ELSEVIER

Journal of Chromatography A, 886 (2000) 245–257

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Statistical evaluation of linear solvation energy relationship models used to characterize chemical selectivity in micellar electrokinetic chromatography

Mark D. Trone[1], Morteza G. Khaledi*

North Carolina State University, Raleigh, NC 27695-8204, USA

Received 2 August 1999; received in revised form 28 March 2000; accepted 28 March 2000

Abstract

Characterization of retention and selectivity differences between surfactants in micellar electrokinetic chromatography (MEKC) using linear solvation energy relationships (LSERs) has been given a significant amount of attention in the last four years. This report evaluates the validity of using the two LSER models that have been used to fit retention in MEKC in the literature. The results and the fit of the revised model and parameters developed by Abraham and coworkers are compared to the original model developed by Kamlet, Taft, and coworkers. LSERs can generally only be used as a comparative tool to describe the selectivity differences between surfactant systems used in MEKC. With this in mind, it was determined that the results of both models essentially provide the same information about these differences. However, the revised model and parameters have been found to yield a statistically better fit of the MEKC retention data as well as providing more chemically sound LSER coefficients.

Keywords: Linear solvation energy relationships; Micellar electrokinetic chromatography; Statistical analysis; Surfactants

1. Introduction

In the last several years, a good deal of attention has been given to the characterization of selectivity in micellar electrokinetic chromatography (MEKC) [1–7]. A significant amount of this work has used linear solvation energy relationships (LSERs). The LSER model was first developed by Kamlet and Taft et al. to describe solvation effects on physicochemical processes [8–11]. The descriptors in this model were later adapted to describe solute characteristics in order to investigate the solubility properties of various media [12–14]. The first MEKC retention and selectivity studies used the original form of the Kamlet and Taft LSER model (Eq. (1)) [1].

$$\log k' = c + mV_1 + s\pi^* + b\beta + a\alpha \quad (1)$$

In this equation, the logarithm of the solute retention factor in MEKC is correlated to known solute descriptors. The terms representing the solute descriptors are V_1, π^*, and β and α. The solute polarity and polarizability are represented by the π^* term.

*Corresponding author. Tel.: +1-919-515-2943; fax: +1-919-515-5079.

E-mail address: morteza_khaledi@ncsu.edu (M.G. Khaledi)

[1]Present address: Merck & Co., Inc., Rahway, NJ 07065, USA.

0021-9673/00/$ – see front matter
PII: S0021-9673(00)00416-7

The solute hydrogen bond accepting and solute hydrogen bond donating abilities are described by β and α, respectively. V_1 represents the intrinsic volume of the solute and is divided by 100 to bring it to scale with the other terms. The coefficients of these descriptors (*m*, *s*, *b*, and *a*) are related to the difference in contribution between the pseudo-stationary phase and the bulk aqueous phase with respect to each type of interaction. The cohesive and dispersive nature of the micellar phase is related to *m*. The difference in dipolarity/polarizability between the micelles and the bulk aqueous phase is represented by *s*. The *a* and *b* terms describe the relative ability of the micellar phase to form hydrogen bond interactions with solute molecules relative to the bulk phase, where *b* represents the hydrogen bond donating ability and *a* represents the hydrogen bond accepting ability. The constant, *c*, contains information about the system that is not accounted for by the other LSER parameters. When Eq. (1) is used to characterize selectivity in MEKC, the main contributor to the system constant is the separation phase ratio.

Abraham et al. later revised the LSER model and solute parameters in an attempt to improve correlation between retention in gas and liquid chromatography. Because the original solute descriptor values (Eq. (1)) were estimated and extrapolated from bulk solvent properties, they derived more thermodynamically sound solute descriptor values [15–18]. The model itself (Eq. (2)) is very similar to the original version by Kamlet and Taft and coworkers.

$$\log k' = c + mV_X + s\pi_A^* + b\Sigma\beta_A + a\Sigma\alpha_A + rR_A \qquad (2)$$

However, in this model, V_x represents the solute characteristic volume in (cm^3 mol^{-1}) which is divided by 100 in order to bring it to scale with the other descriptors [19]. Another modification addresses the solute polarity. The new rR_A term is the solute's excess molar refraction, and is divided by 10 to obtain a rough scaling with the other parameters [17]. In this manuscript, the subscript of '*A*' simply denotes that these symbols represent the solute descriptor values modified by Abraham and coworkers. The *m*, *b*, and *a* coefficients for the revised model contain the same information as discussed above. The *s* coefficient is related to the dipolarity/polarizability of the micellar phase, and *r* coefficient represents the ability of the micelle to interact with the *n*- and π-electrons of the solute. The rR_A term has also been referred to as the polarization/polarizability correction term for the LSER model [20]. In their papers, Poole and Poole used this revised form of the LSER model (Eq. (2)) to describe solute retention in MEKC [5,6,21,22]. Abraham et al. [23], Quina et al. [24], and Vitha et al. [25], have also used this model to describe solute partitioning in micellar solutions.

Since the reports from this group have primarily involved using the older Kamlet and Taft model and solute descriptors (Eq. (1)), we have been interested in determining reliability of each model. Poole et al. have recently addressed some issues important to LSER analyses of MEKC data [21]. Most notably, they have discussed the importance of providing a suitable number of solutes as well as being careful to avoid cross-correlation between the descriptors for the solutes used in the test set. They have also recommended the use of generic experimental conditions in order to standardize the analysis of selectivity in MEKC surfactant systems. This paper extends that discussion as well as including a few topics not addressed by Poole et al. The different models and the different sets of values for the solute descriptors have been evaluated and compared using both old and new MEKC data obtained in this laboratory. Although some of the systems have been studied previously, they have been re-evaluated in this report using different models to aid in comparisons and to allow for a more in-depth evaluation.

