ELSEVIER

Journal of Chromatography A, 965 (2002) 3–23

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Comparative study of the linear solvation energy relationship, linear solvent strength theory, and typical-conditions model for retention prediction in reversed-phase liquid chromatography

Aosheng Wang, Peter W. Carr*

Department of Chemistry, Smith and Kolthoff Halls, University of Minnesota, 207 Pleasant St. S.E., Minneapolis, MN 55455-0431, USA

Abstract

This paper describes two new retention models for predicting retention under different reversed-phase liquid chromatography (RPLC) conditions. The first one is a *global* linear solvation energy relationship (LSER) that expresses retention as a function of both solute LSER descriptors and mobile phase composition. The second is a so-called ''typical-conditions model'' that expresses retention under a given chromatographic condition as a linear function of retention under different so-called ''typical'' conditions. The global LSER was derived by combining the *local* LSER model and the linear solvent strength theory (LSST) of RPLC. Compared to local LSER and the LSST models, the global LSER model requires far fewer retention measurements for calibrating the model when different solutes and different mobile phase compositions are involved. Its fitting performance is equal to the local LSER model but worse than that of LSST. The poor fit of the global LSER results primarily from the local LSER model and not from the LSST model. The typical-conditions model (TCM) was developed based on a concept of multivariate space that is conceptually compatible with LSER. However, no LSER descriptors are used in the TCM approach. The number of input conditions needed in the typical-conditions model is determined by the chemical diversity of the solutes and the conditions involved. Principal component analysis (PCA) and iterative key set factor analysis (IKSFA) were used to find the number of typical conditions needed for a given data set. Compared to LSER, LSST, and global LSER, the typical-conditions model is more precise and requires fewer retention measurements for calibrating the model when different solutes and different stationary and/or mobile phases are involved.

Keywords: Linear solvation energy relationship; Linear solvent strength theory; Typical-conditions model; Retention prediction; Retention models

1. Introduction

Reversed-phase liquid chromatography (RPLC) is the most widely used chromatographic method for separating chemical mixtures [1,2]. One of the most time-consuming tasks in RPLC is the development of optimized methods for separating complex mixtures [3–5]. Changes in the stationary phase type and the eluent composition are often necessary to optimize the separation. To speed up method development, it is often necessary to predict retention of different solutes under different conditions. Such retention prediction requires a model to describe retention as a

*Corresponding author. Tel.: +1-612-624-0253; fax: +1-612-626-7541.

E-mail address: carr@chem.umn.edu (P.W. Carr).

0021-9673/02/$ – see front matter
PII: S0021-9673(01)01472-8

function of both solute properties and experimental conditions.

Retention can be related to the free-energy change associated with the solute distribution between the stationary and mobile phase. Many types of solute–phase interactions, including dispersive, dipolar, and hydrogen bonding interactions, all simultaneously contribute to the free energy of retention. The free energy contributions from these interactions depend not only on the ability of the solute to participate in these interactions, but also on the response of the phases to these interactions. That is, retention is inherently multivariate.

Many have recognized this multivariate nature of retention and have shown [6–21] that the retention of structurally diverse solutes under a given experimental condition can be modeled by a linear solvation energy relationship (LSER):

$$\log k = \log k_0 + vV_2 + s\pi_2^* + a\sum \alpha_2^{\mathrm{H}} + b\sum \beta_2^{\mathrm{H}} + rR_2 \quad (1)$$

where k is the solute retention and the subscript 2 denotes the solute descriptors, which include molar volume (V_2), dipolarity/polarizability (π_2^*), overall hydrogen-bond acidity ($\Sigma \alpha_2^{\mathrm{H}}$), overall hydrogen-bond basicity ($\Sigma \beta_2^{\mathrm{H}}$), and excess molar refraction (R_2). These descriptors represent the ability of the solute to participate in various solute–phase interactions. Each solute descriptor is multiplied by a coefficient that represents the system response (combination of mobile and stationary phases) to these interactions.

The LSER coefficients (v, s, a, b, r, and $\log k_0$) can be ''calibrated'' by measuring the retention of a set of *judiciously* chosen, structurally diverse solutes whose LSER descriptor values are known. The calibrated LSER equation can then be used to predict retention of other solutes whose LSER descriptor values are known. However, the LSER model has to be calibrated for *each* condition examined. Thus, the number of calibrations and the effort involved increase as the number of conditions of interest increases.

Snyder et al. have shown [22,23] that, in the binary aqueous–organic eluents typically used in RPLC, the retention of a solute can be useful approximated as a quasi-linear function of the mobile phase composition expressed as the volume fraction (ϕ) of the organic modifier in the mobile phase (this is termed linear solvent strength theory (LSST)):

$$\log k = \log k_{\mathrm{w}} - S\phi \quad (2)$$

The LSST coefficients ($\log k_{\mathrm{w}}$ and S) can be calibrated by measuring the retention of a solute at two eluent compositions. The calibrated LSST equation can then be used to predict retention of the same solute at other mobile phase compositions. However, LSST has to be calibrated for *each* solute examined, and the number of calibrations increase as the number of solutes of interest increases.

This paper describes two new models that are more general than either LSER or LSST. The first one is a global linear solvation energy relationship (global LSER) that expresses retention as a function of both solute LSER descriptors and eluent composition. The second is a so-called ''typical-conditions model'' (TCM) that expresses retention under a given chromatographic condition as a linear function of retention under other ''typical'' conditions. These other conditions are comprised of either different eluent types or compositions or different types of stationary phase. We developed these models so that fewer retention measurements would be required for calibration compared to either the LSER or LSST model when multiple solutes and multiple conditions are involved.

1.1. The global linear solvation energy relationship (global LSER)

Retention can be related to the equilibrium constant (K) of solute distribution between the stationary and mobile phases and can be further related to the free energy of solute transfer from the mobile phase to the stationary phase (ΔG°):

$$\log k = \log \Phi + \log K = \log \Phi - \Delta G^\circ / 2.3RT \quad (3)$$

where Φ is the ratio of the volume of the stationary phase to the volume of the mobile phase within the column. By definition, $\log k_{\mathrm{w}}$ in the LSST model is the hypothetical retention that the solute would have in a purely aqueous eluent ($\phi = 0$). Therefore, $\log k_{\mathrm{w}}$ can be related to the free energy of solute transfer from water to the stationary phase:

$$\Delta G^{\circ}_{w} = -2.3RT \log \frac{k_w}{\phi} \quad (4)$$

If we set ϕ in the LSST equation equal to unity, from Eq. (2) we see that

$$S = \log k_w - \log k_{org} \quad (5)$$

where $\log k_{org}$ denotes the retention in a purely organic eluent. Assuming that the stationary phase is not modified by sorption of mobile phase and using Eqs. (2), (4) and (5), S can be related to the free energy of solute transfer from water to pure organic mobile phase:

$$\Delta G^{\circ}_{S} = -2.3RTS \quad (6)$$

Since both $\log k_w$ and S are free energy parameters, we can, in principle, model $\log k_w$ and S using the LSER approach:

$$\log k_w = \log k_{0,w} + v_w V_2 + S_w \pi_2^* + a_w \sum \alpha_2^H + b_w \sum \beta_2^H + r_w R_2 \quad (7)$$

$$S = \log k_{0,S} + v_S V_2 + S_S \pi_2^* + a_S \sum \alpha_2^H + b_S \sum \beta_2^H + r_S R_2 \quad (8)$$

where v_w, s_w, a_w, b_w, r_w, and $\log k_{0,w}$ are LSER coefficients for $\log k_w$ and v_s, s_s, a_s, b_s, r_s, and $\log k_{0,s}$ are LSER coefficients for S. By replacing the two coefficients in the LSST model with the two LSER models [Eqs. (7) and (8)] and collecting the terms, we have effectively extended the *local* LSER to the *global* LSER:

$$\log k = (\log k_{0,w} - \log k_{0,S}\phi) + (v_w - v_S\phi)V_2 + (s_w - s_S\phi)\pi_2^* + (a_w - a_S\phi)\sum \alpha_2^H + (b_w - b_S\phi)\sum \beta_2^H + (r_w - r_S\phi)R_2 \quad (9)$$

A global LSER expresses retention as a concurrent function of both the solute LSER descriptors and mobile phase composition (ϕ) with a maximum of 12 coefficients regardless of the number of solutes and the number of mobile phase compositions under consideration. This is obviously a tremendous advantage over both the local LSER and LSST models, because the local LSER requires six coefficients for *every* value of ϕ examined, and LSST requires two coefficients for *every* solute examined. That is, fewer retention measurements are required for calibrating global LSER than for calibrating either local LSER when more than two values of ϕ are to be examined or LSST when more than six solutes are to be examined.

Kaliszan et al. [24,25] proposed a similar ''global'' approach based on quantum mechanically calculated molecular properties, the total molecular energy and the maximal excess electronic charge difference. They then tested their global approach with RPLC retention obtained on C_{18} stationary phases using methanol–water mobile phases. In prior work we have shown that the global approach is consistent with linear free energy theory, and tested our global LSER approach with RPLC retention obtained on C_8 stationary phases using three different types of mobile phases (methanol, acetonitrile, and tetrahydrofuran mixtures with water) [26]. The present paper gives the results of the global LSER approach applied to RPLC retention obtained on five different types of stationary phases using the same three types of eluent types (Table 1). We then compare the results with those of the typical-conditions model.

