ELSEVIER

Journal of Chromatography A, 965 (2002) 301–314

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Comparison of methods for characterization of reversed-phase liquid chromatographic selectivity

Lawrence A. Lopez, Sarah C. Rutan*

Department of Chemistry, Virginia Commonwealth University, 1001 West Main Street, Richmond, VA 23284-2006, USA

Abstract

The goal of the present work is to obtain a better understanding of the chemical factors affecting liquid chromatographic retention. One of the most commonly used formats for liquid chromatographic separations is based on a nonpolar stationary phase, typically an octadecyl-derivatized silica material. A wide variety of these reversed-phase columns are commercially available that differ significantly in their chromatographic retention and selectivity. We seek to quantitatively characterize these differences. Retention data for a range of compounds with many diverse characteristics have been measured on several different octadecyl silica columns (*J. Chromatogr. A*, submitted for publication). Principal components analysis is used to characterize the different properties of these stationary phases and predict retention factors. The key set factor analysis method and the typical solute method are used in conjunction with the principal components analysis to identify small subsets of solutes that can be used to quantitatively describe the retention of a broad range of compounds. In addition, a quantitative comparison to alternative data analysis methods is made, including linear solvation energy relationships and an iterative subtraction method based on linear regression techniques. Although many earlier studies have reported the application of these methods, this study is the first to make a quantitative comparison of these methods using a highly precise and structurally variable set of test compounds.

Keywords: Principal component analysis; Linear free energy relationships; Iterative subtraction; Selectivity; Linear solvation relationships

1. Introduction

Reversed-phase liquid chromatography (RPLC), based on a nonpolar alkyl phase (usually an octadecyl silica) and an aqueous–organic mobile phase, is used for a wide range of practical analytical separations. There are several potential chemical interactions that can occur between the C_{18} phase and the solutes, including dispersion, dipole–dipole, dipole–induced dipole, H bonding (van der Waals interactions), ion–ion, ion–dipole, ion–induced dipole (electrostatic interactions), and steric interactions. Our goal is to encode these interactions in a logical way to allow for quantitative prediction of retention and selectivity. A convenient way to explain the contribution of these different interactions to retention is by using a linear free energy formalism, which is represented by the following equation:

$$\log k_{i,j} = \log k_{\mathrm{ref},j} + \sum_{k=1}^{n} z_{i,k} \cdot Z_{j,k} \quad (1)$$

Here $k_{i,j}$ is the retention factor for the ith compound under the jth set of conditions, $z_{i,k}$ is the kth ''polarity'' parameter for the ith solute, $Z_{j,k}$ is the

*Corresponding author. Tel.: +1-804-828-7517; fax: +1-804-828-8599.

E-mail address: srutan@saturn.vcu.edu (S.C. Rutan).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)00002-X

complementary ''polarity'' parameter for the *j*th chromatographic condition, and log $k_{\mathrm{ref},j}$ is the log *k* for a reference compound. As an example, $z_{i,3}$ might represent the hydrogen bond basicity of the *i*th solute and $Z_{j,3}$ would then represent the complementary property, the hydrogen bond acidity of the stationary phase. While Eq. (1) can be applied to study variations in either mobile phase [2,3] or stationary phase [4–6] conditions, in this work we have focused on the characterization of stationary phase variations. Ideally, a small set of parameters could allow for quantitative prediction of chromatographic retention factors and selectivities.

1.1. Linear solvation energy relationships

Several different approaches have been described in the literature to find sets of linear parameters than span the variance in chromatographic data. One of these methods, based on linear solvation energy relationships (LSERs), has been described by Carr and co-workers [2,4,7,8], Abraham and co-workers [9–11], and Poole et al. [12,13]. In these studies, retention is represented by the following equation:

$$\log k = \underset{\text{(i)}}{\log k_0} + \underset{\text{(ii)}}{rR_2} + \underset{\text{(iii)}}{s\pi_2} + \underset{\text{(iv)}}{a\sum\alpha_{\mathrm{H}^2}} + \underset{\text{(v)}}{b\sum\beta_{\mathrm{H}^2}} + \underset{\text{(vi)}}{vV_x} \tag{2}$$

Term (i) is a solute independent constant (the intercept) that includes the phase ratio. This term varies with the stationary and mobile phase conditions. Terms (ii)–(vi) account for the intermolecular interactions between the solute and the mobile and stationary phases. The subscripted symbols in Eq. (2) (R_2, π_2, etc.) represent solute properties, which have been measured for a large number of simple compounds [9]. The rR_2 term represents dispersion interactions, the $s\pi_2$ term represents hydrogen-bonding interactions and the vV_x term represents ''hydrophobic'' interactions that are sometimes related to the energy of cavity formation. The $a\Sigma\alpha_2^{\mathrm{H}}$ and $b\Sigma\beta_2^{\mathrm{H}}$ terms represent hydrogen-bond interactions. The solute parameters for thousands of solutes have been compiled [9], such that the system dependent coefficients (log k_0, *r*, *s*, *a*, *b* and *v*) are easily determined via multiple linear regression. Studies have focused on mobile phase properties [2,10] and stationary phase properties [4]. Efforts have been made to develop global descriptions that include both mobile phase and stationary phase contributions [3]. One limitation of the LSER approach is the characterized solutes are usually much simpler structurally than the broad, complex array of pharmacologically active compounds that are idea targets for quantitative prediction of retention and selectivity. While some studies have included some complex pharmaceuticals [5,6], the fit quality of these regressions is generally worse when multi-function drug molecules are included in the fit.

1.2. Principal components analysis

Another approach that has been used to characterize retention and selectivity patterns in RPLC is principal components analysis (PCA) [14–19]. This analysis, like LSER analysis, is based on the linear format of Eq. (1). In the case of PCA however, the $z_{i,k}$ and $Z_{j,k}$ parameters are determined by a purely mathematical approach, which seeks orthogonal parameter scales (scales with the lowest possible cross-correlations). In the PCA literature the $z_{i,k}$ and $Z_{j,k}$ parameters are referred to as principal components (PCs).

Here we represent this relationship as follows:

$$\log k_{i,j} = \log k_{\mathrm{ref},j} + \sum_{k=1}^{n} \mathrm{SP}_{i,k} \cdot \mathrm{CP}_{j,k} \tag{3}$$

where log $k_{\mathrm{ref},j}$ is again the log *k* value of the reference compound. The SPs (solute properties) represent the solute PCs and the CPs (column properties) represent the stationary phase PCs.

The singular value decomposition (SVD) algorithm is frequently used in PCA to decompose the data matrix into the corresponding linear parameters. SVD gives stable results for a wide range of applications. In SVD notation the data set is described by the following equation:

$$\mathbf{D} = \mathbf{USV}^T \tag{4}$$

The rows of matrix **D** contain the log $k_{i,j} - \log k_{\mathrm{ref},j}$ values for each individual solute in the data set and the columns of **D** represent the different chromatographic stationary phases. The matrix **S** contains a diagonal arrangement of values ($s_{k,k}$) that indicate

the variance contributions from each principal component. A large singular value represents a large factor, which has physical significance, but may have contributions from several different interaction types. A small singular value contains small, unimportant factors attributed to noise. The columns of **U** represent the variations in retention as a function of the solute. The columns of **V** describe the variations in retention between the different columns. These are abstract orthonormal factors that describe the variations in the entire data set as a function of column behavior.