2. Experimental

For this report, the surfactants originally used by Yang and Khaledi [1,2], a mixed surfactant system [7], and two new surfactant systems [26,27] have been evaluated. The original surfactant solutions include 40 m*M* sodium dodecyl sulfate (SDS), 40 m*M* lithium perfluorooctanesulphonate (LiPFOS), 10 m*M* tetradecyltrimethylammonium bromide (TTAB), and 60 m*M* sodium cholate (SC). The mixed surfactant solution consisted of 30 m*M* SC–30 m*M* SDS in

a 50 m*M* phosphate buffer solution (pH 7). These MEKC runs were performed in a 50 m*M* phosphate buffer (pH 7) at 25°C. Sodium *N*-parmitoyl sarcosinate (SPN) and (*S*)-dodecoxycarbonylvaline were also investigated. These surfactants were used in a 10 m*M* phosphate buffer (pH 7) also at 25°C. The other experimental conditions used for all of these systems have been described previously [1–3,7,27].

3. Results and discussion

3.1. Correlation of test solutes

The test solutes and their descriptors used are given in Tables 1–4. In the original publications, 60 solutes were used in the test set [1]. However, six of these solutes were consistent outliers using the Kamlet and Taft et al. LSER model (Eq. (1)). Therefore, a test set of 54 of the original solutes have been used to re-evaluate the LSER model and solute descriptors used in the original publications. In addition, the descriptor values for some of these solutes have not been modified using Abraham et al.'s procedures (Eq. (2)). Therefore, a reduced test set of 36 solutes has been used to evaluate the Abraham's model.

As mentioned previously, Poole et al. have discussed the importance of orthogonality in the solute descriptors [21]. Descriptors that are unintentionally correlated with one another can significantly bias the model making it an ineffective analysis tool. However, one possibility that Poole did not discuss is that one descriptor may be linearly related to two or more of the other parameters. The simple cross-correlation results for descriptors with respect to one another are also presented (Tables 2 and 4). As can be seen, pairwise descriptor correlation is not a problem in either set. Analysis also shows that multiple correlation of one parameter against the remaining terms is also not a problem (not shown). Therefore, the parameters in both models are adequately orthogonal to be used in the model without fear of biasing.

3.2. LSER results

The LSER constants for all of the surfactant

Table 1
Test solutes and their solvatochromic parameters based on Kamlet and Taft's original LSER model (Eq. (1))

Solutes	V_i	π^*	β	α
1 Benzene	0.491	0.59	0.10	0
2 Toluene	0.592	0.55	0.11	0
3 Ethyl benzene	0.668	0.53	0.12	0
4 Propylbenzene	0.769	0.51	0.12	0
5 *p*-Xylene	0.668	0.51	0.12	0
6 Acetophenone	0.690	0.90	0.49	0.04
7 Propiophenone	0.788	0.88	0.49	0
8 Butyrophenone	0.886	0.86	0.49	0
9 Valerophenone	0.984	0.84	0.49	0
10 Benzonitrile	0.590	0.90	0.37	0
11 Nitrobenzene	0.631	1.01	0.30	0
12 Anisole	0.639	0.73	0.32	0
13 Ethoxybenzene	0.727	0.69	0.30	0
14 Methyl benzoate	0.736	0.75	0.39	0
15 Ethyl benzoate	0.834	0.74	0.41	0
16 Chlorobenzene	0.581	0.71	0.07	0
17 Bromobenzene	0.624	0.79	0.06	0
18 Iodobenzene	0.671	0.81	0.05	0
19 4-Dichlorobenzene	0.671	0.70	0.03	0
20 2-Dichlorobenzene	0.671	0.80	0.03	0
21 2-Chloronitrobenzene	0.721	1.11	0.26	0
22 4-Chloronitrobenzene	0.721	1.01	0.26	0
23 4-Chlorotoluene	0.679	0.67	0.08	0
24 4-Chloroanisole	0.720	0.73	0.22	0
25 4-Bromonitrobenzene	0.764	1.01	0.26	0
26 4-Nitrotoluene	0.729	0.97	0.31	0
27 4-Chloroacetophenone	0.780	0.90	0.45	0.06
28 Methyl 2-methylbenzoate	0.834	0.71	0.40	0
29 Phenyl acetate	0.736	1.14	0.52	0
30 Phenol	0.536	0.72	0.33	0.61
31 4-Methylphenol	0.634	0.68	0.34	0.58
32 4-Ethylphenol	0.732	0.66	0.35	0.58
33 4-Fluorophenol	0.562	0.73	0.28	0.65
34 4-Chlorophenol	0.626	0.72	0.23	0.67
35 4-Bromophenol	0.669	0.79	0.23	0.67
36 Benzyl alcohol	0.634	0.99	0.52	0.39
37 4-Methylbenzyl alcohol	0.732	0.93	0.53	0.39
38 4-Chlorobenzyl alcohol	0.724	1.11	0.42	0.40
39 Aniline	0.562	0.73	0.50	0.26
40 *N*-Ethylaniline	0.758	0.82	0.47	0.17
41 4-Chloroaniline	0.653	0.73	0.40	0.31
42 Naphthalene	0.753	0.70	0.15	0
43 1-Methylnaphthalene	0.851	0.66	0.16	0
44 2-Methylnaphthalene	0.851	0.66	0.16	0
45 Biphenyl	0.920	1.18	0.20	0
46 3-Chlorophenol	0.626	0.77	0.23	0.69
47 3-Methylphenol	0.634	0.68	0.34	0.58
48 2- Methylphenol	0.634	0.68	0.34	0.54
49 3-Bromophenol	0.669	0.84	0.23	0.69
50 3-Methylbenzyl alcohol	0.732	0.95	0.53	0.39
51 3-Chlorobenzyl alcohol	0.724	1.11	0.42	0.40
52 Phenethyl alcohol	0.732	0.97	0.55	0.33
53 3-Phenyl-1-propanol	0.830	0.95	0.55	0.33
54 3,5-Dimethyiphenol	0.732	0.64	0.35	0.56