1.2. Typical-conditions model

Although the global LSER is more general than the local LSER, the values of the LSER descriptors must be known for all solutes of interest. At present, the values of LSER descriptors are not available for most existing and new chemicals, and their determination is experimentally difficult especially for the highly polar solutes typical of those of interest in biochemistry and the pharmaceutical industry.

Since multiple conditions are often necessary to optimize the separation of complex mixtures, predicting retention under one condition from retention under other conditions can be useful for minimizing the number of trial runs. Since similar solute–phase interactions can operate under different conditions, there must be an inherent correlation of retention from different conditions. That is, the retention under one condition might be useful as an empirical ''solute descriptor'' to model the retention under other conditions.

Many years ago, Collander suggested [27] that there might be a linear relationship between two

Table 1
Thirty-two conditions comprised of five stationary phases and three types of eluents at four compositions

	Condition	Stationary phase	Mobile phase		Condition	Stationary phase	Mobile phase
A	1	Betasil-C_{18}	30% ACN	E	17	POMA-ZrO_2	60% MeOH
	2	Betasil-C_{18}	40% ACN		18	POMA-ZrO_2	50% MeOH
	3	Betasil-C_{18}	50% ACN		19	POMA-ZrO_2	40% MeOH
	4	Betasil-C_{18}	60% ACN		20	POMA-ZrO_2	30% MeOH
B	5	Betasil-C_{18}	40% MeOH	F	21	PS-ZrO_2	50% ACN
	6	Betasil-C_{18}	50% MeOH		22	PS-ZrO_2	40% ACN
	7	Betasil-C_{18}	60% MeOH		23	PS-ZrO_2	30% ACN
	8	Betasil-C_{18}	70% MeOH		24	PS-ZrO_2	20% ACN
C	9	Betasil-C_{18}	20% THF	G	25	Phenyl-SiO_2	70% ACN
	10	Betasil-C_{18}	30% THF		26	Phenyl-SiO_2	60% ACN
	11	Betasil-C_{18}	40% THF		27	Phenyl-SiO_2	50% ACN
	12	Betasil-C_{18}	50% THF		28	Phenyl-SiO_2	40% ACN
D	13	POMA-ZrO_2	40% ACN	H	29	PRP-1	80% ACN
	14	POMA-ZrO_2	30% ACN		30	PRP-1	70% ACN
	15	POMA-ZrO_2	20% ACN		31	PRP-1	60% ACN
	16	POMA-ZrO_2	10% ACN		32	PRP-1	50% ACN

partition coefficients (K_1 and K_2) from two different but related partition systems:

$$\log K_2 = a \log K_1 + b \tag{10}$$

If we treat retention in RPLC as a partition process, retention of solutes can be related to their partition coefficients (K) for the partitions between the stationary and mobile phases:

$$\log k = \log \Phi + \log K \tag{11}$$

where Φ is the ratio of the stationary phase volume to the mobile phase volume within the column. The Collander type relation has been extended to correlate retention under a given condition with retention under another condition [28]. However, good correlations in RPLC are obtained for structurally diverse solutes *only when* the two conditions are chemically very similar or *only for* structurally related solutes if the two conditions are chemically different [29].

LSER models retention of structurally diverse solutes under each condition as a linear combination of solute properties that represent the ability of the solutes to participate in various solute–phase interactions. The LSER equations established for different conditions differ only in the combination of coefficients that can be regarded as the phase response to these interactions. If LSER is a realistic model for RPLC retention, then from the theories of linear algebra and multivariate statistics, there must exist multivariate relationships to enable the correlation of retention under one condition to retention under several other conditions. That is, retention under several (hopefully only a few) conditions can be used as empirical but precise ''descriptors'' to model and predict the retention under other desired conditions.

The number of conditions that must be included in this multivariate correlation is dictated by the number of *active* solute–phase interactions responsible for the systematic variations in the retention of all the solutes and conditions involved. An intermolecular interaction is active if the ability of different solutes to participate in the interaction and the phase response to the interaction under different conditions *are both significantly different*. However, including an excessive number of conditions introduces the ''multicollinearity'' problem in the correlation [30], causing large uncertainties in predictions using the correlation; including fewer conditions loses pertinent information in the correlation, causing poor correlation performance.

If we arrange the retention data for all the solutes

and conditions of interest in a matrix, the number of solute–phase interactions active in the matrix can be determined from principal component analysis (PCA) of the matrix. Geometrically, each solute in the matrix can be considered as a point in the multidimensional *data space* whose coordinate axes are defined by the conditions in the matrix. Assuming that the number of conditions in the matrix is larger than the number of active solute–phase interactions, the solutes will form a lower dimensional pattern in the high dimensional data space. PCA allows us to extract the *minimum* number of new axes (factors) as linear combinations of the original axes (retention under different conditions) to describe the inherent pattern in the data space. The factors are determined such that the first factor points in the direction of maximum variation in the multidimensional cloud of data points. Each successive factor points at right angles (orthogonal) to the prior ones and explains the maximum data variation not accounted for by the prior factors. These factors are ordered in terms of the fraction of the retention variation which they explain.

The number of factors needed to describe the data pattern is only as large as the number of active solute–phase interactions that are responsible for the data pattern. Any additional PCs explain the remaining variation contributed mainly by the experimental uncertainties associated with the retention measurements. To determine the number of active solute–phase interactions that are responsible for the data pattern, we should retain a minimum number of factors to describe only the *systematic* variations in the retention and exclude those factors that describe the variations caused mainly by the experimental uncertainties.

In principle, once the minimum number of factors are determined for a retention data matrix, retention under any condition in the matrix can be modeled by the retention under the same number of other conditions in the same matrix. However, the retention under various conditions are often highly correlated. If the highly correlated conditions are included in the correlation, large uncertainties can result in predicting retention using that correlation. Therefore, the least correlated conditions must be selected from the matrix to provide the most reliable predictions. These least correlated conditions are called ‘‘typical conditions’’, and the multivariate correlation using these conditions is thus termed the ‘‘typical-conditions model’’.

Lochmüller et al. found that three factors are required in their PCA models to describe retention obtained on similar RPLC columns (all C_{18} stationary phases) with binary and ternary eluents of different compositions. They then used target factor analysis (TFA) to select a ‘‘key set’’ of three conditions from the same data matrix to approximate the factor space for retention prediction [31–33]. The key conditions were selected such that they reproduce the data with minimum error. We believe that there is a strong possibility for highly correlated conditions to be selected as key conditions using this minimum-error approach and the resulting correlation using these key conditions can cause large uncertainties in predicting the retention of other solutes using the correlation.

Geometrically, each condition in the matrix can be considered as a vector in the multidimensional *data space* whose coordinate axes are defined by the solutes in the matrix. Correlation between two conditions can be measured by the angle spanned by the two condition vectors. The typical conditions are the most orthogonal conditions available in a retention data matrix. When a retention data matrix is small and the dimension of the inherent data pattern is low, finding the most orthogonal conditions poses little difficulty, because we can quickly compare all possible combinations for the most orthogonal conditions. However, for a large data matrix having c conditions and an n-dimensional inherent data pattern, there would be $c!/(c-n)!n!$ possible combinations. The number of combinations can become very large and the time needed to perform an exhaustive search can be too long to use it routinely. In this case, iterative key set factor analysis (IKSFA) [34] can be used to speed up the search for the most orthogonal conditions. IKSFA uses PCs as guides to find the most orthogonal conditions such that the number of searches required is significantly reduced compared to the exhaustive search.

2. Experimental

The retention data in this paper were taken from

Refs. [35–37] which give detailed descriptions of the experimental conditions employed. All the chromatographic measurements were made on a Hewlett-Packard 1100 chromatograph, equipped with a quaternary pump, a vacuum degasser, an autosampler, a thermostatted-column compartment controlled at 30°C, a variable wavelength UV detector and Chemstation software (Hewlett-Packard, Wilmington, DE). The averages of triplicate measurements of the retention time were used to calculate the retention factors. The average experimental uncertainty is estimated to be 0.02 in log k units.

Five different RPLC stationary phases were used in this data set. They were chosen to be chemically diverse (Table 1) and responsive to different types of intermolecular interactions. Betasil-C_{18} is a typical ODS-bonded phase based on silica. The particles (5 μm) were a gift from Keystone Scientific (Bellefonte, PA). POMA-ZrO_2 is a homemade zirconia-based phase with poly(octadecene-co-maleic acid) coated and cross-linked on bare zirconia particles (3.1 μm) [36]. PS-ZrO_2 is another homemade zirconia-based phase with polystyrene coated and cross-linked on bare zirconia particles (~2.5 μm) [38]. Phenyl-SiO_2 is a phenyl-bonded phase based on silica (5 μm) and was donated by Alltech Associates (DeerField, IL). Hamilton PRP-1 is a poly(styrene–divinylbenzene) resin (10 μm) phase, and was purchased from Chrom Tech (Apple Valley, MN).

The most common organic modifiers, methanol (MeOH), acetonitrile (ACN), and tetrahydrofuran (THF), were used in binary aqueous mobile phases. Four compositions for each organic modifier were chosen to cover a wide and linear range of the LSST model so that the performance of the global LSER and typical-conditions models could be compared to that of the more conventional LSST approach. HPLC-grade solvents were used for the mobile phases.