The following equations are used to relate the SVD results to the solute and chromatographic stationary phase properties. The linear parameters dependent on the SPs are defined in Eq. (5):

$$\mathrm{SP}_i = U_i \tag{5}$$

and the linear parameters dependent on the CPs are given in Eq. (6):

$$\mathrm{CP}_j = S_{jj}V_j \tag{6}$$

A more detailed explanation of the application of PCA to chemical problems can be found in Malinowski's work [20].

1.3. Key factor method

The key solute method is based on the identification of one of the best subsets of actual solutes whose behavior reflects the retention of all the compounds in the data set. In essence, each set of retention factors for a given column serves as indicators of the different interactions driving the retention process. In Eq. (7), $\log k_{\mathrm{ref},j}$ is the log of the retention factor for the reference solute on the *j*th column. The index *k* represents the *n* different key solutes, each representing a distinctly different combination of molecular interactions:

$$\log k_{i,j} = \log k_{\mathrm{ref},j} + \sum_{k=1}^{n} z_{j,k} \cdot \log k_{i,k} \tag{7}$$

The results from the PCA can be used to identify sets of key solutes, according to the iterative key set factor analysis (IKSFA) procedure published by Malinowski et al. [21,22]. This approach considers sets of solutes together as predictors of retention. Wang and Carr have explored the used of the key set method as applied to the selection of the most typical of a set of chromatographic conditions. This approach, called the typical condition method (TCM), was used to identify the minimal set of chromatographic conditions to precisely predict the retention factors for the remaining conditions [23]. The approach used in this paper is the typical solute method, where the minimal set of solutes, selected from the entire data set, is identified with the key set method, and the data for this subset of solutes enables prediction of the retention factors of the remaining solutes. This variation of IKSFA is the typical solute method (TSM).

1.4. Test solute method

One method that has been widely used for characterization of chromatographic selectivity is the test solute method. This method consists of the selection of a small set of solutes selected specifically to focus on the different properties of the stationary phases. Some of the test solute methods that have been developed include the shape selective selectivity test of Sander and Wise (benzo[*a*]pyrene, phenanthrophenanthrene, and tetrabenzonaphthalene) [24–26], the Engelhardt test (toluene, ethylbenzene, phenol, toluidine, *N*,*N*-dimethylaniline, aniline, and ethylbenzoate) [27,28], and the Neue selectivity chart (acenaphthene and amitriptyline) [29–31]. While these scales are useful for describing variations in behavior of different stationary phases, they are not normally used for quantitative retention predictions.

1.5. Iterative subtraction method

The iterative subtraction method (ISM) attempts to find a form of Eq. (1) based on the empirical linear-free-energy equation given by Eq. (8) [1]:

$$\log k_{i,j} = \log k_{\mathrm{ref},j} + \eta_i'\mathbf{H}_j + \sigma_i'\mathbf{S}_j + \beta_i'\mathbf{A}_j + \alpha_i'\mathbf{B}_j + \kappa_i'\mathbf{C}_j \tag{8}$$

The parameters η', σ', β', α', and κ' are properties of the solute molecules. The parameters **H**, **S**, **A**, **B**, and **C** account for the complementary properties of the column, mobile phase, and temperature. The quantity $k_{\mathrm{ref},j}$ represents retention of the refer-

ence compound on the jth column. The inclusion of the log $k_{\mathrm{ref},j}$ term is to intended to correct for differences in the column phase ratio. The $\eta_i'\mathbf{H}_j$, $\sigma_i'\mathbf{S}_j$, $\beta_i'\mathbf{A}_j$, $\alpha_i'\mathbf{B}_j$, and $\kappa_i'\mathbf{C}_j$ terms describe different solute–column interactions that affect retention because, in this case, the mobile phase conditions are held constant. The parameters are determined by grouping solutes into chemically similar sets, and defining the parameters from the average log k values of each set. The details of this procedure are given in the Experimental section. The terms can be approximately related to physicochemical interactions as follows. ''Hydrophobicity'' is represented by the term $\eta'\mathbf{H}$, and depends roughly on the size of the solute and the cohesivity of the stationary phase. Shape selectivity is described by the term $\sigma'\mathbf{S}$ and includes the effects from steric hinderance. In this case, steric hindrance decreases solute retention. The $\beta'\mathbf{A}$ term describes hydrogen bonding interactions between hydrogen bond acceptor solutes and non-ionized silanols groups. The interactions between hydrogen bond donor solutes and hydrogen bond acceptor groups in the stationary phase are represented by the term $\alpha'\mathbf{B}$. The $\kappa'\mathbf{C}$ term describes the ion-exchange interactions of protonated bases with ionized silanol groups.

The goal of the present paper is to quantify the similarities and differences among the approaches mentioned above for the quantitative prediction of retention in RPLC. A key aspect of this study is the inclusion of an extensive, well-characterized data set with precise and accurate retention values for both simple solutes as well as complex multifunctional drug molecules [1].

2. Experimental

The data that are investigated in this paper are the results from measurements of retention for 67 solutes on 10 different C_{18} columns in a 50% acetonitrile mobile phase. The primary data are provided in Ref. [1]. The solutes are listed in Table 1 and the stationary phase materials are described in Table 2. The retention factors for solutes 1–45 were measured using an unbuffered mobile phase, while the retention factors for solutes 46–67 were obtained with the aqueous component of the mobile phase buffered at pH 2.80. At pH 2.80, the strong base compounds (solutes 46–50) are ionized. This adds potential complexity to the models required to fit the data, but this pH was selected in part due to the fact that well-behaved (i.e., nontailing) peaks are observed for the strong bases for the most commonly used reversed-phase columns at this pH. The details for the acquisition of this data set are given by Wilson et al. [1].

LSER analyses were carried with the regression

Table 1
Test solutes used in the present study (Ref. [1])