Table 2
Simple correlation matrix for Kamlet, Taft et al. solute parameters (R^2)

	V_1	π^*	β	α
V_1	1.00			
π^*	0.08	1.00		
β	0.09	0.23	1.00	
α	0.11	0.00	0.07	1.00

systems using Eq. (1) and the original solute parameter values are listed in Table 5. The older model and parameter values yield an acceptable fit to the experimental data. The adjusted coefficient of determination (R^2_{adj}), the predicted retention vs. the measured retention plots (Fig. 1) and the residual plots (not presented) all show that the model provides a reasonable fit and that the linear model of Eq.

Table 3
Test solutes and their solvation descriptors for the revised LSER model (Eq. (2))[a]

Solutes	V_x	π^*_A	$\Sigma\beta_A$	$\Sigma\alpha_A$	R_A
1 Benzene	0.716	0.52	0.14	0.00	0.610
2 Toluene	0.857	0.52	0.14	0.00	0.601
3 Ethylbenzene	0.998	0.51	0.15	0.00	0.613
4 Propylbenzene	1.139	0.50	0.15	0.00	0.604
5 *p*-Xylene	0.998	0.52	0.16	0.00	0.613
6 Acetophenone	1.014	1.01	0.48	0.00	0.818
7 Benzonitrile	0.871	1.11	0.33	0.00	0.742
8 Nitrobenzene	0.891	1.11	0.28	0.00	0.871
9 Methyl benzoate	1.073	0.85	0.46	0.00	0.733
10 Ethyl benzoate	1.214	0.85	0.46	0.00	0.689
11 Chlorobenzene	0.839	0.65	0.07	0.00	0.718
12 Bromobenzene	0.891	0.73	0.09	0.00	0.882
13 Iodobenzene	0.975	0.82	0.12	0.00	1.188
14 4-Chlorotoluene	0.980	0.67	0.07	0.00	0.705
15 4-Chloroanisole	1.038	0.86	0.24	0.00	0.838
16 4-Citrotoluene	1.032	1.11	0.28	0.00	0.870
17 4-Chloroacetophenone	1.136	1.09	0.44	0.00	0.955
18 Methyl 2- methylbenzoate	1.214	0.87	0.43	0.00	0.772
19 Phenylacetate	1.073	1.13	0.54	0.00	0.661
20 Phenol	0.775	0.89	0.30	0.60	0.805
21 4-Methylphenol	0.916	0.87	0.31	0.57	0.820
22 4-Ethylphenol	1.057	0.90	0.36	0.55	0.800
23 4-Fluorophenol	0.793	0.97	0.23	0.63	0.670
24 4-Chlorophenol	0.898	1.08	0.20	0.67	0.915
25 4-Bromophenol	0.950	1.17	0.20	0.67	1.080
26 Benzyl alcohol	0.916	0.87	0.56	0.33	0.803
27 4-Chloroaniline	0.939	1.13	0.31	0.30	1.060
28 Naphthalene	1.085	0.92	0.20	0.00	1.360
29 1-Methylnaphthalene	1.226	0.90	0.20	0.00	1.344
30 Biphenyl	1.324	0.99	0.22	0.00	1.360
31 3-Chlorophenol	0.898	1.06	0.15	0.69	0.909
32 3-Methylphenol	0.916	0.88	0.34	0.57	0.822
33 3-Bromophenol	0.950	1.15	0.16	0.70	1.060
34 3-Methylbenzyl alcohol	1.057	0.90	0.59	0.33	0.815
35 Phenethyl alcohol	1.057	0.83	0.66	0.30	0.784
36 3,5-Dimethylphenol	1.057	0.84	0.36	0.57	0.820

[a] Solute descriptors from Ref. [25].

Table 4
Simple correlation matrix for Abraham et al.'s solute parameters (R^2)

	V_X	π_A^*	$\Sigma\beta_A$	$\Sigma\alpha_A$	R_A
V_X	1.00				
π_A^*	0.01	1.00			
$\Sigma\beta_A$	0.10	0.12	1.00		
$\Sigma\alpha_A$	0.13	0.14	0.00	1.00	
R_A	0.13	0.24	0.02	0.01	1.00

(1) is adequate for MEKC retention data. As reported previously, the LSER *m* and *b* coefficients in Table 5 show that solute size and hydrogen bond accepting ability play the most important roles in MEKC retention. In this model, the more positive (or less negative) the coefficient, the more favorable that solute characteristic is for interacting with a micellar phase. The original LSER model (Eq. (1)) shows that LiPFOS and SDS have the strongest interactions with hydrogen bond accepting solutes, and TTAB and SC have the weakest (*b*). The large and positive *m* coefficient shows that all of the surfactants form pseudo phases that are quite organic like, and the *s* coefficient shows that all of the micelles have similar polarity/polarizability. The smaller *m* coefficient for LiPFOS suggests that it forms micelles that are more structured and much less organic-like than the other surfactants. Finally, Table 5 shows that TTAB micelles have the strongest and LiPFOS micelles have the weakest interactions with hydrogen bond donating solutes, respectively.