A total of 22 structurally diverse solutes were carefully chosen to cover a wide range of chemical and structural properties in terms of LSER solute descriptors (Table 2). This small set of solutes does a very good job of reproducing the LSER fitting coefficients obtained by Tan in a study with over 80 structurally diverse solutes [7,9]. However, no amines or other ionic solutes were included in order to avoid complexities from ionic interactions during

Table 2
Twenty-two test compounds and their LSER descriptor values

No.	Solute	$V_x/100$[a]	π_2^*	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	R_2
1	*N*-benzyl formamide	1.1137	1.80	0.40	0.63	0.990
2	Benzylalcohol	0.9160	0.87	0.33	0.56	0.803
3	Phenol	0.7751	0.89	0.60	0.30	0.805
4	3-Phenyl propanol	1.1978	0.90	0.30	0.67	0.821
5	*p*-Chlorophenol	0.8975	1.08	0.67	0.20	0.915
6	Acetopheonone	1.0139	1.01	0.00	0.48	0.818
7	Benzonitrile	0.8711	1.11	0.00	0.33	0.742
8	Nitrobenzene	0.8906	1.11	0.00	0.28	0.871
9	methyl benzoate	1.0726	0.85	0.00	0.46	0.773
10	Anisole	0.9160	0.75	0.00	0.29	0.708
11	Benzene	0.7164	0.52	0.00	0.14	0.610
12	*p*-Nitrotoluene	1.0315	1.11	0.00	0.28	0.870
13	*p*-Nitrobenzyl chloride	1.1539	1.34	0.00	0.40	1.080
14	Toluene	0.8573	0.52	0.00	0.14	0.601
15	Benzophenone	1.4808	1.50	0.00	0.50	1.447
16	Bromobenzene	0.8914	0.73	0.00	0.09	0.882
17	Naphthalene	1.0854	0.92	0.00	0.20	1.340
18	Ethylbenzene	0.9982	0.51	0.00	0.15	0.613
19	*p*-Xylene	0.9982	0.52	0.00	0.16	0.613
20	*p*-Dichlorobenzene	0.9612	0.75	0.00	0.02	0.825
21	Propylbenzene	1.1391	0.50	0.00	0.15	0.604
22	*n*-Butylbenzene	1.2800	0.51	0.00	0.15	0.600

[a] Values of V_x were taken from Ref. [43], while values of π_2^*, $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and R_2 were obtained from Refs. [44,45].

the retention process. Modeling of retention involving ionic interactions is left for future studies. All solutes were obtained commercially. The retention values of propylbenzene and butylbenzene on Betasil-C_{18}/20% THF were missing. The retention values were estimated from the retention of toluene and ethylbenzene using the Martin equation [39–41]. The retention values of *p*-nitrobenzyl chloride on Phenyl-SiO_2/70% ACN and Phenyl-SiO_2/60% ACN were also missing. They were estimated from retention at 40 and 70% ACN mobile phase compositions using LSST. The original retention value of 3-phenyl propanol on PRP-1/60% ACN deviated tremendously from the straight line of linear solvent strength theory and therefore was excluded as an outlier. The retention value of the solute used in the analysis was estimated from its retention at 50, 70 and 80% ACN mobile phase compositions using LSST.

Data analysis package in Excel of Microsoft® Office 97 was used for multiple linear regression and related statistical calculations. Singular value decomposition and other matrix functions in MATLAB® 5.2 from MathWorks were used for principal component analysis and other matrix calculations.

3. Results and discussions

3.1. Global LSER

The log *k* values of all 22 solutes at each mobile phase composition were used to fit the local LSER model. Therefore, there are a total of 32 separate local LSER fits for all 32 conditions, and a total of 192 (32×6) fitting coefficients are to be determined for this data set using local LSER. The log *k* values of each solute for each pair of the stationary and mobile phases were used to fit the LSST model. Therefore, there are a total of 176 (22×8) separate LSST fits for all 22 solutes and eight pairs of the stationary and mobile phases, and a total of 352 (22×8×2) fitting coefficients are to be determined for this data set using LSST. The log *k* values of all 22 solutes at all mobile phase compositions were concatenated to fit the global LSER model. Therefore, there are only eight separate global LSER fits for all 22 solutes and eight pairs of the stationary and mobile phases, and a total of 96 (12×8) fitting coefficients needs to be determined for this data set using the global LSER. Obviously, more coefficients would be needed for the LSER fits if more mobile phase compositions were used, and more coefficients would be needed for the LSST fits if more solutes were included.

Typical local LSER, LSST, and global LSER fits are given in Figs. 1–3. We see from the plot that local LSER can explain most of the data variation, but the standard deviation of the LSER fit is much larger than the typical experimental uncertainty (0.02 in log *k* unit). The fitting performance of LSST is much better than that of LSER, and the fitting performance of global LSER is as good as that of LSER but worse than that of LSST. Given the smaller average residuals using LSST, it is clear that a large fraction of the residuals of the LSER fits is due to model defects and not to random error in the measured retention factors. The fitting performance of LSST is adequate for method development, at least over the range in the mobile phase compositions covered here, but LSER is not good enough for method development.

Since fewer regression coefficients are needed in a

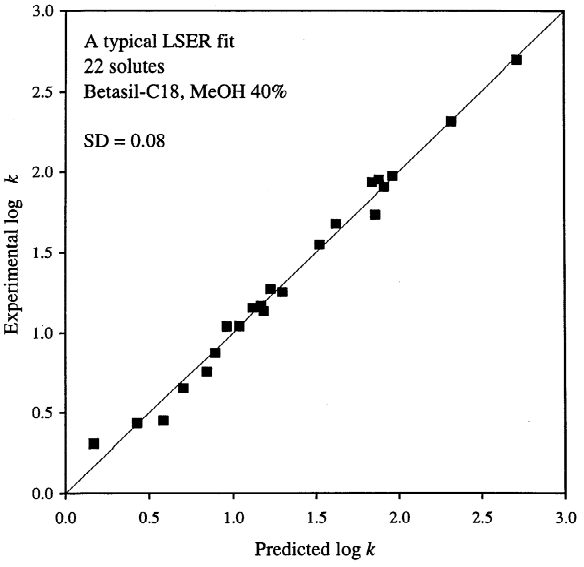

Fig. 1. A typical LSER fit.

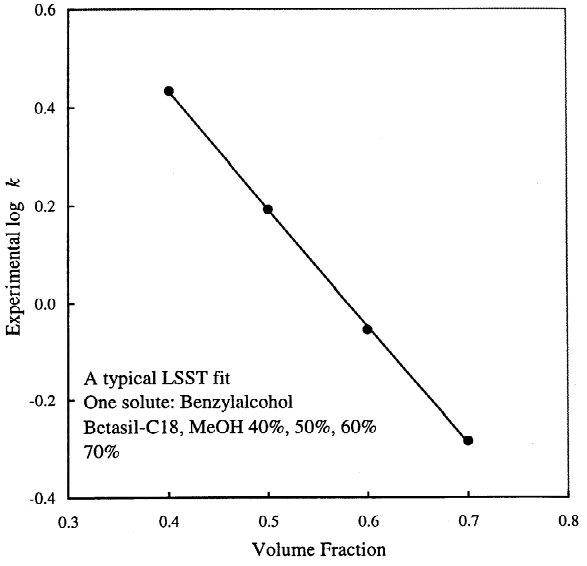

Fig. 2. A typical LSST fit.

global LSER fit than in a series of local LSER or LSST fits for the same data, we expect that the goodness-of-fit of the global LSER fits should be worse than that of the local LSER or LSST fits. To test *if* the goodness-of-fit of the global LSERs is significantly worse than that of local LSERs and that of LSSTs, we did a one-tailed *F*-test on the residual

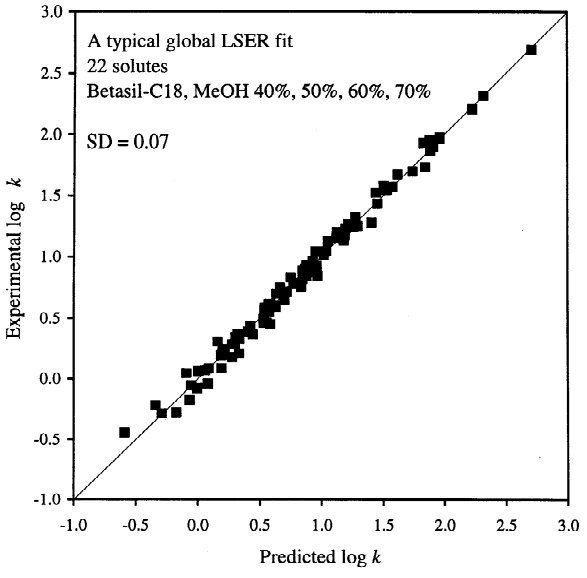

Fig. 3. A typical global LSER fit.

mean squares pooled from all eight global LSER fits, from all 32 local LSER fits, and from all 176 LSST fits, respectively (Table 3). Despite the two-fold decrease in the number of fitting coefficients used, the *F*-test shows that the goodness-of-fit of the global LSERs is *not statistically worse* than that of the local LSERs.

These results confirm that the local LSER model for a single mobile phase composition can be effectively extended to the global LSER model for multiple mobile phase compositions within the range of mobile phase compositions considered here. Hence, after a total of 12 regression coefficients are empirically determined for each type of organic modifier, the global LSER model can be used to predict the retention of other solutes at other mobile phase compositions within the range of mobile phase compositions where LSST is valid. In contrast, the local LSER model requires a different set of six regression coefficients at *every* mobile phase composition. When retention prediction at more than two mobile phase compositions is attempted, the global LSER approach will be more efficient than the local LSER model applied at the same number of mobile phase compositions.

