid			X
20	*p*-Fluorobenzoic acid			X
21	4-Aminobenzoic acid			X
22	2-Chlorobenzoic acid			X
23	4-Chlorobenzoic acid			X
24	2,5-Dimethylbenzoic acid			X
25	3,4-Dimethylbenzoic acid			X
26	Salicylic acid			X

Theoretical

There are two main processes controlling the retention of acids and bases in a reversed-phase LC system: the extent of the ionization of acids and bases and the distribution of the sample compounds between the mobile phase and the stationary phase.

Under suitable conditions of pH, the acids (HA), and bases (B) are in equilibrium with their ions:

$$HA \rightleftharpoons H^+ + A^-$$
$$B + H^+ \rightleftharpoons BH^+$$

The extent of the ionization of acids and bases during a chromatographic analysis can be controlled by varying the pH of the mobile phase. Increasing the pH increases the ionization of the acids and decreases the ionization of the bases. In reversed-phase LC, an increase in the ionization of the sample decreases the retention of the sample because the ionic form is more hydrophilic than the molecular form. If a buffered mobile phase is used in a reversed-phase LC system, the ratios of the pairs of HA/A^- or B/BH^+ should be controlled by the constant pH of the mobile phase, and this should permit the factor analytical model to be extended to ionic compounds.

In the factor analytical model, it is assumed that the natural logarithm of the capacity factor ($\ln k'$) can be expressed as a linear sum of product terms of n functions of solvent composition and n functions of solute structure as shown in Equation 2:

$$\ln k' = \sum_{i=1}^{n} f_i^c f_i^s \qquad \text{Eq 2}$$

where f^c is a function of solute structure, f^s is a function of the solvent composition, and

n is the number of significant factors. This is consistent with most of the retention theories that have been proposed in reversed-phase HPLC.

The experimental retention data can be represented in matrix notation as **D**, where d_{ij}, an element of **D**, is the natural logarithm of the capacity factor of compound i in the mobile phase j. The data matrix **D** may be factored, using singular-value decomposition, into three matrices, **U**, **S**, and $\mathbf{V}^T$ (Equation 3):

Table II. Mobile Phases Used at pH 7, pH 5, and pH 3

	% Volume		
Solvent no.	Water	Methanol	Acetonitrile
1	60	40	0
2	60	30	10
3	60	0	40
4	50	50	0
5	50	25	25
6	50	12.5	37.5
7	50	37.5	12.5
8	50	0	50
9	40	60	0
10	40	45	15
11	40	30	30
12	40	15	45
13	30	70	0
14	30	52.5	17.5
15	30	35	35
16	30	17.5	52.5
17	20	80	0
18	20	60	20
19	20	40	40

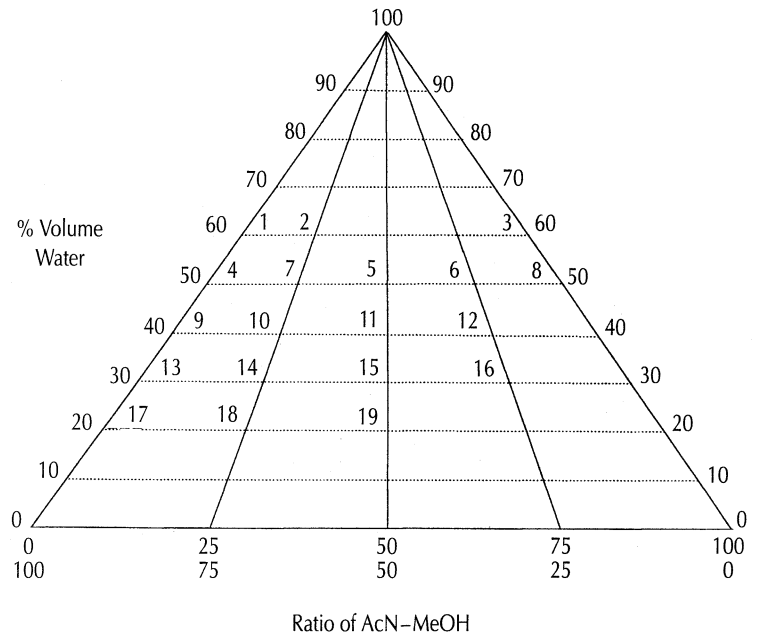

Figure 1. The solvent composition map. The vertical scale shows the percent volume of water and the horizontal scale shows the ratios of acetonitrile and methanol in the mobile phase. The numbers shown in the map are the solvent numbers listed in Table II.

$$\mathbf{D} = \mathbf{USV}^T \qquad \text{Eq 3}$$

where **U** is an $r \times r$ orthogonal matrix containing the eigenvectors that define the compound retention space, **V** is a $c \times c$ orthogonal matrix containing the eigenvectors that span the mobile phase retention space, and **S** is an $r \times c$ diagonal matrix that contains the singular values. The singular values are the square roots of the eigenvalues of the matrix $\mathbf{D}^T\mathbf{D}$.