Table 6 shows that using the revised model and solute parameters (Eq. (1)) has a dramatic effect on the values of the LSER coefficients. This is especially true when comparing the micellar polarities (*s*). Unlike the results from the original model, using Eq. (2) suggests that the surfactants form micellar phases with significantly different polarities. The polarization/polarizability correction term (rR_A) also shows that the surfactants possess a wide range of ability to interact with the solute *n*- and π-electrons. However, the trends observed for micelle cavity formation and hydrogen bonding remain consistent with those in Table 5. For the sake of comparison, the LSER results using the same reduced set of test solutes and the original solute parameters (from Tables 1 and 2) are listed in Table 7. It is clear from the coefficient of determination and the standard errors of the coefficients that using a reduced set of solutes severely limits the effectiveness of the LSER model when the older solute parameters are used.

Although the comparative results between the two models are similar, it is worth comparing the fit that each model provides. The standard error values, the R^2_{adj}, and the predicted vs. measured retention plots (Figs. 1 and 2) all suggest that the revised model (Eq. (2)) provide a better fit for the MEKC data even though a smaller set of solutes is used. The smaller residual scale using Eq. (2) relative to that using Eq. (1) also indicates that this model fits the data more accurately than Eq. (1). In addition to the statistical evidence, the LSER results in Table 6 are also more chemically sound. Using Eq. (1) suggests that the hydrogen bond donating ability of LiPFOS and the

Table 5
LSER coefficients using the original solute parameters and the original Kamlet, Taft et al. model (Eq. (1)) ($n=54$)[a]

Surfactant system	*c*	*m*	*s*	*b*	*a*	SE	R^2_{adj}	SSE
40 m*M*	−1.62	4.19	−0.28	−1.99	−0.12	0.09	0.968	0.37
SDS		(0.14)	(0.08)	(0.10)	(0.12)			
40 m*M*	−1.70	2.77	−0.28	0.05	−0.90	0.13	0.909	0.78
LiPFOS		(0.16)	(0.12)	(0.14)	(0.08)			
10 m*M*	−1.83	4.11	−0.29	−2.90	0.96	0.11	0.941	0.64
TTAB		(0.19)	(0.11)	(0.13)	(0.07)			
60 m*M*	−1.63	3.96	−0.28	−2.99	0.19	0.11	0.952	0.62
SC		(0.19)	(0.11)	(0.13)	(0.07)			
30/30 m*M*	−1.49	3.83	−0.20	−2.65	0.20	0.11	0.944	0.60
SC/SDS		(0.18)	(0.11)	(0.12)	(0.07)			

[a] Numbers in parentheses are the standard errors of the coefficients.

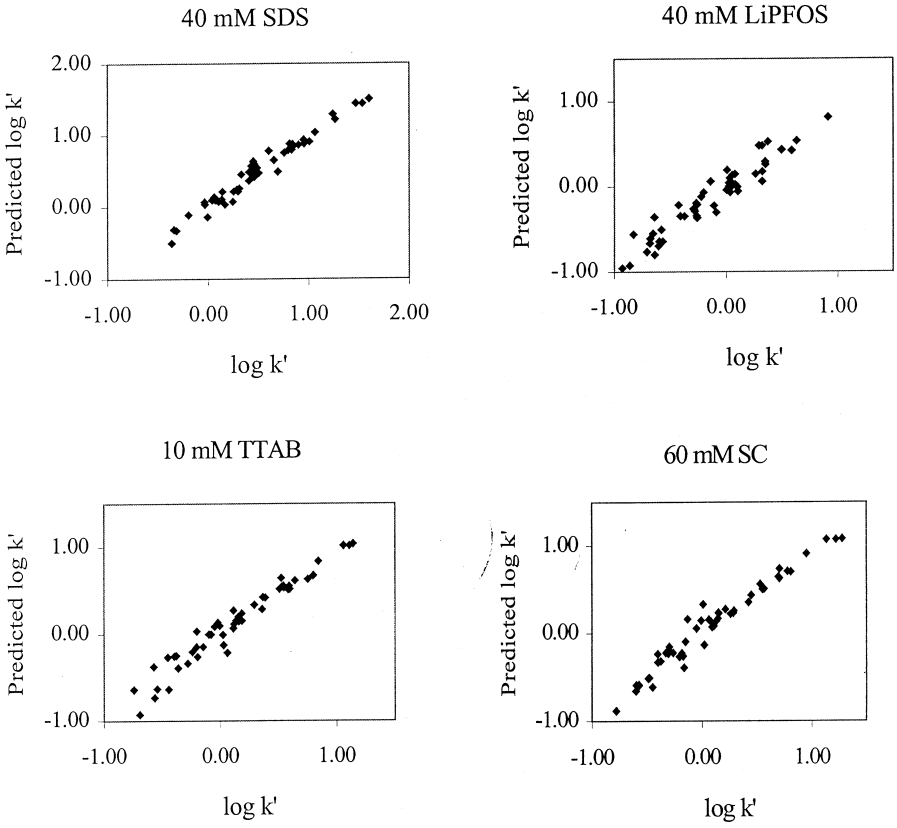

Fig. 1. Predicted versus experimental log k' values for SDS, LiPFOS, TTAB, and SC using Kamlet, Taft and coworkers' model (Eq. (1)) and solute parameters ($n=54$). The coefficient of determination for each is: (SDS) $y=0.967x+0.016$, $R^2=0.967$; (LiPFOS) $y=0.916x-0.011$, $R^2=0.910$; (TTAB) $y=0.946x+0.008$, $R^2=0.946$; (SC) $y=0.956x+0.003$, $R^2=0.956$.