The *F*-test indicates that the global LSER fits are significantly poorer than the LSST fits. Since the global LSER is derived from both LSER and LSST, and the standard deviation of the pooled LSST fits is significantly better than that of the local LSER fits for the same data, we conclude that the large residuals of the global LSER fits must result primarily from the local LSERs, and the performance of global LSER is limited by that of the local LSER model. We are convinced that for the global LSER approach to achieve the same precision that is possible with LSST, significant improvements in the LSER model and/or solute's descriptor values are necessary. We now proceed to develop the typical-conditions model which does not use the LSER descriptors.

3.2. Typical-conditions model

To model the retention under various combinations of different conditions, the 32 conditions (Table 1) were divided into 14 groups (Table 4) each including a different number and types of stationary

Table 3
F-tests on the residual mean squares pooled from global LSER fits, from local LSER fits, and from LSST, respectively

Model	Number of fits	*SD*[a]	s^{2}[b]	df[b]	*F*-ratio	F_c ($\alpha=0.1$)
Local LSER	32	0.077	0.0059	512	1.07	1.11
LSST	176 (22×8)	0.031	0.0010	352	5.77	1.14
Global LSER	8	0.074	0.0055	608		

[a] *SD* is square root of residual mean squares.
[b] s^2 and df are residual mean squares and degree of freedom for the *F*-test, respectively.

and mobile phases. The same 22 solutes (Table 2) were used in each group.

First, PCA was applied to the matrix from each group to find the minimum number of factors required to explain the systematic variations in the matrix. Next, the same number of the most orthogonal conditions (typical conditions) was then selected from the matrix, and the retention under these typical conditions was used to model the retention under other conditions in the same matrix. Finally, a quadratic model of the relationship between log *k* and ϕ was used together with a typical-conditions model to reduce the number of retention measurements required for calibration when multiple compositions of each organic modifier are involved.

3.2.1. The number of conditions needed in the typical-conditions model

In principle, the number of factors needed to explain the systematic variations in a retention data matrix can be determined graphically from the change in the percent residual variance (PRV) on the number of factors used in a PCA model [34]. PRV is the part of retention variation that is not accounted for by the factors included in a PCA model:

$$PRV(j) = 100 \cdot \left(\frac{\sum_{i=1}^{r} \sum_{k=1}^{c} \left(\log k_{ik} - \log \hat{k}_{ik}(j)\right)^2}{\sum_{i=1}^{r} \sum_{k=1}^{c} \log k_{ik}^2} \right) \quad (12)$$

where $\log k_{ik}$ is the original retention data for solute *i* on condition *k* and $\log \hat{k}_{ik}(j)$ is the retention data calculated from the PCA model using *j* factors. When the PRV is plotted against the number of factors in the PCA model, the plot should show a continuous decrease with each additional PC added to the model because of the smaller retention variation explained by each successive factor. The retention variations contributed by the experimental uncertainties should be relatively small and similarly sized, if the experimental uncertainties associated with the data are truly random. Therefore, a plot of the PRVs versus number of factors should level off when all the significant factors have been included in the PCA model [34,42].

Plots of the percent residual variance versus number of factors for each group are given in Figs. 4–7. The breakpoints in the reduction of percent residual variance in the plots are quite clear. However, we can see from expanded plots that the decreases in the percent residual variance after the breakpoints continue and become smaller with each successive factor included in the PCA models. That is, with the number of factors at the breakpoints retained in the PCA model, some minor systematic variations in the residuals are still left unmodeled.

Retaining too many factors imports more noise than signal, and retaining too few factors excludes pertinent information. Therefore, a trade-off between the number of factors retained in the PCA models (which should be minimal) and the variances accounted for by these factors (which should be maximal) has to be made.

Since the purpose of this work is to use the retention under typical conditions to model and predict retention under other conditions, it makes no sense to predict retention beyond the experimental uncertainties in the retention measurements. Therefore, we compared the residual standard deviations at the breakpoints with the estimated experimental uncertainties in our data (~0.02 in log *k* unit) to decide if more factors were needed to model the retention properly.

We can see from Figs. 4–7 that the residual standard deviations at the breakpoints are all lower than the estimated experimental uncertainty. That is,

Table 4
Fourteen groups of conditions for principal component analysis, the number of factors needed in a PCA model to explain the systematic variations in the retention data matrix of each group, and residual standard deviations of the PCA models

Group	Stationary phases[a]	Mobile phases[a]	Number of conditions	Number of factors	Standard deviation	Typical conditions column type/Eluent type volume fraction
1	Betasil-C_{18} POMA-ZrO_2 PS-ZrO_2 Phenyl-SiO_2 PRP-1	ACN MeOH THF	32	6	0.013	Betasil-C_{18}/ACN 0.3 Betasil-C_{18}/MeOH 0.4 Betasil-C_{18}/THF 0.2 POMA-ZrO_2/MeOH 0.6 PS-ZrO_2/ACN 0.2 PRP-1/ACN 0.8
2	Betasil-C_{18}	ACN MeOH THF	12	4	0.011	Betasil-C_{18}/ACN 0.3 Betasil-C_{18}/MeOH 0.5 Betasil-C_{18}/THF 0.2 Betasil-C_{18}/THF 0.5
3	Betasil-C_{18}	ACN MeOH	8	3	0.009	Betasil-C_{18}/ACN 0.3 Betasil-C_{18}/MeOH 0.4 Betasil-C_{18}/MeOH 0.7
4	Betasil-C_{18}	ACN THF	8	3	0.010	Betasil-C_{18}/ACN 0.3 Betasil-C_{18}/THF 0.2 Betasil-C_{18}/THF 0.5
5	Betasil-C_{18}	MeOH THF	8	3	0.015	Betasil-C_{18}/MeOH 0.4 Betasil-C_{18}/THF 0.2 Betasil-C_{18}/THF 0.5
6	POMA-ZrO_2	ACN MeOH	8	3	0.009	POMA-ZrO_2/ACN 0.1 POMA-ZrO_2/ACN 0.4 POMA-ZrO_2/MeOH 0.6
7	Betasil-C_{18}	ACN	4	2	0.005	Betasil-C_{18}/ACN 0.3 Betasil-C_{18}/ACN 0.6
8	Betasil-C_{18}	MeOH	4	2	0.004	Betasil-C_{18}/MeOH 0.4 Betasil-C_{18}/MeOH 0.7
9	Betasil-C_{18}	THF	4	2	0.009	Betasil-C_{18}/THF 0.2 Betasil-C_{18}/THF 0.5
10	POMA-ZrO_2	ACN	4	2	0.010	POMA-ZrO_2/ACN 0.1 POMA-ZrO_2/ACN 0.4
11	POMA-ZrO_2	MeOH	4	2	0.007	POMA-ZrO_2/MeOH 0.3 POMA-ZrO_2/MeOH 0.6
12	PS-ZrO_2	ACN	4	2	0.009	PS-ZrO_2/ACN 0.2 PS-ZrO_2/ACN 0.5
13	Phenyl-SiO_2	ACN	4	2	0.005	Phenyl-SiO_2/ACN 0.4 Phenyl-SiO_2/ACN 0.7
14	PRP-1	ACN	4	2	0.007	PRP-1/ACN 0.5 PRP-1/ACN 0.8

[a] Each pair of stationary and mobile phases has four conditions of different mobile phase compositions.

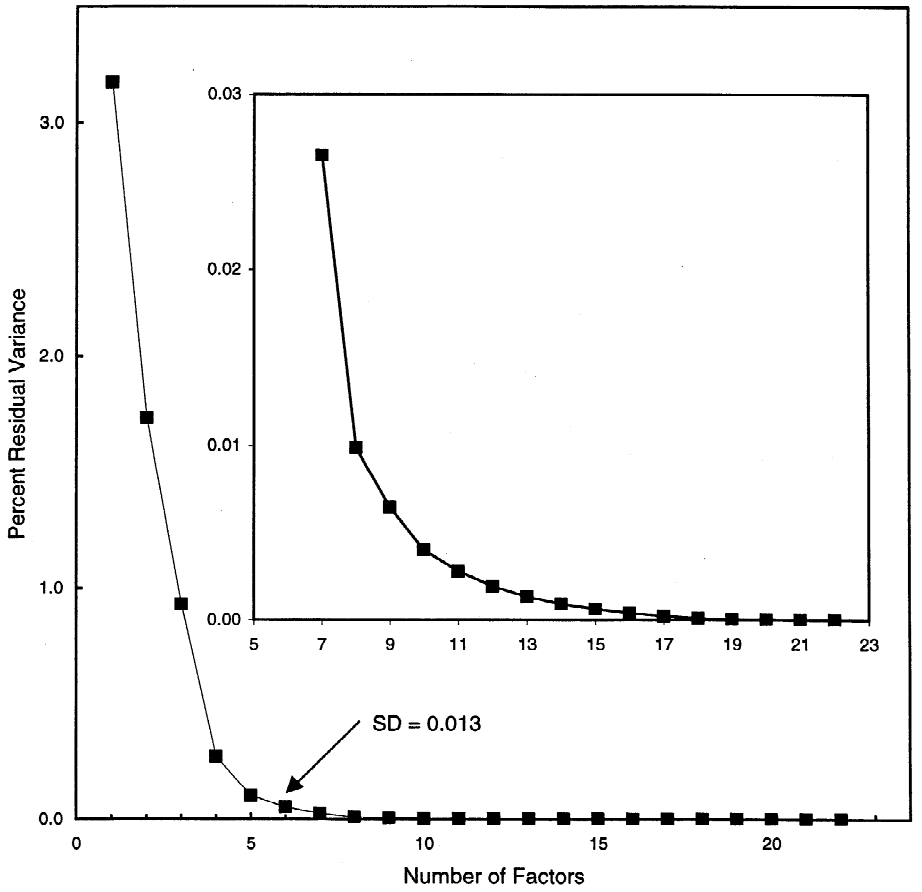

Fig. 4. Plot of percent residual variance versus the number of factors for the retention data matrix of 32 conditions comprised of five stationary phases and three types of organic modifiers each at four mobile phase compositions. Six factors are needed for the PCA model to explain the systematic variations in the matrix.