The data matrix may be reproduced using $n \leq m$ factors (the smaller number of r and c). When the correct value of n is employed, the agreement between the data matrix and the reproduced data matrix should be within experimental error. This procedure compresses the data space from m dimensions to n dimensions. When the correct value of n is determined, Equation 4 may be used to reproduce the data:

$$\bar{\mathbf{D}} = \bar{\mathbf{U}}\bar{\mathbf{S}}\bar{\mathbf{V}}^T \qquad \text{Eq 4}$$

where $\bar{\mathbf{D}}$ is the reproduced data matrix using n factors only, $\bar{\mathbf{U}}$ contains the first n columns of matrix **U**, $\bar{\mathbf{S}}$ contains the first n rows and columns of matrix **S**, and $\bar{\mathbf{V}}^T$ contains the first n rows of matrix $\mathbf{V}^T$ where the superscript T denotes the transposed matrix. After the correct number of significant factors has been determined, the significant row eigenvectors, $\bar{\mathbf{U}}$, may be used to perform target testing on the test vectors.

The target transformation may be used to find new axes that are aligned with fundamental parameters of the mobile phases (or compounds). The data points can be designated by the coordinates on any set of n axes that span the data space. The newly transformed axis $\mathbf{X}_j$, which is called the predicted vector, can be obtained from Equation 5:

$$\hat{\mathbf{X}}_j = \bar{\mathbf{U}}\bar{\mathbf{U}}^T\mathbf{X}_j \qquad \text{Eq 5}$$

where $\mathbf{X}_j$ is a test vector that is suspected to be a real factor. If the agreement between $\mathbf{X}_j$ and $\hat{\mathbf{X}}_j$ is within the experimental error, then this test vector is tentatively accepted as a real factor.

Columns of the experimental data may be used as test vectors. A matrix, $\mathbf{D}^\ddagger$, which contains n typical factors (columns) of the data matrix, may serve as a target test matrix. Although each typical factor lies in the data space and obviously will yield a successful test, it is not guaranteed that all combinations of n typical factors will successfully reproduce the data as only certain combinations may span the total factor space. Therefore, a combination test is performed on the data matrix to identify the set of typical factors that best reproduces the data matrix. This combination set, **D**, is called the key combination set, and the solvents (columns) contained in the key combination set are called the key solvents. This is the approach we have adapted to achieve the retention behavior of the solvents used in this study.

The transformation matrix associated with the key solvents can be obtained using Equation 6:

$$\mathbf{T} = \lambda^{-1}(\bar{\mathbf{U}}\bar{\mathbf{S}})^{\mathrm{T}}\ddot{\mathbf{D}} \qquad \text{Eq 6}$$

where λ contains the n largest eigenvalues of the matrix $\mathbf{D}^{\mathrm{T}}\mathbf{D}$. The derivation of Equation 6, which was used to get the transformation matrix **T**, was described by Malinowski and Howery (3).

A projection matrix, **P**, containing the projections of the other solvents on the key solvents can be obtained using Equation 7:

$$\mathbf{P} = \mathbf{T}^{-1}\bar{\mathbf{V}}^{\mathrm{T}} \qquad \text{Eq 7}$$

Then, the data matrix can be reproduced using Equation 8:

$$\tilde{\mathbf{D}} = \ddot{\mathbf{D}}\mathbf{P} \qquad \text{Eq 8}$$

where $\tilde{\mathbf{D}}$ is the reproduced data. Once the solvent projections are available, only n reversed-phase HPLC retention measurements are needed for each compound to predict the retention behavior in all the solvents.