A. Neutral solutes			B. Basic solutes	C. Acidic solutes (weak acids)
1. Benzene	16. *N*-Benzylformamide	31. Acetophenone	B.1. Strong bases	56. Diclofenate acid
2. Toluene	17. Anisole	32. Benzophenone	46. Amitriptyline	57. Mefenamic acid
3. Ethylbenzene	18. Benzyl alcohol	33. *cis*-Chalcone	47. Diphenhydramine	58. Ketoprofen
4. *p*-Xylene	19. 3-Phenylpropanol	34. *trans*-Chalcone	48. D,L-Propanolol	59. Diflunisal
5. Propylbenzene	20. 5-Phenylpentanol	35. *cis*-4-Nitrochalcone	49. Nortriptyline	60. 4-*n*-Butylbenzoic acid
6. Butylbenzene	21. Phenol	36. *trans*-4-Nitrochalcone	50. Prolintane	61. 4-*n*-Pentylbenzoic acid
7. Naphthalene	22. *p*-Chlorophenol	37. *cis*-4-Methoxychalcone	B.2. Weak bases	62. 4-*n*-Hexylbenzoic acid
8. *p*-chlorotoluene	23. 2,3-Dihydroxynaphthalene	38. *trans*-4-Methoxychalcone	51. 4-Pentylaniline	63. 3-Cyanobenzoic acid
9. *p*-Dichlorobenzene	24. 1,3-Dihydroxynaphthalene	39. Prednisone	52. 4-Hexylaniline	64. 2-Nitrobenzoic acid
10. Benzotrichloride	25. Eugenol	40. Hydrocortisone	53. 4-Heptylaniline	65. 3-Nitrobenzoic acid
11. Bromobenzene	26. Danthron	41. Mephenytoin	54. *N*-Ethylaniline	66. 2,6-Dimethylbenzoic acid
12. 1-Nitropropane	27. *n*-Propylformate	42. Oxazepam	55. 2-Phenylpyridine	67. 2-Fluorobenzoic acid
13. Nitrobenzene	28. Methylbenzoate	43. Flunitrazepam		
14. *p*-Nitrotoluene	29. Benzonitrile	44. 5,5-Diphenylhydantoin		
15. *p*-Nitrobenzylchloride	30. Coumarin	45. *N*,*N*-Dimethylacetamide		

Table 2
Characteristics of C_{18} columns used in present study; 5 μm particles, 150×4.6 mm (Ref. [1])

Column	Abbreviation[a]	Surface area (m^2/g)	Pore diameter (nm)	% C	μmol/m^2
1. GL Inertsil ODS-3	Inertsil	420–450	9.4–10.2	14.3	1.8
2. Waters Symmetry C_{18}	Symmetry	343	9	19.7	3.13
3. HP Zorbax SB C_{18}	SB-100%	186	8	10.4	2.08
4. HP Zorbax SB C_{18}	SB-90%	188	8	9.20	1.79
5. HP Zorbax SB-300 C_{18}	SB-300	52	30	3.25	2.09
6. HP Eclipse XDB-C_{18}	Eclipse	186	8	10.7	3.0
7. YMC Pack Pro C_{18}	YMC 15	322	12.5	15.5	2.51
8. YMC Pack Pro C_{18}	YMC 16	321	12.5	16.3	2.68
9. YMC Pack Pro C_{18}	YMC 17	322	12.5	17.0	2.82
10. Supelco Discovery C_{18}	Discovery	190–220	17–20	12.5	3.12

[a] Shorthand designation of each column.

tool in EXCEL, version 97 (Microsoft, Redmond, WA, USA). PCA analysis was carried out in MATLAB, version 6.0 (Mathworks, South Natick, MA, USA). The iterative key set factor analysis method was implemented in MATLAB according to the algorithm given by Schostack and Malinowski [21]. The ISM technique (Eq. (8)) was implemented using EXCEL spreadsheets using the following procedure [1].

(1) For each individual column, the log $\alpha_{i,j} = \log(k_{i,j}/k_{\mathrm{ref},j})$ for each solute is calculated, where $k_{\mathrm{ref},j}$ is the retention of the reference compound, ethylbenzene, on the jth column.

(2) The values of log $\alpha_{i,j}$ for all columns vs. log $\alpha_{i,1}$, where 1 represents the reference column (SB-100), are plotted, and the linear regression parameters are calculated (intercept forced to 0). Ideal solutes are identified as those with a standard error of <0.010 in log k units.

(3) Using these ideal solutes, the linear regression parameters for the plots of log $\alpha_{i,j}$ vs. log $\alpha_{i,1}$ are obtained, normalizing the slope (**H**) found for the SB-100 column to one.

(4) Values for $\eta'_{i,j} = (\log\ \alpha_{i,j})/\mathbf{H}_j$ are calculated for each solute and column separately. The values of $\eta'_{\mathrm{ave},i}$ for each solute are averaged and the standard deviations of these averages are calculated.

(5) Deviations of $\Delta_{i,j} = \log\ \alpha_{i,j} - (\eta_{\mathrm{ave},i}\mathbf{H}_j)$ are calculated for each solute and column. Depending on the solute structure and the similarities in retention behavior, solutes are assigned to a specific group, each group corresponding to a term in Eq. (8) ($\sigma'\mathbf{S}$, $\beta'\mathbf{A}$, $\alpha'\mathbf{B}$, and $\kappa'\mathbf{C}$).

(6) The values of the column parameters are calculated from the values of Δ: the term **A** represents the value Δ for the solute 45 (*N*,*N*-dimethylacetamide); the term **C** is the average value of Δ for solutes 46–50 (ionized bases); the term **B** is the average value of Δ for solutes 56–58, 60–65 (acids); the term **S** is the average value of Δ for solutes 32–40, and 43–44.

(7) Linear regression fits (forcing the intercept to 0) of the log $\alpha_{i,j}$ values vs. the column parameters **H**, **S**, **A**, **B** and **C** provides coefficients which correspond to the solute parameters η', σ', β', α' and κ'.

(8) A second set of linear regression parameters are determined by fitting the values of log $\alpha_{i,j}$ vs. the solute parameters obtained from step 7, and resulting coefficients are the final column parameters **H**, **S**, **A**, **B** and **C**.

3. Results and discussion

There is a wide range of commercially available C_{18} columns currently on the market. Due to efforts by the manufacturers, the lot-to-lot variability of these materials has decreased tremendously, such that replication of selectivity and retention are possible to a good level of precision [32,33]. The fundamental approach for comparing column be-

havior is to compare their selectivities. In fact, many stationary phase characterization methods are based on the evaluation of selectivities for selected pairs of solutes. The selectivity scale of Neue and co-workers is a good example [29,30]. Another method for evaluation column similarity is to plot log $\alpha_{i,j}$ for one column versus another column. The more linear the plot, the more similar the retention mechanism is between the two columns. Fig. 1 shows this relationship for the SB-90 column relative to the reference column (SB-100). It can be seen that because the column chemistry of these two stationary phases is similar (differing only by the extent of reaction with silyl ligands), the log $\alpha_{i,j}$ values are highly correlated. Fig. 2 shows this relationship for two columns with different selectivity characteristics (SB-100 and Inertsil), and it can be seen that the correlation is much lower. The six compounds significantly displaced from the perfect correlation line are the weakly retained, basic compounds (46–50), amitriptyline, diphenhydramine, propanolol, nortryptyline and prolintane, and (45) *N*,*N*-dimethylacetamide.

The initial technique used to characterize this data set was PCA. PCA was applied to the data after subtracting the log *k* of the reference compound, ethyl benzene. The retention of ethylbenzene for a given column is greater than the average retention of all solutes on that column. However, the PCA results obtained from this approach are essentially the same

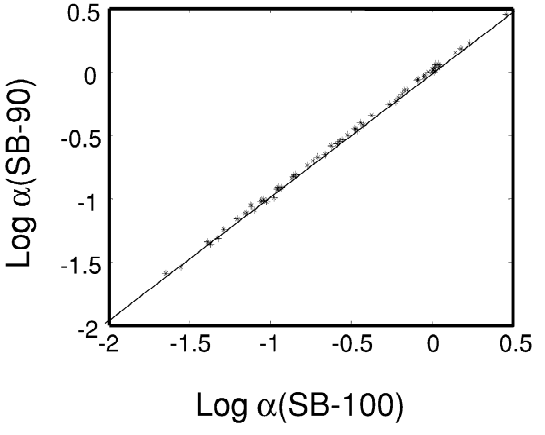

Fig. 1. Plot of the log of the selectivity of the solutes relative to ethylbenzene (log $\alpha_{\mathrm{ref},j}$) on the SB-90 column relative to the SB-100 column.