six factors are needed to explain the systematic variations in retention when all 32 conditions are considered simultaneously, and only two, three, and four factors are required when a single stationary phase is used with one, two, and three types of mobile phases, respectively. The results are summarized in Table 4. We can see from the table that the number of required factors decreases as fewer stationary or mobile phases are included in each group. This is consistent with the fact that there is less variability in the phase chemical properties in a group when fewer stationary or mobile phases are included in the group. The reduction in the number of required factors is important for practical applications of the typical-conditions approach to retention prediction. As the number of TCs decrease, the less data one needed to collect for retention prediction when only those conditions from one of these groups are involved. We expect that additional factors will be necessary to account for possible ionic and other interactions when amines and/or metal complexing agents are included in the data sets. Similarly, if one were to consider a novel phase in which charge transfer interactions were significant, additional solute and phase factors would be required.

Only six factors are needed for the 32 conditions comprised of five different stationary phases and three different organic modifiers, each at four volume fraction compositions. In contrast, modeling the same data using the six-terms, LSER would require

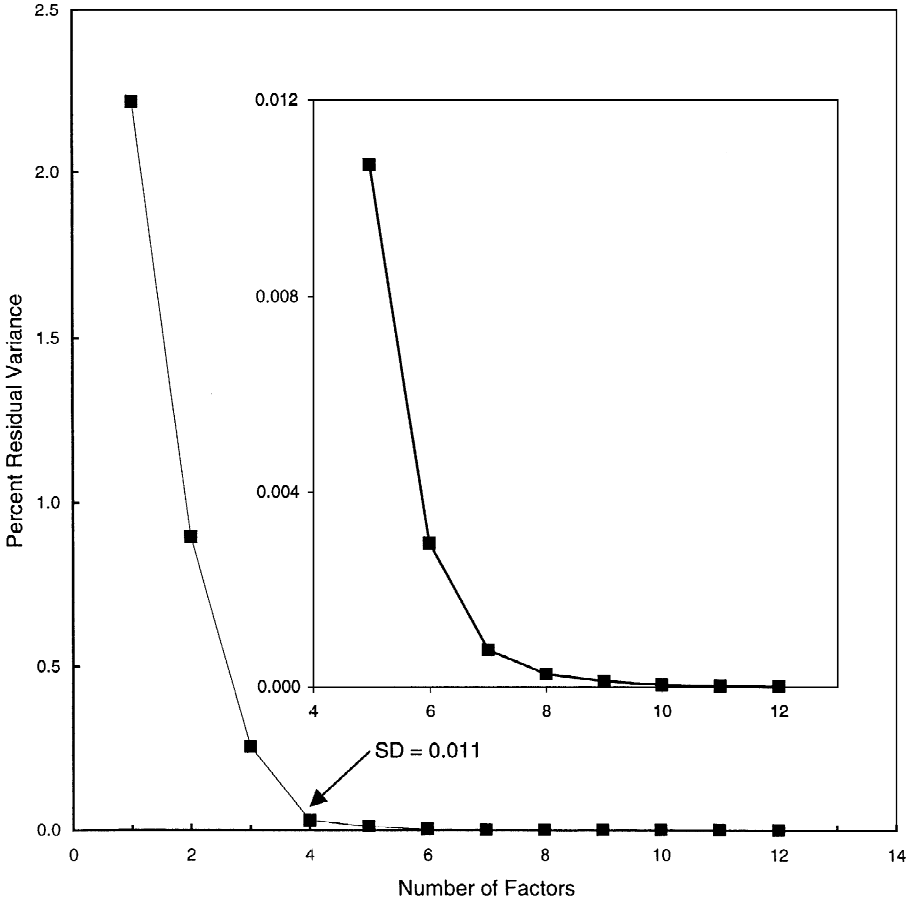

Fig. 5. Plot of percent residual variance versus the number of factors for the retention data matrix of 12 conditions comprised of one Betasil-C_{18} stationary phase and three types of organic modifiers each at four mobile phase compositions. Four factors are needed for the PCA model to explain the systematic variations in the matrix.

192 (6 parameters×32 conditions) parameters to be determined *independently*. Using the two-terms LSST would require 176 (22 solutes×8 pairs of stationary and mobile phases) parameters to be determined *independently*. Based on the PCA results, the LSER parameters determined for different conditions and the LSST parameters determined for different solutes and different pairs of stationary phase and organic modifiers are *not independent*.

3.2.2. Using the typical conditions to model retention under other conditions

Once the minimum number of factors had been determined for each group, iterative key set factor analysis (IKSFA) was used to select the same number of the most orthogonal conditions from the group as the typical conditions for the calibration (Table 4). These typical conditions represent the maximally different and thus chemically most informative conditions available from all the conditions in the group.

The six typical conditions selected for the group of 32 conditions (group 1 in Table 4) contain all the stationary and mobile phase types used except for the Phenyl-SiO_2 stationary phase. That is, although Phenyl-SiO_2 is quite different from the other two polymeric aromatic phases (PS-ZrO_2 and PRP-1), its contribution to the retention variation can be quantitatively described by the retention under conditions

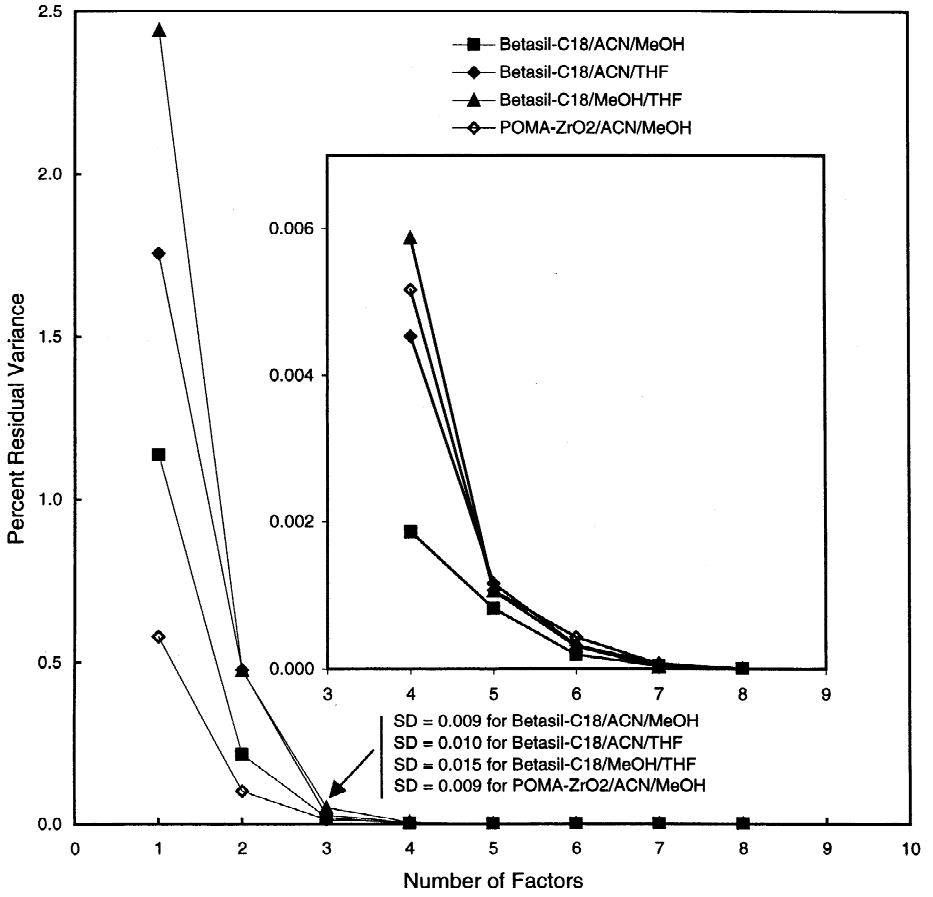

Fig. 6. Plot of percent residual variance versus the number of factors for the retention data matrix for each of four groups of eight conditions comprised of one stationary phase and two types of organic modifiers each at four mobile phase compositions. Three factors are needed for the PCA model to explain the systematic variations in each matrix.

using the other two polymeric aromatic phases. Each of three mobile phase modifiers is unique in its chemical properties, so all three must be included in the typical-conditions model as expected. For other groups of conditions where a single stationary phase is involved (groups 7–14 in Table 4), conditions with the same mobile phase modifier but with the two extreme compositions were selected as the typical conditions.