Table 6
LSER coefficients using Abraham et al.'s revised parameters and model (Eq. (2)) ($n=36$)[a]

Surfactant system	c	m	s	b	a	r	SE	R^2_{adj}	SSE
40 m*M* SDS	−1.86	2.98	−0.30	−1.85	−0.18	0.24	0.05	0.987	0.08
		(0.09)	(0.07)	(0.08)	(0.04)	(0.06)			
40 m*M* LiPFOS	−2.01	2.36	0.46	−0.61	−0.80	−0.68	0.09	0.947	0.27
		(0.16)	(0.12)	(0.14)	(0.07)	(0.11)			
10 m*M* TTAB	−2.26	2.99	−0.20	−2.71	0.87	0.30	0.07	0.982	0.13
		(0.11)	(0.08)	(0.10)	(0.05)	(0.08)			
60 m*M* SC	−1.82	2.74	−0.65	−2.51	0.08	0.55	0.07	0.984	0.14
		(0.12)	(0.08)	(0.10)	(0.05)	(0.08)			
30/30 m*M* SC/SDS	−1.67	2.67	−0.51	−2.24	0.08	0.45	0.08	0.970	0.21
		(0.14)	(0.10)	(0.12)	(0.06)	(0.10)			
40 m*M* SDCV	−1.65	2.99	−0.58	−2.43	0.14	0.42	0.07	0.983	0.14
		(0.12)	(0.08)	(0.10)	(0.05)	(0.08)			
40 m*M* SPN	−1.72	3.11	−0.45	−2.58	0.48	0.42	0.07	0.982	0.14
		(0.12)	(0.09)	(0.10)	(0.05)	(0.08)			

[a] Numbers in parentheses are the standard error of the coefficients.

Table 7
LSER coefficients for the reduced solute set using Kamlet and Taft's solute parameters ($n=36$)[a]

Surfactant system	c	m	s	b	a	SE	R^2_{adj}	SSE
40 m*M*	−0.14	2.80	−1.40	−1.45	−0.50	0.29	0.631	3.74
SDS		(0.57)	(0.36)	(0.40)	(0.22)			
40 m*M*	−0.70	1.98	−1.16	0.88	−1.09	0.21	0.743	4.95
LiPFOS		(0.42)	(0.27)	(0.29)	(0.16)			
10 m*M*	−0.34	2.33	−0.99	−1.06	0.68	0.40	0.343	4.54
TTAB		(0.75)	(0.50)	(0.50)	(0.28)			
60 m*M*	−0.34	2.71	−1.22	−1.53	−0.19	0.34	0.572	4.65
SC		(0.65)	(0.43)	(0.43)	(0.26)			
30/30 m*M*	−1.68	4.06	−0.13	−2.70	0.16	0.10	0.958	0.30
SC–SDS		(0.20)	(0.12)	(0.14)	(0.07)			

[a] Numbers in parentheses are the standard errors of the coefficients.

hydrogen bond ''accepting'' ability of three of the systems are all identical to that of the bulk aqueous phase (statistically insignificant b and a coefficients, respectively). Although this is possible it is not likely. The older model (Eq. (1)) also suggests that all of these surfactants have similar interactions with polar solutes. It is also improbable that surfactants that possess hydrocarbon, fluorocarbon, and bile salt hydrophobic groups would all produce micellar phases with identical polarity characteristics.

3.3. Parameter evaluation

The revised LSER model (Eq. (2)) and solute descriptors appear to result in a more sound fit to MEKC retention data. In addition, it has been shown that all of the terms can be included in the model without fear of collinearity. However, it is still useful to evaluate the relative importance of all the parameters in the model. The omission of terms can result in undesirable biasing in the remaining coefficients. As a result, it is common to 'over fit' the model by including extraneous parameters that do not add a significant amount of new information.

There is no outlined manner in which to select the proper variables used to define a model. Statistical tests do not define the best model, but they can be used to aid the decision. Showing that there is good multiple correlation (e.g. R and/or R^2) between the predicted value from the full model and the measured retention does not rule out the possibility that individual parameters fail to yield a substantial improvement of the descriptive power of the model. The coefficient of determination is equal to the ratio of the regression sum of squares (SSR) over the total sum of squares (S_{yy}):

$$R^2 = \frac{\mathrm{SSR}}{S_{yy}} = 1 - \frac{\mathrm{SSE}}{S_{yy}} \tag{3}$$

In this equation, SSR is defined as the sum of squared difference between the predicted responses and the average of all responses, S_{yy} is the squared difference between the measured response and the average of all responses, and the error sum of squares (SSE) is the sum of squared deviations from the predicted values:

$$\mathrm{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{4}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{5}$$

In these equations, n is the number of observations (e.g. the number of solutes in the test set). Adding independent variables to a model will always increase R^2 due to an decrease in SSE. Instead, evaluating the adjusted coefficient of determination (R^2_{adj}) gives a better indication as to whether or not an additional parameter improves the model or simply over fits it. Unlike the correlation coefficient (R) and the coefficient of determination (R^2), R^2_{adj} takes the degrees of freedom of the model into account.