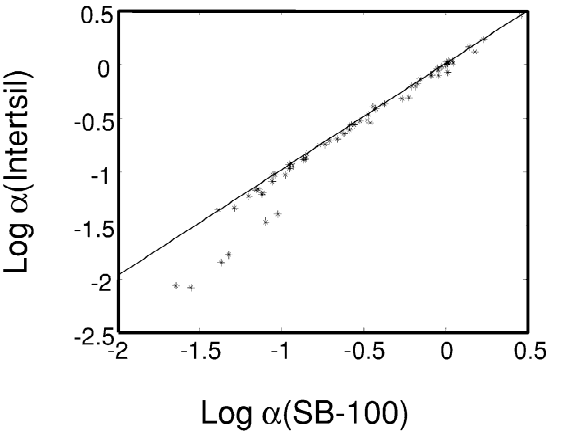

Fig. 2. Plot of the log of the selectivity of the solutes relative to ethylbenzene (log $\alpha_{\mathrm{ref},j}$) on the Intersil column relative to the SB-100 column.

as those obtained using mean-centering of the data. A plot of SP_2 vs. SP_1 (the values of the first two solute parameters obtained with PCA) is shown in Fig. 3. The percentages given in the axis labels are the percent of the total data set variance described by the corresponding property. Hence, 99.67% of the variation in the data can be described by SP_1; 0.24% of the variation in the data can be described by SP_2 for a total of 99.91%. The clustering of the points is also indicative of the retention patterns. The trends in the SP_1 parameter are almost completely described

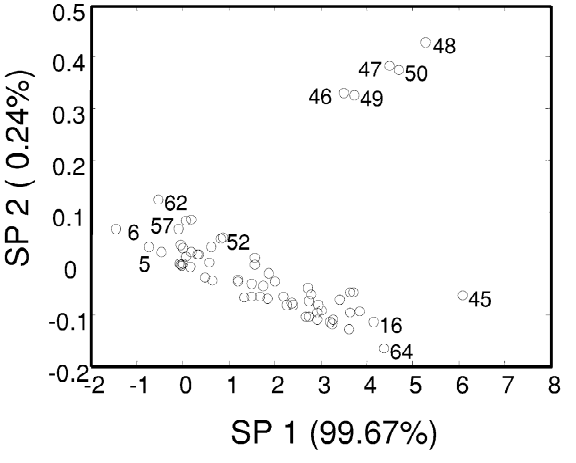

Fig. 3. PCA results. Plot of the second PC describing the solute retention variations (SP_2) vs. the first solute PC (SP_1). Selected solutes are defined in Table 1.

by the average retention of each solute on all 10 columns. The group of points in the upper right hand corner of the plot corresponds to solutes 46–50, and this pattern mimics the deviations from linear behavior shown in Fig. 2. These compounds are all bases that should be protonated under these chromatographic conditions and are presumed to be retained by an ion-exchange mechanism with the residual ionized silanols on the stationary phase. A plot of SP_4 vs. SP_3 is shown in Fig. 4. From this plot it can be seen that these solute parameters are determined largely by the properties of solute 45 (*N*,*N*-dimethylacetamide) and solute 59 (diflunisal, an acidic drug). *N*,*N*-Dimethylacetamide is weakly retained ($k=0.022$), while diflusinal shows moderate retention ($k=0.364$). An alternate way of examining the results from a PCA analysis is to plot the selected PC vs. the sample (solute) number. Fig. 5 shows a plot of the SP_5 parameter vs. solute number. The bulky chalcones (solutes 35–38) have positive SP_5 values. In addition, a trend towards decreasing value of SP_5 with increasing homologue number (i.e., the *n*-alkylbenzoic acids, 60–62) can be seen. This behavior indicates that SP_5 may be reflecting shape-selective contributions to retention, with the more globular chalcones giving positive SP_5 values, and the more linear alkylbenzoic acids giving negative SP_5 values.

A model containing five PCs (i.e., the solute parameters SP_1–SP_5, along with the corresponding column parameters, CP_1–CP_5) accounts for 99.999% of all the variation in the data set and allows for prediction of log *k* values to ±0.0042. Wilson et al. [1] have estimated that the experimental precision of their data is ±0.002. Another way to evaluate the relative importance of these principal components is to plot the square root of the variance of the information in each component (the diagonal elements of matrix **S** in Eq. (4)) vs. the number of the component. This is normally done with a logarithmic scale on the *y*-axis, as shown in Fig. 6. The variances associated with components 6–10 are small, and of

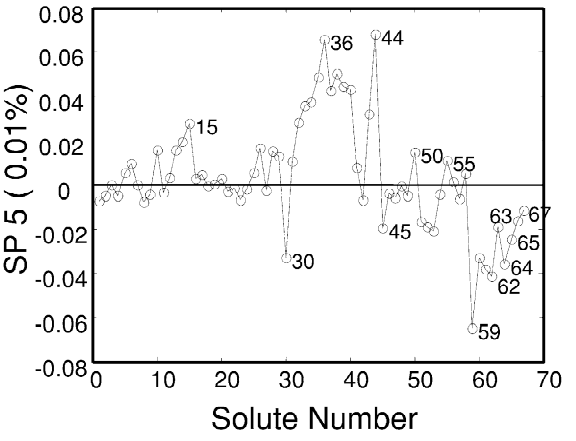

Fig. 5. PCA results. Plot of the fifth PC (SP_5) vs. solute number. Selected solutes are defined in Table 1.

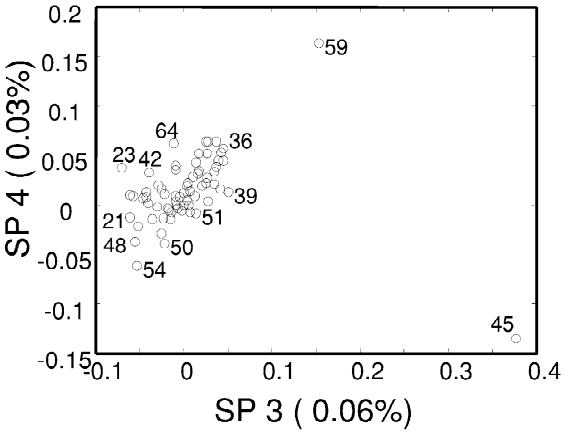

Fig. 4. PCA results. Plot of the fourth PC describing the solute retention variations (SP_4) vs. the first third PC (SP_3). Selected solutes are defined in Table 1.