Once the typical conditions had been determined for each group of data, we linearly correlated the retention under each of the remaining conditions in a group with the retention under the typical conditions for the same group. Obviously data obtained under the typical conditions are not used as dependent variables because of their use as independent variables in the regressions.

A typical TCM fit is shown in Fig. 8, where the retention from Phenyl-SiO_2 and 40% ACN are modeled by the retention from the six typical conditions (group 1 in Table 4). Although Phenyl-SiO_2 was not included as a typical condition, retention on this phase is precisely predicted from the retention under the six typical conditions.

To compare the fitting performance of the typical-conditions approach with that of both the LSER and LSST models, we applied both the LSER and LSST models to the same data in each group, and did one-sided F-tests on the residual mean squares pooled from the three models (Table 5). The F-tests

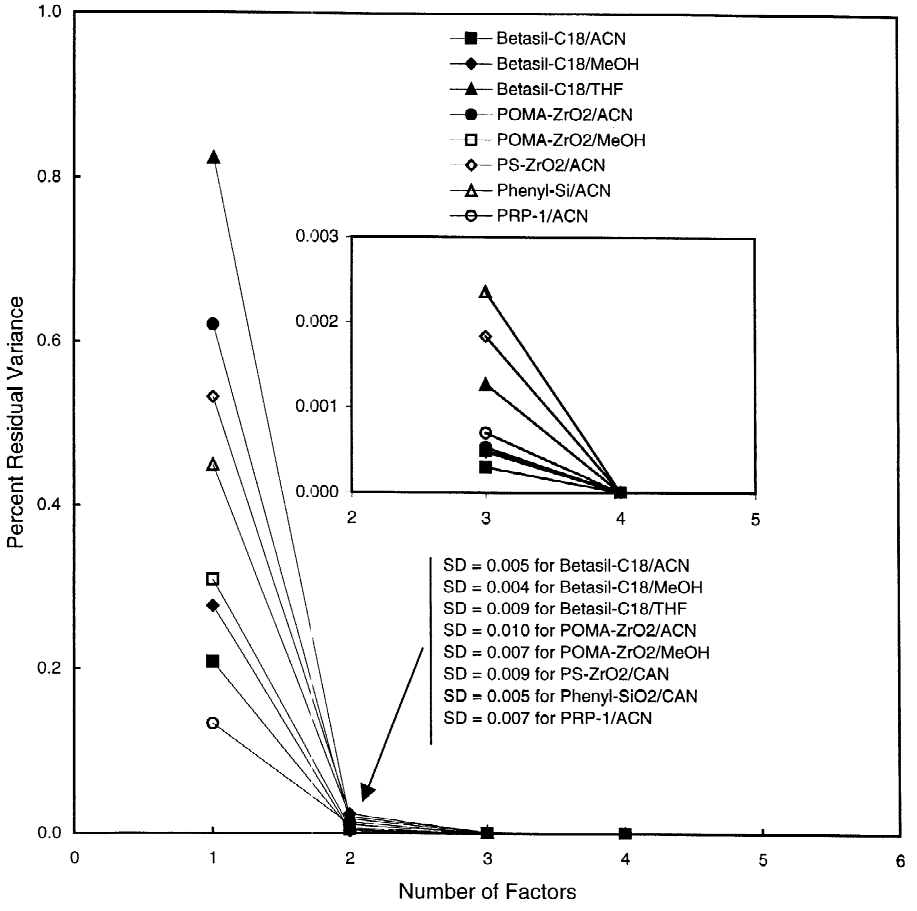

Fig. 7. Plot of percent residual variance versus the number of factors for the retention data matrix for each of eight groups of four conditions comprised of one stationary phase and one type of organic modifier at four mobile phase compositions. Two factors are needed for the PCA model to explain the systematic variations in each matrix.

show that the goodness-of-fit of the typical-conditions model is statistically better than that of either the LSER or LSST models applied to the same data in every group of conditions except the PRP-1/ACN group where the goodness-of-fits of the typical-conditions and LSST models are not significantly different due to the exceptionally good LSST fits for this group.

Although LSER has the advantage of providing chemical insight and LSST is simpler and easier to apply in method development, the quantitative performance of both the LSER and LSST models is poorer than the typical-conditions model as developed here. LSER is an approximate retention model in which key interactions go unmodeled. For example, there are no explicit terms representing silanophilic interactions, quadrapole interactions, and molecular shape factors. LSST approximates a generally curved relationship between log k and mobile phase composition with a quasi-linear relationship over a limited range of mobile phase compositions [23]. When the curvature becomes significant, the accuracy of the LSST retention prediction deteriorates.

Since the LSST fits for the PRP-1/ACN group are so much better than for the other groups, we suspected that the curvature in other groups might be significant enough to cause poorer LSST fits. There-

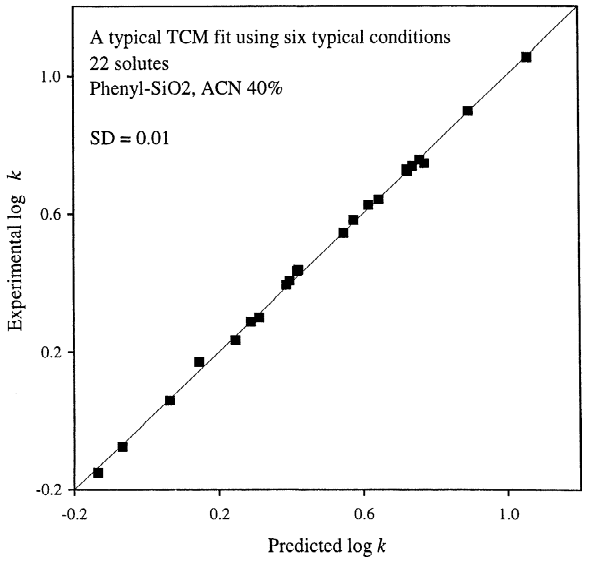

Fig. 8. A typical TCM fit using six typical conditions.

fore, we refitted each group using a quadratic relationship [23] between log k and mobile phase composition:

$$\log k = A\phi^2 + B\phi + C \quad (13)$$

and did one-sided F-tests on the residual mean squares pooled from the quadratic and typical-conditions fits (Table 6). The F-tests show that the improvement in goodness-of-fit of the quadratic model over the LSST model makes the goodness-of-fits for some groups close to or even better than the typical-conditions fits.

3.2.3. Retention as a function of both mobile phase composition and retention under typical conditions

In principle, retention under the typical conditions can be used to predict the retention in any mobile phase composition over the same range of mobile phase compositions used to find the typical conditions, because no new factor should be introduced by just changing the composition in the same composition range. The typical-conditions calibration for a new composition can be established from the retention measurements of a set of standard solutes under the typical conditions and at that new composition. The standard solutes must be structurally diverse, and the number of the standard solutes must be larger than the number of typical conditions included in the calibration (two or three times more may be necessary to cover adequate ranges of solute properties for the calibration). The calibrated typical-conditions equation can then be used to predict the retention of new solutes at that new composition from their retention measurements under the typical conditions.

However, when retention at many new compositions is to be predicted using the typical-conditions model, the number of retention measurements for calibrating the model at these compositions can be overwhelming. We found that the number of retention measurements for the calibration can be reduced when the retention is predicted from a quadratic retention model [Eq. (13)]. The retention of each of the standard solutes at a new composition is first predicted from the retention of the same solute at three mobile phase compositions, and the *predicted* retention is then used to calibrate the typical-conditions model for that new composition. We applied the quadratic retention model to the same 14 retention data matrices after PCA and IKSFA, and found that the performance of this combined quadratic and typical-conditions approach is not statistically different from that of the typical-conditions model using measured retention data. More retention measurements would be required if only the quadratic model was used for the same data.

The combined use of both the quadratic and typical-conditions models is better explained using the diagram below for a hypothetical group of conditions comprised of two organic modifiers, assuming that three typical conditions are required for the calibration. The retention of a set of standard solutes is used to calibrate both the quadratic and typical-conditions models. First, the retention of all the standard solutes under three typical conditions (TC1, TC2, and TC3) is measured. Second, the retention of each standard solute at three compositions for each organic modifier (QC1, QC2, and QC3 for the first organic modifier, and QC1′, QC2′, and QC3′ for the other organic modifier) is measured to calibrate the quadratic model for the solute. Some of the retention measurements can be used in both the quadratic and typical-conditions models. Next, the

Table 5
F-test on the residual mean squares pooled from typical-conditions regressions, from LSER regressions, and from LSST regressions using the retention data from each of the 14 groups