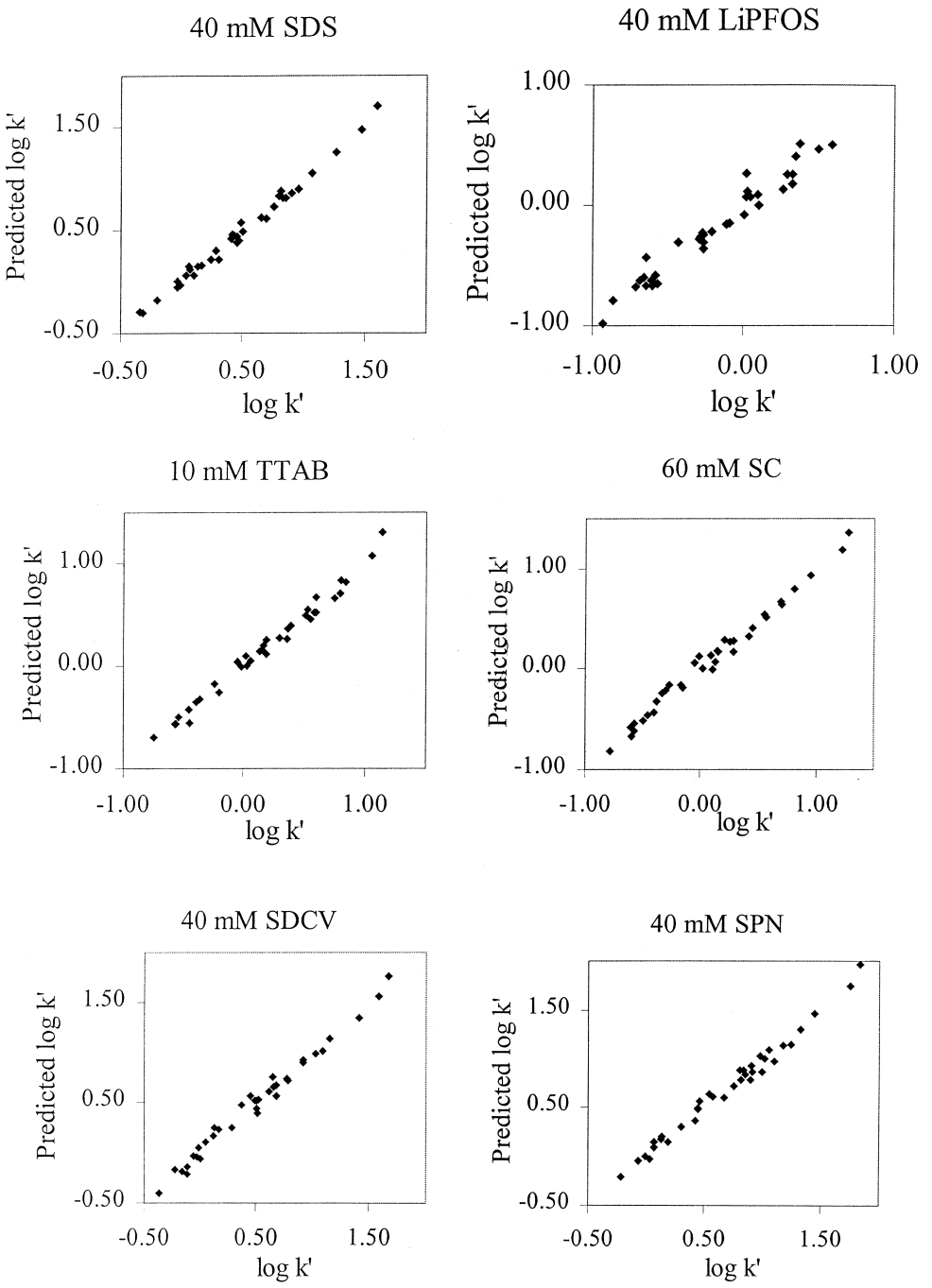

Fig. 2. Predicted versus experimental log k' values for SDS, LiPFOS, TTAB, SC, SDCV and SPN using Abraham and coworker's revised model (Eq. (2)) and solute parameters ($n=36$). The coefficient of determination for each is: (SDS) $y=0.989x+0.005$, $R^2=0.989$; (LiPFOS) $y=0.955x-0.008$, $R^2=0.956$; (TTAB) $y=0.985x+0.003$, $R^2=0.985$; (SC) $y=0.986x+0.001$, $R^2=0.986$, (SDCV) $y=0.985x+0.007$, $R^2=0.985$; (SPN) $y=0.984x+0.011$, $R^2=0.984$.

$$R^2_{adj} = 1 - \frac{SSE/(n-p)}{S_{yy}/(n-1)}$$
$$= 1 - \frac{(n-1)}{(n-p)} \cdot (1 - R^2) \quad (6)$$

In this equation, n has the same definition, and p is the number of independent parameters (including the regression constant) used to model the data.