Table III. Results of Principal Component Factor Analysis

No. of factors	Eigenvalue	% of Variance explained	Reduced eigenvalues	Indicator function ($\times 10^2$)	Probability test
pH 7 Data excluding the acid compounds					
1	4.90e+02	89.86	1.43e+00	0.127	0.000
2	5.20e+01	9.53	1.70e–01	0.036	0.000
3	3.02e+00	0.55	1.11e–02	0.012	0.000
4	1.39e–01	0.03	5.78e–04	0.011	0.057
5	9.08e–02	0.02	4.33e–04	0.008	0.024
6	2.50e–02	0.00	1.38e–04	0.007	0.116
7	1.32e–02	0.00	8.49e–05	0.007	0.163
8	1.01e–02	0.00	7.63e–05	0.006	0.120
9	4.61e–03	0.00	4.19e–05	0.006	0.194
10	2.80e–03	0.00	3.11e–05	0.006	0.220
pH 5 Data excluding the acid compounds					
1	4.69e+002	89.07	1.37e+00	0.130	0.000
2	5.35e+001	10.15	1.75e–01	0.040	0.000
3	3.68e+000	0.70	1.35e–02	0.015	0.000
4	2.36e–001	0.04	9.85e–04	0.012	0.029
5	1.20e–O01	0.02	5.71e–04	0.008	0.015
6	3.87e–002	0.01	2.13e–04	0.006	0.035
7	1.66e–002	0.00	1.06e–04	0.005	0.037
8	4.68e–003	0.00	3.54e–05	0.005	0.142
9	3.36e–003	0.00	3.05e–05	0.004	0.089
10	1.02e–003	0.00	1.14e–05	0.004	0.248
pH 3 Data					
1	4.85e+02	77.90	9.82e–01	0.167	0.000
2	1.30e+02	20.95	2.90e–01	0.044	0.000
3	5.81e+00	0.93	1.42e–02	0.022	0.000
4	6.74e–01	0.11	1.83e–03	0.018	0.030
5	2.82e–01	0.05	8.54e–04	0.017	0.068
6	1.53e–01	0.02	5.22e–04	0.015	0.092
7	9.31e–02	0.01	3.58e–04	0.015	0.103
8	7.44e–02	0.01	3.27e–04	0.012	0.046
9	2.80e–02	0.00	1.41e–04	0.012	0.107
10	1.50e–02	0.00	8.82e–05	0.011	0.136

Experimental

Instrumentation

The retention data were collected using a Perkin-Elmer Series 4 LC pump (Norwalk, CT), a PE ISS-100 autosampler, and a PE LC-235 diode-array detector. A PE LCI-100 integrator and a Hewlett-Packard 3390A integrator (Wilmington, DE) were used to measure the retention times.

Columns

The data were collected using PE 3 × 3 C_{18} cartridge columns; the columns were especially prepared by the manufacturer in two batches; each batch was prepared from the same lot of packing material at the same time. Each column was characterized by running eight compounds in three mobile phases before use. The columns were sorted, and the most similar were selected to collect the retention data. The standard characteristics of the column in use were periodically examined, and the column was replaced when a change of more than 3% was observed.

Materials and Procedures

The solutes and mobile phases used at pH 7, 5, and 3 are listed in Tables I and II, respectively. The solvent compositions are also graphically shown in Figure 1. All retention measurements were collected at 25°C using a wavelength of 255 nm. In order to maintain constant column characteristics, a Whatman ODS-2 C_{18} precolumn was placed between the pump and the autosampler to presaturate the mobile phase with stationary phase. Samples were dissolved in the mobile phase for injection. The retention time of each compound in each mobile phase was measured at least three times, and the average standard deviation was always less than 1.5%. Mobile phases were prepared from commercial HPLC-grade solvents. The composition of the mobile phase was determined by dividing the mass of each solvent in the mobile phase by its density and using this value to compute the volume percentage of each solvent. It is customary to report the composition of mobile phases as the volume percent of each component and to use the volume before mixing.

A phosphate buffer was used to control the pH of the mobile phase. The pH of a mixed

Table IV. Key Combination Set and Reproduction Error Based on Four Factors at Each pH Level

	Key solvent		Key solute	
pH	% Volume*	Reproduction error (%)	Compound	Reproduction error (%)
7	60:40:00	1.6	*o*-Dichlorobenzene	1.8
	50:25:25		Caffeine	
	50:00:50		2,6-Dimethylaniline	
	30:52.5:17.5		Pyridine	
5	60:30:10	1.8	*o*-Dichlorobenzene	1.8
	50:50:00		*p*-Dinitrobenzene	
	50:00:50		Caffeine	
	30:52.5:17.5		3,4-Dimethylaniline	
3	60:40:00	3.0	Benzonitrile	2.7
	60:00:40		Caffeine	
	30:70:00		2,5-Dimethylbenzoic acid	
	20:40:40		Salicylic acid	

* Percent volume of water–methanol–acetonitrile.