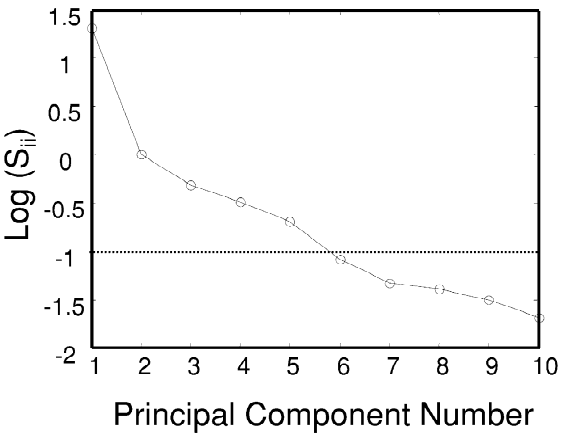

Fig. 6. PCA results. Plot of the log of the diagonal elements of the **S** matrix (Eq. (4)) vs. component number.

the same order of magnitude, indicating that these components are probably primarily associated with noise.

A quantitative comparison of the PCA results to the scales derived by the ISM represented by Eq. (8) can be made by evaluating the η', σ', β', α' and κ' parameter scales in terms of their relationship with the PCA determined SPs. Table 3 shows the coefficients for SP_1–SP_5 required to predict the η', σ', β', α' and κ' parameters from the ISM. In addition, the mean and standard deviation of the parameter scales are shown. It can be seen that η' is primarily determined by SP_1 and σ' is primarily determined by SP_5. The relationships between the remaining parameters and the SP values are much weaker, with β' being partially related to SP_3, α' correlating with both SP_4 and SP_5, and κ' showing a contribution from SP_2. Although the individual scales from the PCA and ISM models are not the same, the overall prediction capabilities of the two models are virtually identical, with the prediction error of the η', σ', β', α' and κ' of the ISM model is 0.0043 log k units and the prediction error of the PCA model is 0.0042 log k units. However, chemical rationalization of the parameter scales is easier in the case of the ISM, as these scales were developed with target solutes of known physicochemical properties (i.e., hydrogen bond acids, hydrogen bond bases, and ionized bases), rather than the purely mathematical scales resulting from the PCA method.

A further comparison of the PCA analysis can be made to the column selectivity chart developed by Neue and co-workers [29–31]. Neue's chart is a plot of the log of the selectivity of amitryptiline and acenaphthene vs. the log of the retention factor of acenaphthene, and is reproduced in Fig. 7. This plot can be seen to be related to a plot of the column principal components, $-CP_2$ vs. CP_1, shown in Fig. 8. The relationships between columns 1 (Intersil OD3), 2 (Symmetry C_{18}) and 3 (Zorbax SB18) is virtually the same as that seen on the Neue plot. In addition, the Eclipse XDB C_{18} phase (6) and the YMC PackPro C_{18} phase (7–9) are found in very similar positions on both plots, although the ''hydrophobicity'' order of these materials is reversed.

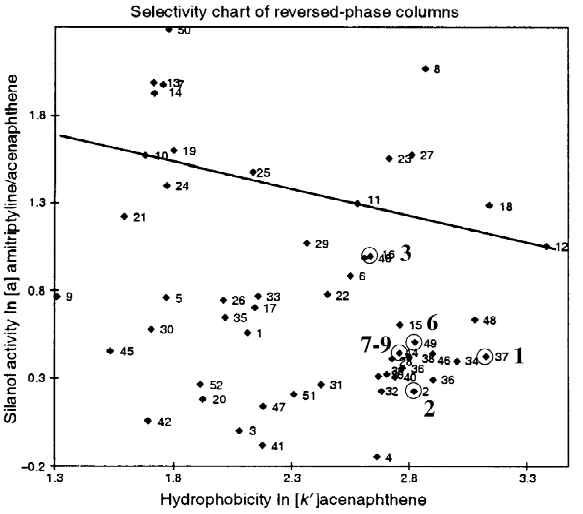

Fig. 7. Neue plot. Column numbers of interest here are 37 (Intersil ODS-3), 2 (Symmetry C_{18}), 16 (Zorbax SB-C_{18}), 49 (Zorbax XDB C_{18}), (44) YMC Pack Pro C_{18}. Large numbers correspond to column numbers given in Table 2. Permission to reprint figure from Ref. [29] by Wiley (pending).

In a similar fashion, the column parameters from the ISM can also be compared to the Neue chart and the PCA results. The analogous plot of **C** vs. **H** is

Table 3
Regression equation coefficients for the η', σ', β', α' and κ' parameters to the principal components SP_1–SP_5

	SP_1	SP_2	SP_3	SP_4	SP_5	Mean	Standard deviation
η'	**0.3161**	−0.0254	0.0149	−0.0195	−0.0015	−0.58	0.53
σ'	−0.0248	−0.0151	2.2166	−4.9289	**−16.1203**	0.18	0.43
β'	−0.0177	0.0800	1.8035	1.3825	1.1487	0.02	0.13
α'	−0.0645	0.2329	**2.5171**	**−11.8735**	**12.1283**	0.27	0.54
κ'	−0.0514	**−1.9732**	−1.0347	1.5373	−0.4617	0.07	0.28
% Total variance	99.67	0.24	0.06	0.03	0.01		

Values in bold indicate which of the ISM parameters are most important in relationship to the PCA SP factors.

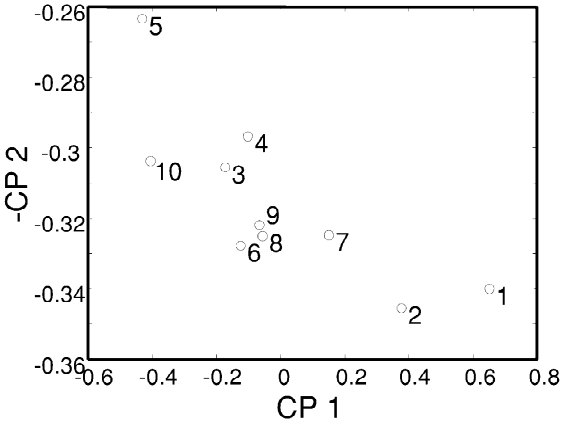

Fig. 8. PCA results. Plot of the second PC describing the column retention variations ($-CP_2$) vs. the first column PC (CP_1). Column numbers are given in Table 2.

shown in Fig. 9. There are some differences in the local arrangement of columns between Figs. 8 and 9, indicating that the relative contributions of the other parameters on hydrophobic and ion-exchange interactions are expressed differently in the two parameter scales.