Table 8 evaluates the importance of each parameter for all of the surfactant systems studied with the revised model. One of the most common methods is to simply inspect the goodness of fit by looking at the coefficient of determination (R^2) and the standard errors. Using these criteria, the full model is the 'best' fit to the MEKC retention data, it is also clear that the rR_A, $s\pi^*_A$, and $a\Sigma\alpha_A$ descriptors can be omitted from the model individually or together without severely affecting the overall fit for most surfactant systems (Table 8). In addition, using the partial models that omit rR_A or $s\pi_A$ only compromise the standard errors modestly. The $a\Sigma\alpha_A$ parameter can be removed without any dramatic changes in the model for SDS, SC, SDCV, and SC–SDS solutions. However, the large range of a coefficient values appears to make it an important descriptor when modeling selectivity differences between a large number of surfactant systems. LiPFOS is the only system that shows truly unique characteristics. Table 6 shows that removing any of the parameters in the revised model can cause a significant loss in model fit. The observed retention and selectivity differences can be still be completely rationalized using the mV, $b\Sigma\beta_A$, and $a\Sigma\alpha_A$ terms.

Another general goal for descriptor selection is to minimize the residual sum of squares (*SSE*). The model that yields the lowest sum of squares value is typically considered to be the most descriptive model. Evaluating each system and model based solely on this criteria also suggests that using all five descriptors provides the best model. However, it is also permissible to omit the $a\Sigma\alpha_A$ term in many of the anionic hydrocarbon surfactant systems. The relevance of all of the parameters was also tested using the F significance test [28]. By examining the residual sum of squares for the full model with respect to the residual sum of squares of the partial model after removing one or more of the parameters, the relative importance of each parameter can be determined (Eq. (7)).

$$F_c = z\frac{(SSE(PM) - SSE(FM))/(f-p)}{SSE(FM)/(n-f)} \quad (7)$$

In this equation, SSE is the residual sum of squares for the full model (FM) and the partial model (PM). The number of parameters (including the system constant) in the full model and the partial model are represented by f and p respectively, and n is the number of measurements (e.g. number of solutes) in the study. A comparison of F_c to F_{table} ($F_{1,30,0.05} = 4.17$) shows that each of the individual parameters are significant at the 95% confidence level after all other parameters have been accounted for (Table 8). Not surprisingly, the F_c for the model that only consists of the mV_x and the $b\Sigma\beta_A$ terms also shows that the other terms should not be omitted simultaneously ($F_{3,30,0.05} = 2.92$). Therefore, all of the descriptors in Abraham et al.'s model (Eq. (2)) bring unique information in describing the observed retention and selectivity in MEKC.

It is important to note that care must be taken when evaluating non hydrocarbon based surfactants. The fluorocarbon hydrophobic tail of LiPFOS causes it to show truly unique interaction characteristics with solute molecules, and only the full model seems to adequately describe retention in these solutions. The high electronegativity of fluorine atoms results in LiPFOS having poor interactions with the n- and π-electrons of solute molecules. Therefore, LiPFOS is the only surfactant studied that has a negative r coefficient. Another unique characteristic of LiPFOS is its unusually high degree of interaction with polar/polarizable solutes. The full model suggests that LiPFOS micelles have a polarity greater than that of bulk aqueous phases (positive s coefficient). Although this does not seem likely, LSER still shows that LiPFOS is certainly in a surfactant class of its own and all possible parameters should be considered when using it.

4. Conclusions

These results show that LSER provides a good model of retention and selectivity differences be-

Table 8
Comparison of LSER coefficients and statistical analysis of Abraham et al.'s model[a]

	Parameters used									
	c	m	s	b	a	r	SE	R^2_{adj}	SSE	F_{calc}
(A) 40 m*M* SDS										
Full model	−1.86	2.98	−0.30	−1.85	−0.18	0.24	0.05	0.987	0.08	–
		(0.09)	(0.07)	(0.08)	(0.04)	(0.06)				
Omitting rR_A	−1.92	3.16	−0.16	−2.01	−0.17	–	0.06	0.982	0.12	13.93
		(0.09)	(0.06)	(0.08)	(0.05)					
Omitting $s\pi^*_A$	−2.00	3.07	–	−2.03	−0.23	0.06	0.07	0.979	0.15	21.69
		(0.12)		(0.08)	(0.05)	(0.07)				
Omitting $a\Sigma\alpha_A$	−1.96	3.15	−0.39	−1.88	–	0.21	0.07	0.978	0.15	22.86
		(0.11)	(0.08)	(0.10)		(0.08)				
Omitting $b\Sigma\beta_A$	−1.27	1.87	−1.12	–	−0.27	1.09	0.23	0.752	1.71	580.00
		(0.35)	(0.25)		(0.17)	(0.23)				
Omitting mV_x	0.30	–	−0.75	−0.58	−0.68	1.33	0.32	0.554	3.07	1065.71
			(0.38)	(0.39)	(0.21)	(0.32)				
Omitting π^*_2, rR_A, and $a\Sigma\alpha_A$	−2.21	3.1	–	−2.14	–	–	0.08	0.965	0.26	62.50
		(0.12)								
(B) 40 m*M* LiPFOS										
Full model	−2.01	2.36	0.46	−0.61	−0.80	−0.68	0.09	0.974	0.27	–
		(0.16)	(0.12)	(0.14)	(0.07)	(0.11)				
Omitting rR_A	−1.83	1.84	0.04	−0.15	−0.83	–	0.14	0.887	0.59	36.08
		(0.20)	(0.14)	(0.17)	(0.10)					
Omitting $s\pi^*_A$	−1.79	2.22	–	−0.33	−0.72	−0.41	0.12	0.922	0.41	15.59
		(0.19)		(0.14)	(0.08)	(0.19)				
Omitting $a\Sigma\alpha_A$	−2.47	3.12	0.09	−0.76	–	−0.79	0.22	0.175	1.49	136.08
		(0.35)	(0.26)	(0.32)		(0.26				
Omitting $b\Sigma\beta_A$	−1.82	1.99	0.19	–	−0.83	−0.41	0.12	0.915	0.45	19.71
		(0.18)	(0.13)		(0.09)	(0.12)				
Omitting mV_x	−0.30	–	0.11	0.40	−1.19	0.18	0.26	0.591	2.14	208.46
			(0.32)	(0.33)	(0.17)	(0.27)				
Omitting π^*_2, rR_A, and $a\Sigma\alpha_A$	−2.61	2.56	–	−0.43	–	–	0.25	0.617	2.14	207.91
		(0.34)		(0.29)						
(C) 10 m*M* TTAB										
Full model	−2.26	2.99	−0.20	−2.71	0.87	0.30	0.07	0.982	0.13	–
		(0.11)	(0.08)	(0.10)	(0.05)	(0.08)				
Omitting rR_A	−2.34	3.22	−0.02	−2.91	0.88	–	0.08	0.974	0.20	14.41
		(0.11)	(0.08)	(0.10)	(0.06)					
Omitting $s\pi^*_A$	−2.36	3.05	–	−2.83	0.84	0.18	0.07	0.979	0.16	6.53
		(0.12)		(0.09)	(0.05)	(0.07)				
Omitting $a\Sigma\alpha_A$	−1.76	2.16	0.20	−2.54	–	0.41	0.23	0.792	1.59	327.48
		(0.36)	(0.27)	(0.33)		(0.27)				
Omitting $b\Sigma\beta_A$	−1.41	1.36	−1.40	–	0.74	1.55	0.34	0.525	3.62	785.81
		(0.51)	(0.36)		(0.25)	(0.34)				
Omitting mV_x	−0.10	–	−0.66	−1.44	0.37	1.40	0.32	0.589	3.13	675.90
			(0.38)	(0.40)	(0.21)	(0.33)				
Omitting π^*_2, rR_A, and $a\Sigma\alpha_A$	−1.50	2.45	–	−2.61	–	–	0.24	0.757	1.97	414.19
		(0.32)		(0.28)						