Table V. Average Prediction Errors Associated with Each Compound

	% Average prediction error*					
	pH 7		pH 5		pH 3	
Compound	C	S	C	S	C	S
Acetopheonone	1.9	1.7	2.0	2.1	1.0	2.9
Benzene	2.7	2.9	2.2	2.5	1.5	4.4
Benzonitrile	2.3	1.7	3.0	2.9	k	2.9
o-Dichlorobenzene	k	2.0	k	0.8	3.0	3.6
p-Dinitrobenzene	2.8	1.3	k	1.2	4.0	4.4
Methylbenzoate	1.0	1.0	1.4	1.4	1.4	2.6
2-Phenylethyl alcohol	3.0	3.1	2.6	3.6	2.0	2.5
Toluene	2.8	2.8	2.2	2.5	1.4	2.6
Caffeine	k	2.8	k	2.7	k	3.3
4-Chloroaniline	1.6	2.1	1.5	1.6	2.9	3.2
N,N-Dimethylaniline	2.5	1.8	1.5	1.8	7.1	9.7
2,6-Dimethylaniline	k	0.9	0.7	0.8	2.3	2.1
3,4-Dimethylaniline	0.7	1.0	k	0.8	4.4	5.2
Pyridine	k	2.3	2.8	3.0	4.4	3.9
4-*tert*-Butylpyridine	3.4	2.5	1.2	2.7	6.1	7.3
2-Aminopyridine	1.9	2.6	6.3	4.6	3.5	4.6
Quinoline	2.6	1.6	2.3	2.4	3.1	3.4
3-Aminoquinoline	3.1	2.5	2.7	2.4	4.8	5.5
Benzoic acid					1.6	1.4
p-Fluorobenzoic acid					3.4	4.3
4-Aminobenzoic acid					6.7	6.7
2-Chlorobenzoic acid					2.8	4.4
4-Chlorobenzoic acid					1.2	1.4
2,5-Dimethylbenzoic acid					k	1.3
3,4-Dimethylbenzoic acid					0.5	1.5
Salicylic acid					k	4.6

* Symbols: C, data were predicted by using four key compounds; S, data were predicted by using four key solvents; k, key compounds.

aqueous-organic solvent was taken as the same as the pH of the aqueous fraction in this study. The total ionic strength of the buffered mobile phase was controlled at 5mM. Void volumes were determined as the elution volume of ammonium nitrate.

Computations

PCFA and TTFA were performed using MATLAB (Mathworks, Inc.; Natick, MA) script programs.

Data preparation

In order to generate the data matrix for factor analysis, the data were prepared as follows:

- The mean value of the retention time measurements was calculated.
- The mean retention time was converted to a capacity factor, k', using Equation 9:

$$k' = \frac{(t_R - t_0)F}{Ft_0} \qquad \text{Eq 9}$$

where t_R is the retention time of each compound in each solvent, t_0 is the void time of ammonium nitrate, and F is the flow rate of the mobile phase.

- The natural logarithm of the capacity factor was calculated.

The acidic compounds were excluded from both the pH 7 and the pH 5 data because they were essentially nonretained.

Results and Discussion

Column stability

There were two reasons to characterize the columns before using them. First, the characteristics of two C_{18} columns provided by the same manufacturer are not guaranteed to be exactly the same, and because of column degradation in conditions of extreme pH, a single column may not be sufficiently stable to collect all of the data at the accuracy level required in these experiments. Two batches of columns (five columns per batch) were characterized. Benzene, benzonitrile, dimethyl phthalate, 3,4-dinitrotoluene, *m*-fluoronitrobenzene, *o*-fluoronitrobenzene, sec-phenylethyl alcohol, and 3-phenyl-1-propanol were used as standard solutes. The characteristics of each column were

examined by running the eight standard solutes in three solvent systems: 70:30:00, 60:00:40, and 50:37.5:12.5 (percent volume water–methanol–acetonitrile). The results showed that the characteristics of the columns in the same batch were similar to each other but slightly different from the other batch. So the columns with the most similar characteristics were selected to collect the data. In order to assure reliable results, the characteristics of the column in use were periodically checked. If the retention behavior of the standard solutes in the three mobile phases changed significantly, then the column was replaced with a new column. The characteristics of one column changed significantly after running pH 2.3 buffered mobile phases; therefore, pH 3 was the most acidic system studied. One column had to be discarded when it was contaminated by buffered solvent in which microorganisms had started to grow.