Group	Stationary phases	Mobile phases	Number of conditions	Model	SD^a	S^{2b}	df[b]	*F*-ratio	F_c ($\alpha = 0.1$)
1	Betasil-C_{18} POMA-ZrO_2 PS-ZrO_2 Phenyl-SiO_2 PRP-1	ACN MeOH THF	32	LSER LSST TCM	0.077 0.031 0.021	0.00591 0.00096 0.00044	512 352 390	13.37 2.16	1.13 1.14
2	Betasil-C_{18}	ACN MeOH THF	12	LSER LSST TCM	0.077 0.041 0.017	0.00591 0.00165 0.00030	192 132 136	19.87 5.55	1.23 1.25
3	Betasil-C_{18}	ACN MeOH	8	LSER LSST TCM	0.061 0.028 0.018	0.00368 0.00079 0.00031	128 88 90	11.77 2.52	1.29 1.31
4	Betasil-C_{18}	ACN THF	8	LSER LSST TCM	0.078 0.049 0.017	0.00608 0.00241 0.00029	128 88 90	20.80 8.24	1.29 1.31
5	Betasil-C_{18}	MeOH THF	8	LSER LSST TCM	0.089 0.042 0.027	0.0080 0.00175 0.00073	128 88 90	10.84 2.38	1.29 1.31
6	POMA-ZrO_2	ACN MeOH	8	LSER LSST TCM	0.100 0.020 0.013	0.00993 0.00039 0.00018	128 88 90	54.56 2.14	1.29 1.31
7	Betasil-C_{18}	ACN	4	LSER LSST TCM	0.043 0.038 0.008	0.00181 0.00145 0.00006	64 44 38	28.31 22.73	1.47 1.51
8	Betasil-C_{18}	MeOH	4	LSER LSST TCM	0.075 0.011 0.007	0.00556 0.00012 0.00004	64 44 38	130.08 2.92	1.47 1.51
9	Betasil-C_{18}	THF	4	LSER LSST TCM	0.102 0.058 0.014	0.01036 0.00338 0.00021	64 44 38	49.54 16.15	1.47 1.51
10	POMA-ZrO_2	ACN	4	LSER LSST TCM	0.090 0.023 0.016	0.00815 0.00055 0.00026	64 44 38	31.82 2.15	1.47 1.51
11	POMA-ZrO_2	MeOH	4	LSER LSST TCM	0.108 0.015 0.010	0.01172 0.00023 0.00011	64 44 38	109.14 2.14	1.47 1.51
12	PS-ZrO_2	ACN	4	LSER LSST TCM	0.086 0.039 0.012	0.00746 0.00152 0.00015	64 44 38	49.76 10.17	1.47 1.51

Table 5. Continued

Group	Stationary phases	Mobile phases	Number of conditions	Model	SD^a	S^{2b}	df[b]	F-ratio	F_c ($\alpha=0.1$)
13	Phenyl-SiO_2	ACN	4	LSER	0.037	0.00139	64	33.28	1.47
				LSST	0.016	0.00026	44	6.32	1.51
				TCM	0.006	0.00004	38		
14	PRP-1	ACN	4	LSER	0.030	0.00088	64	8.89	1.47
				LSST	0.011	0.00013	44	1.30	1.51
				TCM	0.010	0.00010	38		

[a] SD is square root of residual mean squares.
[b] s^2 and df are residual mean squares and degree of freedom for the F-test, respectively.

calibrated quadratic equations are used to predict the retention of all the standard solutes at a new composition where retention of new solutes is to be predicted from the typical-conditions model. The retention of the standard solutes predicted from the quadratic equations is then used to calibrate the typical-conditions model for that new composition. Finally, a typical-conditions equation calibrated for that new composition is used to predict retention of new solutes at that new composition from their retention measurements at the typical conditions. If the number of new solutes whose retention is to be predicted is significantly larger than that of the standard solutes, the number of retention measurements for the calibration can be significantly reduced.

	Mobile phase type 1					Mobile phase type 2				
	TC1 QC1	New compositions	QC3	New compositions	QC2	TC2 QC1′	New compositions	QC2′	New compositions	TC3 QC3′
Standard solutes	log k	Quadratic predictions	log k	Quadratic predictions	log k	log k	Quadratic predictions	log k	Quadratic predictions	log k
New solutes	log k	Typical-conditions predictions				log k	Typical-conditions predictions			log k

We tested this combined approach to retention prediction using each of the 14 groups (Table 4). The retention at four mobile phase compositions for each solute were fitted as a quadratic function of the composition. The retention predicted from the quadratic equations for all the solutes under all the conditions in the group were then correlated with the original retention under the typical conditions for that group. The F-test was used to test if the performance of this combined approach is significantly worse than that of the simple typical-conditions model using experimental retention data only (Table 7).

The F-tests show that the performance of the combined approach is statistically better than or not different from that of the typical-conditions model used alone in every group of conditions, except for the Phenyl-SiO_2/ACN group (group 13 in Table 4). In this case, the performance of the combined approach is worse than that of the typical-conditions method used alone. However, the residual standard deviation of the combined approach for the Phenyl-SiO_2/ACN group is much lower than that of the estimated experimental uncertainty for our retention data (~0.02 in log k unit). Therefore, the lower performance of the combined approach in that group makes no practical difference.

4. Conclusions

1. A global LSER model can be derived from both

Table 6
F-test on the residual mean squares pooled from typical-conditions regressions and from quadratic regressions using the retention data from each of the 14 groups

Group	Stationary phases	Mobile phases	Number of conditions	Model	*SD*[a]	s^{2}[b]	df[b]	*F*-ratio	F_c ($\alpha = 0.1$)
1	Betasil-C_{18} POMA-ZrO_2 PS-ZrO_2 Phenyl-SiO_2 PRP-1	ACN MeOH THF	32	Quadratic TCM	0.013 0.021	0.00018 0.00044	176 390	2.43	1.18
2	Betasil-C_{18}	ACN MeOH THF	12	Quadratic TCM	0.012 0.017	0.00015 0.00030	66 136	1.98	1.33
3	Betasil-C_{18}	ACN MeOH	8	Quadratic TCM	0.006 0.018	0.00004 0.00031	44 90	7.40	1.12
4	Betasil-C_{18}	ACN THF	8	Quadratic TCM	0.014 0.017	0.00020 0.00029	44 90	1.46	1.12
5	Betasil-C_{18}	MeOH THF	8	Quadratic TCM	0.014 0.027	0.00021 0.00073	44 90	3.53	1.12
6	POMA-ZrO_2	ACN MeOH	8	Quadratic TCM	0.014 0.013	0.00019 0.00018	44 90	1.06	1.38
7	Betasil-C_{18}	ACN	4	Quadratic TCM	0.006 0.008	0.00003 0.00006	22 38	1.83	1.68
8	Betasil-C_{18}	MeOH	4	Quadratic TCM	0.007 0.007	0.00005 0.00004	22 38	1.16	1.60
9	Betasil-C_{18}	THF	4	Quadratic TCM	0.019 0.014	0.00037 0.00021	22 38	1.75	1.60
10	POMA-ZrO_2	ACN	4	Quadratic TCM	0.014 0.016	0.00019 0.00026	22 38	1.35	1.68
11	POMA-ZrO_2	MeOH	4	Quadratic TCM	0.014 0.010	0.00020 0.00011	22 38	1.83	1.60
12	PS-ZrO_2	ACN	4	Quadratic TCM	0.016 0.012	0.00026 0.00015	22 38	1.72	1.60
13	Phenyl-SiO_2	ACN	4	Quadratic TCM	0.018 0.006	0.00033 0.00004	22 38	8.01	1.60
14	PRP-1	ACN	4	Quadratic TCM	0.005 0.010	0.00003 0.00010	22 38	3.67	1.68

[a] *SD* is square root of residual mean squares.
[b] s^2 and df are residual mean squares and degree of freedom for the *F*-test, respectively.

Table 7
F-test for comparing the performance of retention prediction using the combined quadratic and TCM approach to that of the TCM approach used alone

Group	Stationary phases	Mobile phases	Number of conditions	Model	*SD*[a]	s^{2}[b]	df[b]	*F*-ratio	F_c ($\alpha = 0.1$)
1	Betasil-C_{18} POMA-ZrO_2 Phenyl-SiO_2 PRP-1	ACN THF	32	Quadratic + TCM TCM	0.020 0.021	0.00042 0.00044	480 390	1.05	1.13
2	Betasil-C_{18}	ACN MeOH THF	12	Quadratic + TCM TCM	0.016 0.017	0.00024 0.00030	204 136	1.22	1.22
3	Betasil-C_{18}	ACN MeOH	8	Quadratic + TCM TCM	0.014 0.018	0.00021 0.00031	144 90	1.51	1.27
4	Betasil-C_{18}	ACN THF	8	Quadratic + TCM TCM	0.016 0.017	0.00024 0.00029	144 90	1.22	1.27
5	Betasil-C_{18}	MeOH THF	8	Quadratic + TCM TCM	0.023 0.027	0.00052 0.00073	144 90	1.42	1.27
6	POMA-ZrO_2	ACN MeOH	8	Quadratic + TCM TCM	0.013 0.013	0.00017 0.00018	144 90	1.01	1.27
7	Betasil-C_{18}	ACN	4	Quadratic + TCM TCM	0.006 0.008	0.00004 0.00006	76 38	1.58	1.41
8	Betasil-C_{18}	MeOH	4	Quadratic + TCM TCM	0.006 0.007	0.00003 0.00004	76 38	1.28	1.41
9	Betasil-C_{18}	THF	4	Quadratic + TCM TCM	0.014 0.014	0.00020 0.00021	76 38	1.03	1.41
10	POMA-ZrO_2	ACN	4	Quadratic + TCM TCM	0.013 0.016	0.00018 0.00026	76 38	1.43	1.41
11	POMA-ZrO_2	MeOH	4	Quadratic + TCM TCM	0.010 0.010	0.00011 0.00011	76 38	1.01	1.46
12	PS-ZrO_2	ACN	4	Quadratic + TCM TCM	0.011 0.012	0.00013 0.00015	76 38	1.18	1.41
13	Phenyl-SiO_2	ACN	4	Quadratic + TCM TCM	0.011 0.006	0.00011 0.00004	76 38	2.68	1.46
14	PRP-1	ACN	4	Quadratic + TCM TCM	0.007 0.010	0.00005 0.00010	76 38	1.91	1.41

[a] *SD* is square root of residual mean squares.
[b] s^2 and df are residual mean squares and degree of freedom for the *F*-test, respectively.

local LSER and LSST models. Within the range of mobile phase compositions where the LSST model is valid, the global LSER model can be used to simultaneously model retention in RPLC as a function of both solute LSER descriptors and mobile phase composition. At most 12 coefficients are

required to establish the global LSER model. Many more coefficients would be required if the same data were modeled using the local LSER model or the LSST model, because a different LSER calibration is required for each mobile phase composition and a different LSST calibration is required for each solute. Therefore, the global LSER model requires fewer retention measurements than local LSER or LSST models for the calibration when different solutes and different mobile phase compositions are involved.