For either the ISM or the PCA methods, it would be desirable to identify a small subset of the solutes that can be used reliably to characterize the column parameters for newly designed stationary phases. One approach that has been proposed for this purpose is IKSFA [21,22]. Wang and Carr have used

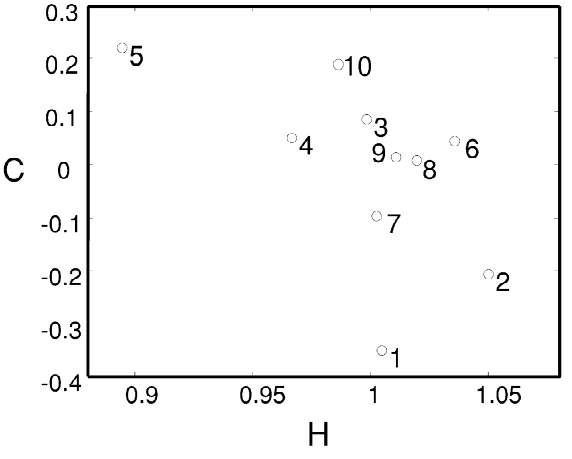

Fig. 9. ISM results. Plot of **C** column parameter vs. **H** column parameter. Column numbers are given in Table 2.

this method to identify selected *chromatographic conditions* that allow for accurate representation of the retention factors for all studied conditions, and called their approach the TCM [23]. In our case, we employ the complementary approach, and use IKSFA to identify the subset of *solutes* that allow for the prediction of the retention factors of the entire data set: the TSM. The results for a five-member key set of solutes are as follows: *p*-chlorotoluene (neutral aromatic), danthron (polar compound), *trans*-4-nitrochalcone (bulky compound), 4-heptylaniline (basic compound), and mefanamic acid (acidic compound). This set of solutes predicts the entire data set with a standard error of 0.088. A more precise prediction is obtained when seeking a seven member key set of solutes, which includes *p*-chlorotoluene, danthron, *trans*-4-methoxychalcone, 4-heptylaniline, mefanamic acid, ethylbenzene (nonpolar reference compound) and amitriptyline (strong base). (The substitution of *trans*-4-nitrochalcone for *trans*-4-methoxychalone is probably not significant). The standard error of prediction based on these compounds is 0.038. Although the error decreases by more than a factor of 2, the fit quality is not at the level need to provide reliable predictions of selectivity. There is very little difference in the prediction of the data compared to the prediction of the column parameters. Clearly, one member of the strong base group (compounds 46–50) is required to model retention, as can be seen from the important deviations of these compounds in the PC plot of the two largest PCs shown in Fig. 3. Schostack and Malinowski mentioned in their report that the IKSFA worked well for data sets containing pure variables [21]. In this case, a pure variable would be a compound whose retention is determined by a single linear factor. This is clearly not the case for the present application, and this may be the reason that the IKSFA algorithm fails to include a strong base compound in the five-component case. This algorithm can also be applied directly to the SPs from PCA or the ISM η', σ', β', α' and κ' parameters. In these cases, the prediction precision of the data by the selected solutes improves to 0.0096 and 0.0066, respectively. This improvement in fit is probably due to the fact that PCA and ISM parameters are more nearly ''pure'' variables. A chemically intuitive set of compounds [acetophenone (neutral aromatic), *trans*-

4-nitrochalcone (bulky compound), *N*,*N*-dimethylacetamide (sterically hindered hydrogen bond base), amitriptyline (strong base), and 2-nitrobenzoic acid (acid compound)], gives a standard error of 0.0045 in predicting the **H**, **S**, **A**, **B**, and **C** parameters, and a standard error of 0.0053 in predicting the overall retention, thus providing the most precise prediction. A summary of these different key solute sets and their corresponding predictive abilities is given in Table 4.

The PCA and ISM results can be compared to the LSER results for this database on a subset of selected compounds. Wang and Carr have recommended 22 compounds that serve as a good basis set for determining the LSER coefficients in Eq. (2) [23]. This set is specifically chosen to include only those compounds with simple structures that result in precise estimation of the LSER coefficients. Significantly more variegated compound sets have also been explored, but the overall fit quality is significantly worse in these cases [5,6]. These compounds are among the 67 solutes whose retention properties have been measured here; these solutes and their corresponding solute parameters are listed in Table 5. These solutes are all neutral, relatively small compounds, and thus these LSERs only describe a limited part of the retention behavior, relative to the methods based on the entire data set. The resulting coefficients calculated by multiple linear regression of the retention data and solute parameters are given in Table 6. As the standard errors of the coefficients were all between 0.04 and

Table 4
Key solute sets, and their standard errors of prediction

Solute selection method	Source data	Standard error	Solutes
TSM	Retention data	0.088	*p*-Chlorotoluene Danthron *trans*-4-Nitrochalcone 4-Heptylaniline Mefanamic acid
TSM	Retention data	0.038	*p*-Chlorotoluene Danthron *trans*-4-Methoxychalcone 4-Heptylaniline Mefanamic acid Ethylbenzene Amitriptyline
IKSFA	η', σ', β', α', κ'	0.0096	Bromobenzene *p*-Nitrobenzylchloride *N*,*N*-Dimethylacetamide Amitriptyline *N*-Ethylaniline
IKSFA	SP_1–SP_5	0.0066	5,5-Diphenylhydantoin *N*,*N*-Dimethylacetamide Diphenhydramine *N*-Ethylaniline 3-Nitrobenzoic acid
Chemical intuition	Retention data	0.0053	Acetophenone *trans*-4-Nitrochalcone *N*,*N*-Dimethylacetamide Amitriptyline 2-Nitrobenzoic acid

Table 5
Solvatochromic parameters for 22 solutes used to determine LSER fits with Eq. (2)

Solute	V_x	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	R_2
N-Benzylformamide	1.1137	1.80	0.40	0.63	0.990
Benzyl alcohol	0.9160	0.87	0.33	0.56	0.803
Phenol	0.7751	0.89	0.60	0.30	0.805
3-Phenylpropanol	1.1978	0.90	0.30	0.67	0.821
p-Chlorophenol	0.8975	1.08	0.67	0.20	0.915
Acetopheonone	1.0139	1.01	0	0.48	0.818
Benzonitrile	0.8711	1.11	0	0.33	0.742
Nitrobenzene	0.8906	1.11	0	0.28	0.871
Methylbenzoate	1.0726	0.85	0	0.46	0.773
Anisole	0.9160	0.75	0	0.29	0.708
Benzene	0.7164	0.52	0	0.14	0.610
p-Nitrotoluene	1.0315	1.11	0	0.28	0.870
p-Nitrobenzylchloride	1.1539	1.34	0	0.40	1.080
Toluene	0.8573	0.52	0	0.14	0.601
Benzophenone	1.4808	1.50	0	0.50	1.447
Bromobenzene	0.8914	0.73	0	0.09	0.882
Naphthalene	1.0854	0.92	0	0.20	1.340
Ethylbenzene	0.9982	0.51	0	0.15	0.613
p-Xylene	0.9982	0.52	0	0.16	0.613
p-Dichlorobenzene	0.9612	0.75	0	0.02	0.825
Propylbenzene	1.1391	0.50	0	0.15	0.604
n-Butylbenzene	1.2800	0.51	0	0.15	0.600

Parameters are from Ref. [3].