Principal component factor analysis

The indicator function (3), reduced eigenvalue (4), Malinowski's probability test (5), and the error in data reproduction were applied to estimate the proper number of factors in this study. The reduced eigenvalue method assumes that the eigenvalues associated with random error factors should be statistically equal and should begin to level off when extra factors are employed. The indicator function, which is an empirical function, should reach a minimum value when the correct number of factors is reached. The probability test estimates the proper number of real factors by using an F-Test. In this study, a significance level of 5% was used to pick out the proper number of real factors; a probability of less than .05 indicated that the factor is a real factor. Table III shows the results of these methods to determine the minimum number of real factors required to reasonably reproduce the data. Unfortunately, there is no clear agreement about the factor numbers for all pHs (7, 5, and 3). The reduced eigenvalue indicated three factors were important for all pHs. The indicator function failed to predict the correct factor numbers for all pHs. The probability test suggested 5, 7, and 8 factors were important for pH 7, pH 5, and pH 3, respectively; however, Table IV shows the reproduction errors based on four typical factors for all pHs were within a reasonable error range. Therefore, four factors were used as the significant factors as they are sufficient to reasonably reproduce the data while keeping the prediction model relatively simple.

The discovery that four factors are needed to successfully predict the retention of ionized species makes sense chemically. Adding the potential to be ionized might be expected to add a factor to an abstract model that is used to describe the behavior of netural species and that requires only three factors. The result is both chemically and chemometrically satisfying.

Target transformation factor analysis and combination test

Target tests were performed on all combinations of four solvents. The combination set with the minimum root-mean-square reproduction error was selected as the key combination set. The solvents in this key combination set are called the key solvents. Then, the data matrix was transposed, and target tests were performed on all combinations of the four compounds. The key compounds were selected in a similar way. The key solvent and compound combination sets for pH 7, pH 5, and pH 3 are listed in Table IV. The procedure of reproducing the data based on the four key solvents was as follows. First, these four key solvents that served as the target matrix were used to obtain the solvent projection matrix, **P**, by using Equation 7. Then, the reproduction was achieved by using Equation 8. Also, the data were reproduced by using the key compounds with the same method. The reproduction errors associated with the key combination sets for pH 7, pH 5, and pH 3 are listed in Table IV. The mean reproduction errors shown in Table IV are the mean values of the errors obtained from Equation 10.

$$\text{Error } \% = 100 \times \frac{\text{abs}(k'_{\text{pred}} - k'_{\text{exp}})}{k'_{\text{exp}}} \quad \text{Eq 10}$$

The average prediction errors associated with each compound and solvent are listed in Tables V and VI, respectively. Also, the results of the reproduction obtained by using the key solvents are graphically shown in Figures 2–4. The predictions were made over capacity factor ranges from 0.3 to 87, 0.3 to 86, and 0.1 to 85 at pH 7, pH 5, and pH 3, respectively. The results of the reproduction based on the key compounds are not shown but were very similar to Figures 2–4.

Table VI. Average Prediction Errors Associated with Each Solvent

	% Average prediction error*					
	pH 7		pH 5		pH 3	
Solvent no.	C	S	C	S	C	S
1	1.9	k	4.2	3.7	4.0	k
2	2.4	1.7	1.6	k	2.5	4.5
3	2.6	1.8	3.5	4.2	3.8	k
4	1.4	2.0	0.9	k	3.9	3.8
5	1.0	k	1.5	1.7	3.1	4.5
6	2.1	1.2	2.9	2.2	3.7	4.0
7	1.8	2.1	2.1	1.5	4.0	4.5
8	2.2	k	1.9	k	4.0	4.4
9	2.7	2.6	3.5	2.3	2.6	3.6
10	2.3	1.6	2.2	1.4	3.2	4.2
11	1.7	1.2	0.8	0.8	2.2	3.9
12	3.7	3.2	2.0	1.4	3.0	4.2
13	2.9	2.4	2.7	2.0	2.6	k
14	1.0	k	1.1	k	3.2	3.2
15	2.3	1.5	1.3	1.5	2.4	3.1
16	2.1	1.9	3.5	2.6	3.1	3.9
17	4.5	3.7	3.1	2.4	4.4	4.1
18	2.3	1.2	2.0	2.2	1.9	1.7
19	2.8	2.1	3.1	3.4	2.2	k

* Symbols: C, data were predicted by using four key compounds; S, data were predicted by using four key solvents; k, key compounds.