2. Even though fewer regression coefficients are used in the global LSER fit than a series of local LSER fits for the same data, the goodness-of-fit of the global LSER is as good as those of the local LSERs. However, the residuals of the LSST fits are smaller than those of both the local LSER fits and the global LSER fit. The residuals of the global LSER fit result mainly from the local LSER model and are not due to the LSST model. Thus, the performance of the global LSER model is limited by the local LSER model.

3. Due to the limitation inherited from LSST, the global LSER model can only be applied in the range of mobile phase compositions where the LSST model is valid, and the model has to be calibrated for each different organic modifier or stationary phase examined.

4. Typical-conditions model (TCM) expresses retention under a given condition as a function of retention under other typical conditions comprised of either different types or compositions of organic modifier in mobile phase or different types of stationary phase. Retention under the typical conditions are used as empirical but precise ''descriptors'' to model the retention under other conditions, which makes cross-modifier or cross-stationary phase prediction of retention possible.

5. The number of typical conditions needed in the typical-conditions model is determined by the chemical diversity of both the solutes and the conditions involved. Principal component analysis (PCA) can be used to determine the number of typical conditions needed for a given data set. Six typical conditions are needed for the retention data for 22 structurally diverse solutes under 32 different RPLC conditions comprised of five different stationary phases and three different types of mobile phases (ACN, MeOH, THF) each in four compositions. However, only two, three, and four typical conditions are needed when a single stationary phase and one, two, and three different types of mobile phases are involved, respectively.

6. Once the number of typical conditions has been determined for a given data set, iterative key set factor analysis (IKSFA) can be used to select a set of the most orthogonal conditions from all the conditions involved as the typical conditions. The retention under each of the remaining conditions in the data can then be linearly correlated with the retention under the typical conditions. The performance of the typical-conditions model is better than that of both LSER and LSST applied to the same data.

7. In principle, a typical-conditions model established for a give data set is applicable to any new condition comprised of the same stationary phase and the same mobile phase type over the same range of mobile phase compositions involved in the data. The regression coefficients for a new condition can be determined from retention measurements for a set of standard solutes under both the typical conditions and the new condition. The calibrated typical-conditions equation can then be used to predict the retention of new solutes under that new condition from their retention measurements under the typical conditions.

8. When retention under multiple mobile phase compositions is to be predicted, the number of retention measurements for the calibration can be reduced by using the retention of standard solutes predicted from a quadratic retention model to calibrate the typical-conditions model. The retention of the standard solutes at a new composition are first predicted from the retention of the same solutes at three different compositions using the quadratic model. The retention predicted from the quadratic model is then used to calibrate the typical-conditions model for that new composition. The performance of the combined quadratic–TCM approach is statistically better than or not different from that of the typical-conditions model using measured retention data.

Acknowledgements

This work is supported in part by grants from the

National Science Foundation and the National Institute of Health.

References

[1] L.R. Snyder, Anal. Chem. 72 (2000) 412A.
[2] J.G. Dorsey, W.T. Cooper, Anal. Chem. 70 (1998) 591R.
[3] J.C. Berridge, Techniques For the Automated Optimization of HPLC Separations, Wiley, New York, 1985.
[4] P.J. Schoenmakers, Optimization of Chromatographic Selectivity. A Guide To Method Development, Elsevier, Amsterdam, 1986.
[5] L.R. Snyder, J.J. Kirkland, J.L. Glajch, Practical HPLC Method Development, Wiley, New York, 1997.
[6] M.H. Abraham, M. Rosés, J. Phys. Org. Chem. 7 (1994) 672.
[7] L.C. Tan, P.W. Carr, M.H. Abraham, J. Chromatogr. A 752 (1996) 1.
[8] M.H. Abraham, M. Rosés, C.F. Poole, S.K. Poole, J. Phys. Org. Chem. 10 (1997) 358.
[9] L.C. Tan, P.W. Carr, J. Chromatogr. A 799 (1998) 1.
[10] M.J. Kamlet, M.H. Abraham, P.W. Carr, R.M. Doherty, R.W. Taft, J. Chem. Soc., Perkin Trans. II (1988) 2087.
[11] P.W. Carr, R.M. Doherty, M.J. Kamlet, R.W. Taft, W. Melander, C. Horvath, Anal. Chem. 58 (1986) 2674.
[12] P.C. Sadek, P.W. Carr, R.M. Doherty, M.J. Kamlet, R.W. Taft, M.H. Abraham, Anal. Chem. 57 (1985) 2978.
[13] D.E. Leahy, P.W. Carr, R.S. Pearlman, R.W. Taft, M.J. Kamlet, Chromatographia 21 (1986) 473.
[14] J.H. Park, P.W. Carr, M.H. Abraham, R.W. Taft, R.M. Doherty, M.J. Kamlet, Chromatographia 25 (1988) 373.
[15] C. Altomare, S. Cellamare, A. Carotti, M. Ferappi, Quant. Struct.-Act. Relatsh. 12 (1993) 261.
[16] J.H. Park, M.D. Jang, S.T. Kim, Bull. Korean Chem. Soc. 11 (1990) 297.
[17] J.H. Park, J.J. Chae, T.H. Nah, M.D. Jang, J. Chromatogr. A. 664 (1994) 149.
[18] H. Zou, Y. Zhang, P. Lu, J. Chromatogr. 522 (1990) 49.
[19] N. Chen, Y. Zhang, P. Lu, J. Chromatogr. 603 (1992) 35.
[20] N. Chen, Y. Zhang, P. Lu, J. Chromatogr. 606 (1992) 1.
[21] L.R. Snyder, P.W. Carr, S.C. Rutan, J. Chromatogr. A 656 (1993) 537.
[22] L.R. Snyder, J.W. Dolan, J.R. Gant, J. Chromatogr. 165 (1979) 3.
[23] K. Valkó, L.R. Snyder, J.L. Glajch, J. Chromatogr. A 656 (1993) 501.
[24] R. Kaliszan, K. Osmialowski, S.A. Tomellini, S.H. Hsu, S.D. Fazio, R.A. Hartwick, Chromatographia 20 (1985) 705.
[25] R. Kaliszan, K. Osmialowski, S.A. Tomellini, S.H. Hsu, S.D. Fazio, R.A. Hartwick, J. Chromatogr. 352 (1986) 141.
[26] A. Wang, L.C. Tan, P.W. Carr, J. Chromatogr. A 848 (1999) 21.
[27] R. Collander, Acta Chem. Scand. 5 (1951) 774.
[28] W.R. Melander, J. Stoveken, C. Horvath, J. Chromatogr. 199 (1980) 35.
[29] J. Zhao, P.W. Carr, Anal. Chem. 71 (1999) 2623.
[30] M. Meloun, J. Militký, M. Forina, PC-Aided Regression and Related Methods, Chemometrics For Analytical Chemistry, Vol. 2, Ellis Horwood, New York, 1994.
[31] C.H. Lochmüller, S.J. Breiner, C.E. Reese, M.N. Koel, Anal. Chem. 61 (1989) 367.
[32] C.H. Lochmüller, C.E. Reese, S.H. Hsu, Anal. Chem. 66 (1994) 3806.
[33] C.H. Lochmüller, C.E. Reese, S.H. Hsu, J. Chromatogr. Sci. 33 (1995) 640.
[34] E.R. Malinowski, Factor Analysis in Chemistry, 2nd ed, Wiley, New York, 1991.
[35] Y. Mao, Ph.D. Thesis, University of Minnesota, 2001.
[36] Y. Hu, Ph.D. Thesis, University of Minnesota, 1998.
[37] J. Zhao, Ph.D. Thesis, University of Minnesota, 1999.
[38] J. Zhao, P.W. Carr, Anal. Chem. 70 (1998) 3619.
[39] A.J.P. Martin, Biochem. Soc. Symp. 3 (1950) 4.
[40] E. Heftman (Ed.), Chromatography: Fundamentals and Applications of Chromatography and Related Differential Migration Methods, Elsevier, New York, 1992.
[41] R. Kaliszan, Structure and Retention in Chromatography: A Chemometric Approach, Harwood Academic, Amsterdam, 1997.
[42] R.B. Cattell, Multivariate Behav. Res. 1 (1966) 245.
[43] J.C. McGowan, J. Chem. Technol. Biotechnol. 34A (1984) 38.
[44] M.H. Abraham, J.C. McGowan, Chromatographia 23 (1987) 243.
[45] M.H. Abraham, Chem. Soc. Rev. 22 (1993) 73.