0.06, these values have been omitted from the table for clarity of the presentation. In addition, it should be noted that the correlations have been carried out using the log α values referenced to ethylbenzene. The standard errors for the LSER fits are all in the range from 0.03 to 0.04, which is consistent with the LSER fits reported in the literature. Upon examination of the coefficients, it can be seen that the rR_2 is not significantly different from zero on all of the columns. The coefficient with the most variability from column to column is the intercept, while the remaining terms are very similar from column to column. The most dissimilar columns as indicated by the intercept are columns 1 (Intertsil ODS-3) and column 5 (Zorbax SB-300 C_{18}). These columns are also on the extremes in Figs. 7 and 9, representing the retention parameters described the PCA-based CPs and the ISM-based **H** and **C** parameters. The Inertsil ODS-3 is a highly retentive column, and the Zorbax SB-300 C_{18} is a wide-pore material with inherently lower overall retention. For RPLC LSERs, the ''hydrophobic'' (vV_x) and hydrogen bond base ($b\Sigma\beta_2^H$) terms are generally the most significant; this is also seen in the present study. The maximum correlations of the ISM **H**, **S**, **A**, **B**, and **C** parameters with the LSER coefficients *s*, *r*, *a*, *b*, *v* and log k_0 were examined. The best correlation is an inverse relationship between the **H** parameter and the *b* coefficient, as shown in Fig. 10. The supposed characteristics represented by these scales are the cohesiveness of the stationary phase (**H**) and the hydrogen bond acidity of the stationary phase (*b*). Both of these scales are strongly correlated with the overall retention, so the correlation between them is not completely unexpected. Column 2 (Symmetry)

Table 6
Coefficients for the LSER fits to Eq. (2)

Column	Log k_0	*r*	*s*	*a*	*b*	*v*	Standard error	R^2
Inertsil	−0.07	0.05	−0.35	−0.52	−1.78	1.58	0.035	0.9955
Symmetry	−0.17	0.05	−0.39	−0.51	−1.82	1.63	0.033	0.9964
SB-100%	−0.27	0.02	−0.33	−0.52	−1.73	1.61	0.032	0.9962
SB-90%	−0.27	0.02	−0.31	−0.51	−1.67	1.57	0.032	0.9959
SB-300	−0.74	0.02	−0.29	−0.45	−1.58	1.49	0.030	0.9957
Eclipse	−0.26	0.03	−0.36	−0.52	−1.80	1.65	0.035	0.9959
YMC 15	−0.23	0.04	−0.36	−0.50	−1.77	1.61	0.036	0.9954
YMC 16	−0.25	0.04	−0.36	−0.50	−1.79	1.63	0.035	0.9957
YMC 17	−0.23	0.05	−0.36	−0.50	−1.78	1.62	0.034	0.9957
Discovery	−0.48	0.04	−0.36	−0.46	−1.75	1.58	0.032	0.9960
Average	−0.30	0.04	−0.35	−0.50	−1.75	1.60	0.03	0.9959
Standard deviation	0.19	0.01	0.03	0.03	0.07	0.05	0.0017	0.0003

Standard deviations of the coefficients are all between 0.04 to 0.06.

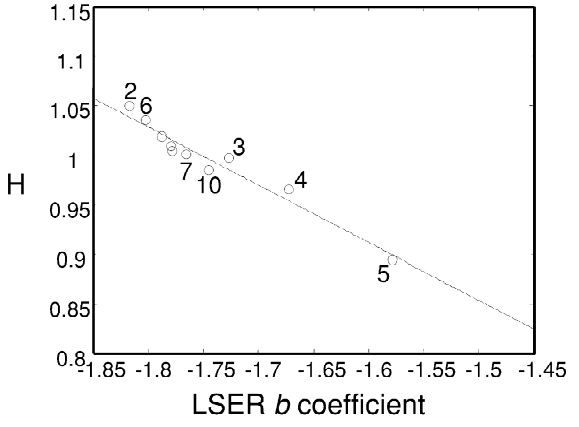

Fig. 10. Comparison of ISM and LSER methods. Plot of **H** parameter (ISM) vs. LSER *b* coefficient.

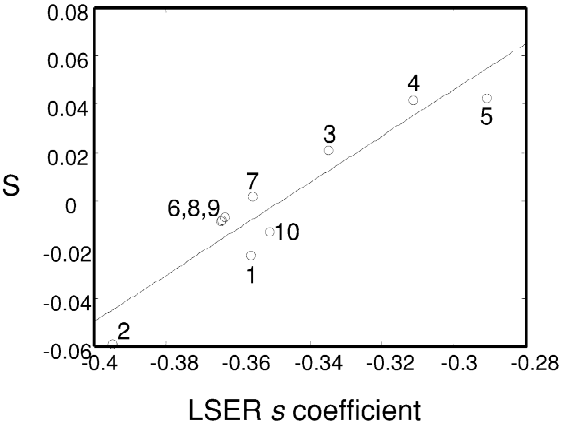

Fig. 12. Comparison of ISM and LSER methods. Plot of **S** parameter (ISM) vs. LSER *s* coefficient.

shows up as having the highest **H**, **S** and **B** parameters and the lowest *b* and *s* coefficients (see Figs. 10–12). This column shows up on the Neue selectivity chart as having a very low amount of silanol activity, as indicated by the low relative retention of amitriptyline. The **H** parameter is also highly correlated with the *v* coefficient in the LSER correlation, as shown in Fig. 11. These scales are both intended to model dispersion interactions between the stationary and the solute. The correlation between the **S** scale (steric effects) and the LSER *s* coefficient (dipolar interactions), shown in Fig. 12, likely indicates that neither the LSER coefficients or the ISM scales reflect a complete partitioning of the underlying physical interactions. The only other correlation of appreciable magnitude ($r^2 = 0.919$) is between the LSER log k_0 coefficient (the intercept) and the **B** parameter, as shown in Fig. 13. The only ISM parameters not partially explained by the LSER coefficients are **A** and **C**. This is entirely reasonable as the **A** parameter is mostly determined by the behavior of *N*,*N*-dimethylacetamide, which is not a member of the 22 solute subset used for the LSER fits, and the **C** parameter is an indicator of ion pairing interactions of protonated bases, which also are not present in the LSER solute subset. It is important to note that the standard errors of the

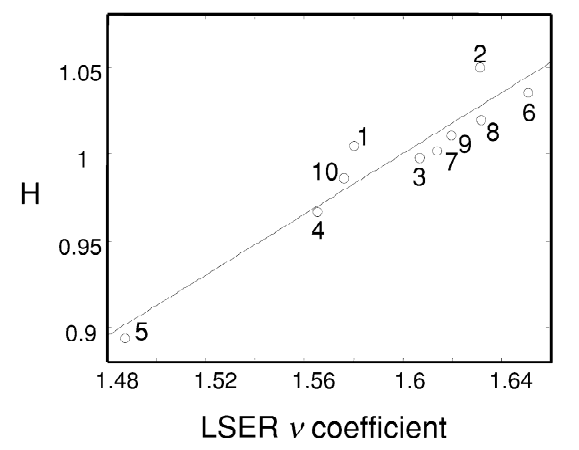

Fig. 11. Comparison of ISM and LSER methods. Plot of **H** parameter (ISM) vs. LSER *v* coefficient.