Predictive strategy

The prediction can be extended to a new compound by using equation 11.

$$a_{ij} = \sum_{n=1}^{4} k_{in}\, p_{nj} \qquad \text{Eq 11}$$

where a_{ij} is the predicted natural logarithm of the capacity factor (ln k') of the new compound i in the mobile phase j, k_{in} is the ln k' of compound i in the key solvent n, and p_{nj} is the projection of solvent j on the key solvent n. Note that the solvent projections can be obtained from Equation 7 and are known. Therefore, only four retention measurements are required (the retention of the new compound in the four key solvents) to get the prediction in the other 15 solvents.

This retention prediction method can also be extended to a new solvent in a similar way. The only required measurements are the retention values of the key compounds in the new solvent, then the prediction of the other compounds of the data matrix can be obtained.

If a new column is used, then a new core matrix is required to construct the solvent/compound projections. The core matrix contains the retention of each key compound in each key solvent. It may be necessary to add a fifth compound to the core matrix to avoid near-singularity problems in singular value decomposition under some conditions. Therefore, the size of the core matrix may be increased to 5×4. The procedure for getting the projections of a new solvent x on each key solvent is first to obtain the retention measurements of the key com-

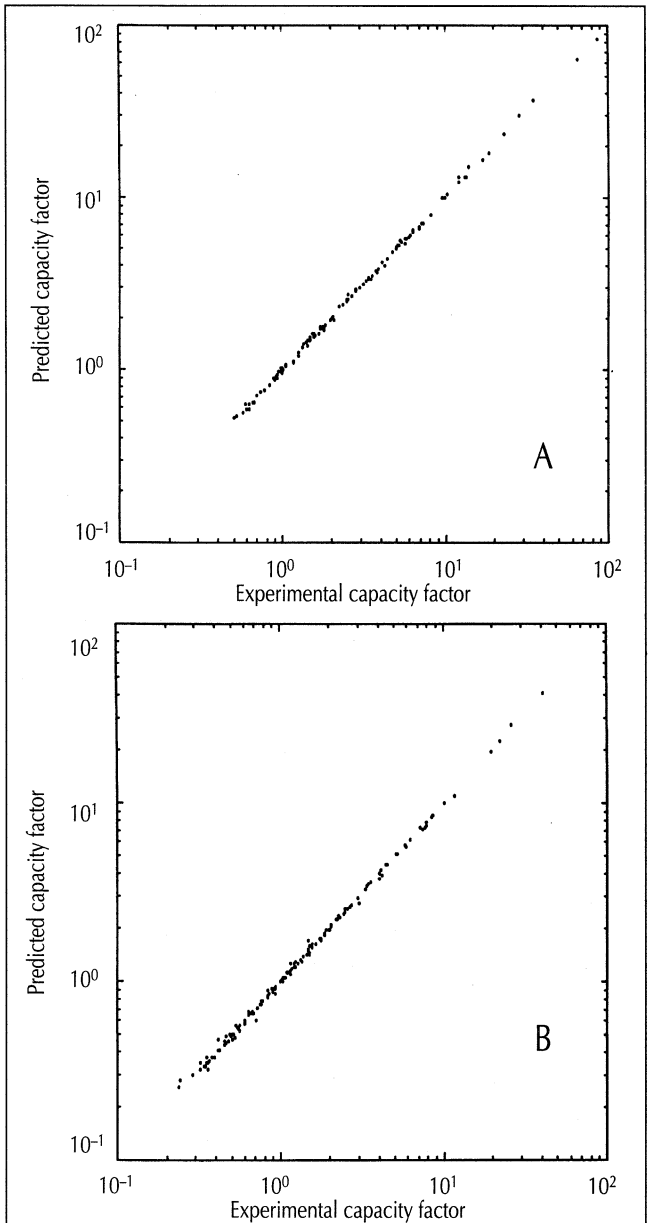

Figure 2. Predicted capacity factors of neutral (a) and basic (b) compounds at pH 7 based on the following percent volumes of key solvents (water–methanol–acetonitrile) versus the experimental capacity factors: 60:40:00, 50:25:25, 50:00:50, and 30:52.5:17.5.

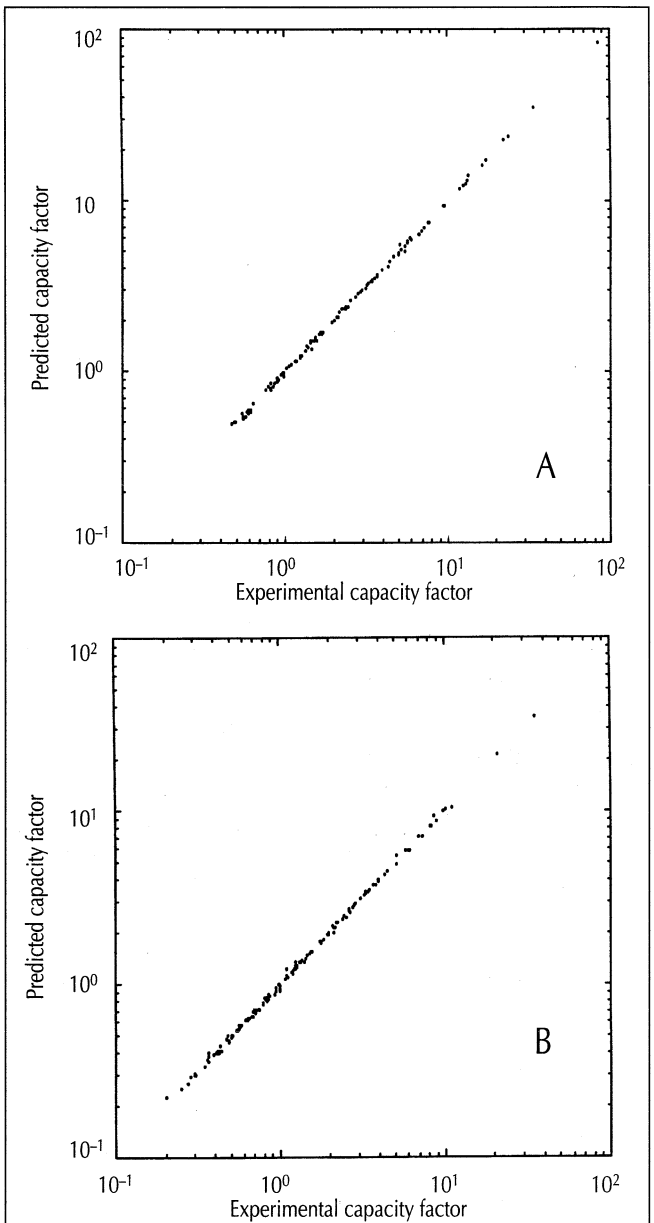

Figure 3. Predicted capacity factors of neutral (a) and basic (b) compounds at pH 5 based on the following percent volumes of key solvents (water–methanol–acetonitrile) versus the experimental capacity factors: 60:30:10, 50:50:00, 50:00:50, and 30:52.5:17.5.