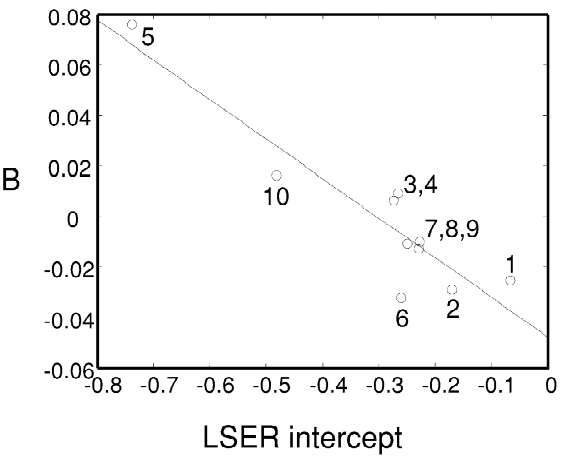

Fig. 13. Comparison of ISM and LSER methods. Plot of **B** parameter (ISM) vs. LSER intercept.

coefficients do not indicate the level of reliability shown by these correlations with the ISM parameters. It should be noted that several solutes are consistent outliers in all 10 of the LSER fits in the same direction and with the same magnitude (acetophenone, *p*-nitrobenzylchloride, *p*-xylene, and *p*-dichlorobenzene), indicating that the standard errors of the LSER regressions are not accurate representations of the error in the data—they represent errors in the model.

4. Conclusions

This comparison of various methods for rationalizing and predicting chromatographic retention lends insight into the separation mechanism. Chemically intuitive approaches, such as the ISM and LSERs offer complementary insights to the mathematical results from PCA. Even though the LSER approach is generally restricted to less complex molecules relative to the PCA and ISM methods, significant correlation between the scales defined by each of these approaches can be seen. Methods for selection of key solutes gave drastically different solute sets and prediction errors. The best predictions were obtained from the intuitively selected solute set (standard error 0.0055) and for the IKSFA method applied directly to the solute principal components (standard error of 0.0066). It appears that the IKSFA method only works well when the source data contain largely ''pure'' variables, in this case, parameters affect by one type of retention interaction. It should be noted that all the analyses reported in this paper are all internal to a single dataset—no true validation of the predictions is provided here. However, in the original evaluation of this data set [1], and in several follow-up papers [34,35], the valid application of these data analyses for other, independent data sets is clearly demonstrated. Either of the PCA or ISM scales demonstrated here should serve as a reliable basis for further quantitative characterization of reversed-phase column selectivities.

Acknowledgements

Support for this work from the National Science Foundation (NSF-007654290) and LC Resources is gratefully acknowledged. L.R. Snyder, E. Bezemer, A. Wang, and P.W. Carr are also thanked for many helpful comments on the manuscript.

References

[1] N.S. Wilson, M.D. Nelson, J.W. Dolan, L.R. Snyder, R.G. Wolcott, P.W. Carr, J. Chromatogr. A, submitted for publication.
[2] L.C. Tan, P.W. Carr, J. Chromatogr. A 799 (1998) 1.
[3] A.S. Wang, L.C. Tan, P.W. Carr, J. Chromatogr. A 848 (1999) 21.
[4] L.C. Tan, P.W. Carr, M. Abraham, J. Chromatogr. A 752 (1996) 1.
[5] K. Valko, M. Plass, C. Bevan, D. Reynolds, M.H. Abraham, J. Chromatogr. A 797 (1998) 41.
[6] K. Valko, C.M. Du, C.D. Bevan, D.P. Reynolds, M.H. Abraham, J. Pharm. Sci. 89 (2000) 1085.
[7] P.W. Carr, Microchem. J. 48 (1993) 4.
[8] M. Reta, P.W. Carr, P.C. Sadek, S.C. Rutan, Anal. Chem. 71 (1999) 3484.
[9] M.H. Abraham, Chem. Soc. Rev. 22 (1993).
[10] M.H. Abraham, M. Roses, J. Phys. Org. Chem. 7 (1994) 672.
[11] M. Abraham, M. Roses, C.F. Poole, S.K. Poole, J. Phys. Org. Chem. 10 (1997) 358.
[12] M. Roses, D. Bolliet, C.F. Poole, J. Chromatogr. A 829 (1998) 29.
[13] D. Bolliet, C.F. Poole, M. Roses, Anal. Chim. Acta 368 (1998) 129.
[14] A. Bolck, A.K. Smilde, in: R.M. Smith (Ed.), Retention and Selectivity in Liquid Chromatography, Elsevier, Amsterdam, 1995, Chapter 12.
[15] M.F. Delaney, A.N. Papas, M.J. Walters, J. Chromatogr. 410 (1987) 31.
[16] C.H. Lochmuller, S.J. Breiner, C.E. Reese, M.N. Koel, Anal. Chem. 61 (1989) 367.
[17] D.V. McCalley, R.G. Brereton, J. Chromatogr. A 828 (1998) 407.
[18] B. Walczak, M. Dreux, J.R. Chretien, K. Szymoniak, M. Lafosse, L. Morin-Allory, J.P. Doucet, J. Chromatogr. 353 (1986) 109.
[19] B. Walczak, J.R. Chretien, M. Dreux, L. Morin-Allory, M. Lafosse, K. Szymoniak, F. Membrey, J. Chromatogr. 353 (1986) 123.
[20] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991.
[21] K.J. Schostack, E.R. Malinowski, Chemometr. Intell. Lab. Syst. 6 (1989) 21.
[22] E.R. Malinowski, Anal. Chim. Acta 134 (1982) 129.
[23] A. Wang, P.W. Carr, J. Chromatogr. A 965 (2002) 3.
[24] L.C. Sander, S.A. Wise, J. Chromatogr. A 656 (1993) 335.
[25] L.C. Sander, S.A. Wise, in: R.M. Smith (Ed.), Retention and

Selectivity in Liquid Chromatography, Elsevier, Amsterdam, 1995, p. 337.
[26] L.C. Sander, S.A. Wise, Anal. Chem. 67 (1995) 3284.
[27] H. Engelhardt, M. Jungheim, Chromatographia 29 (1990) 59.
[28] H. Engelhardt, M. Arangio, T. Lobert, LC–GC Int. 15 (1997) 803.
[29] U.D. Neue, E. Serowik, P. Iraneta, B.A. Alden, T.H. Walter, J. Chromatogr. A 849 (1999) 87.
[30] U.D. Neue, B.A. Alden, T.H. Walter, J. Chromatogr. A 849 (1999) 101.
[31] U.D. Neue, in: R.A. Meyers (Ed.), Encylopedia of Analytical Chemistry, Wiley, Chichester, 2000, p. 11450.
[32] M. Kele, G. Guiochon, J. Chromatogr. A 830 (1999) 41.
[33] M. Kele, G. Guiochon, J. Chromatogr. A 830 (1999) 55.
[34] N.S. Wilson, J.W. Dolan, L.R. Snyder, P.W. Carr, L.C. Sander, J. Chromatogr. A, submitted for publication.
[35] N.S. Wilson, M.D. Nelson, J.W. Dolan, L.R. Snyder, P.W. Carr, J. Chromatogr. A, submitted for publication.