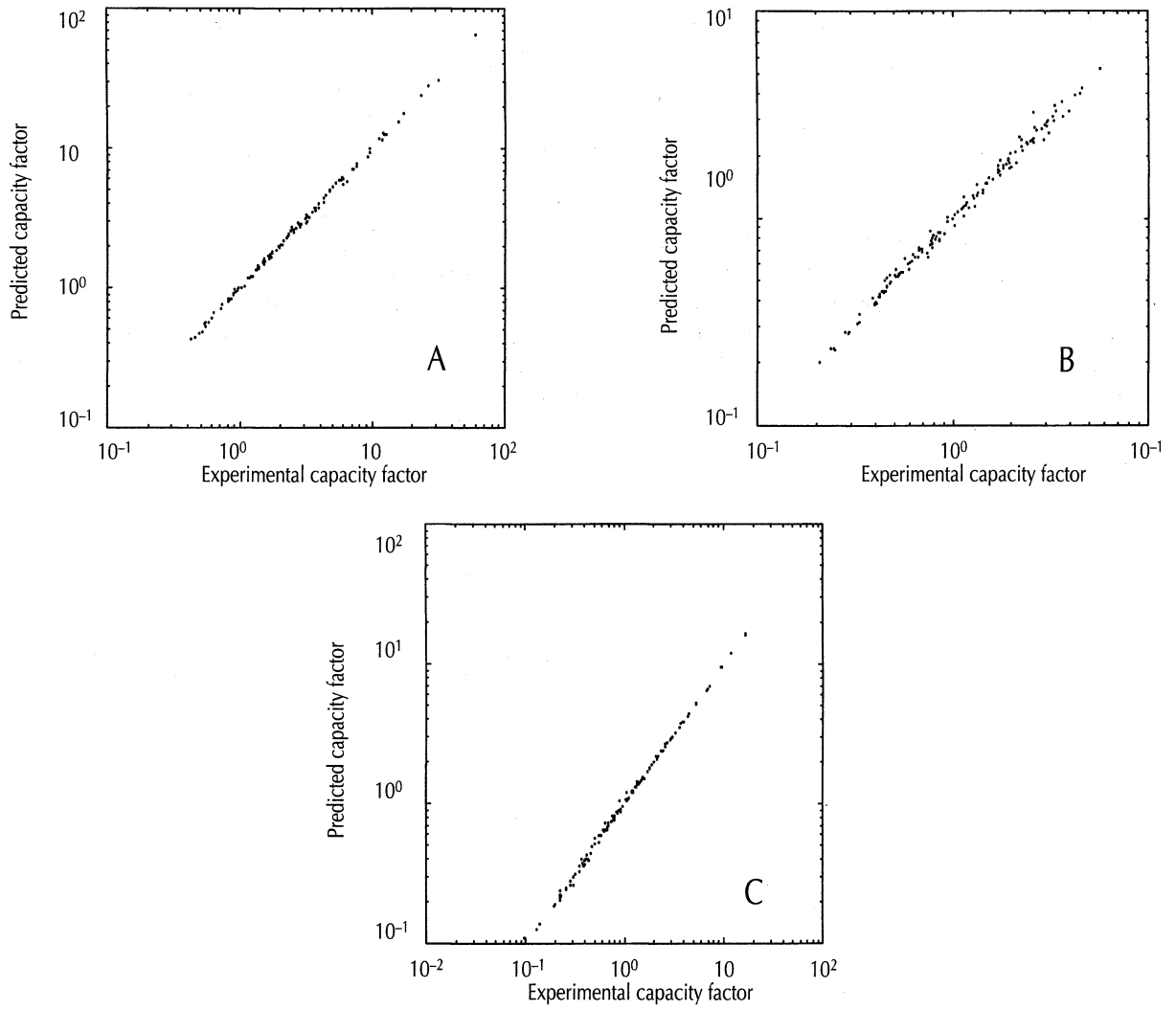

Figure 4. Predicted capacity factors of neutral (a), basic (b), and acidic (c) compounds at pH 3 based on the following percent volumes of key solvents (water–methanol–acetonitrile) versus the experimental capacity factors: 60:40:00, 60:00:40, 30:70:00, and 20:40:40.

pounds in the new solvent x and then append these measurements to the core matrix. By performing a target test on the appended core matrix with the core matrix serving as the target matrix, the projections of the solvent x on the key solvent can be obtained from Equation 7. The method to obtain the compound projections is similar to the method of getting the solvent projections; the retention value of the new compound in the key solvents is appended to the transposed core matrix, and the procedures described earlier are followed.

Conclusion

It has been shown that the factor analytical model can be extended to ionic compounds. The retention behavior of ionic compounds was accurately described with four factors. Controlling the pH with simple buffers permits this method to be extended to ionic compounds, validating the assumption that the change in solvent composition does not significantly affect the degree of ionization of the solutes.

References

1. 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)
2. C.H. Lochmüller, C.E. Reese, and S. Hsu. Cross-column retention prediction in reversed phase liquid chromatography using factor analytical modeling. *Anal. Chem.* 3806–13 (1994).
3. E.R. Malinowski and D.G. Howery. *Factor Analysis in Chemistry.* Wiley, New York, NY, 1980.
4. E.R. Malinowski. Theory of the distribution of error eigenvalues resulting from principal component analysis with applications to spectroscopic data. *J. Chemom.* **1:** 33–51 (1987).
5. E.R. Malinowski. Expert system for knowledge-based modeling of analytical laboratories as a tool for laboratory management. *J. Chemom.* **3:** 49–60 (1988).

Manuscript accepted June 9, 1995.