Table 8. Continued

	Parameters used									
	c	m	s	b	a	r	SE	R^2_{adj}	SSE	F_{calc}
(D) 60 m*M* SC										
Full model	−1.83	2.75	−0.66	−2.52	0.08	0.55	0.07	0.984	0.14	–
		(0.12)	(0.08)	(0.10)	(0.05)	(0.08)				
Omitting rR_A	−1.97	3.16	−0.31	−2.89	0.11	–	0.11	0.960	0.35	45.16
		(0.16)	(0.10)	(0.13)	(0.08)					
Omitting $s\pi_A^*$	−2.13	2.94	–	−2.92	−0.02	0.17	0.12	0.952	0.42	60.65
		(0.20)		(0.14)	(0.08)	(0.11)				
Omitting $a\Sigma\alpha_A$	−1.78	2.67	−0.62	−2.50	–	0.56	0.07	0.983	0.15	2.80
		(0.11)	(0.08)	(0.10)		(0.08)				
Omitting $b\Sigma\beta_A$	−1.03	1.23	−1.76	–	−0.03	1.71	0.32	0.635	3.16	268.92
		(0.48)	(0.34)		(0.23)	(0.32)				
Omitting mV_x	0.16	–	−0.17	−1.35	−0.37	1.56	0.29	0.691	2.67	545.38
			(0.35)	(0.36)	(0.20)	(0.30)				
Omitting π_2^*, rR_A, and $a\Sigma\alpha_A$	−2.11	3.08	–	−2.99	–	–	0.12	0.951	0.45	67.10
		(0.16)		(0.13)						
(E) 30 m*M* SC–30 m*M* SDS										
Full model	−1.67	2.68	−0.51	−2.24	0.08	0.45	0.08	0.970	0.21	–
		(0.14)	(0.10)	(0.12)	(0.06)	(0.10)				
Omitting rR_A	−1.78	3.01	−0.24	−2.54	0.10	–	0.11	0.952	0.34	20.20
		(0.15)	(0.10)	(0.13)	(0.08)					
Omitting $s\pi_A^*$	−1.90	2.83	–	−2.55	−0.00	0.15	0.11	0.947	0.38	25.18
		(0.19)		(0.14)	(0.08)	(0.11)				
Omitting $a\Sigma\alpha_A$	−1.62	2.60	−0.47	−2.23	–	0.46	0.08	0.970	0.22	1.18
		(0.13)	(0.10)	(0.12)		(0.10)				
Omitting $b\Sigma\beta_A$	−0.96	1.32	−1.50	–	−0.03	1.48	0.29	0.637	2.60	350.37
		(0.43)	(0.31)		(0.21)	(0.29)				
Omitting mV_x	0.27	–	−0.92	−1.11	−0.36	1.43	0.29	0.635	2.61	352.12
			(0.35)	(0.36)	(0.19)	(0.30)				
Omitting π_2^*, rR_A, and $a\Sigma\alpha A_A$	−1.87	2.93	–	−2.61	–	–	0.11	0.947	0.40	28.99
		(0.15)		(0.13)						
(F) 40 m*M* SDCV										
Full model	−1.65	2.99	−0.58	−2.43	0.14	0.42	0.07	0.983	0.14	–
		(0.12)	(0.08)	(0.10)	(0.05)	(0.08)				
Omitting rR_A	−1.76	3.31	−0.33	−2.71	0.16	–	0.09	0.968	0.26	25.92
		(0.13)	(0.09)	(0.11)	(0.07)					
Omitting $s\pi_A^*$	−1.92	3.16	–	−2.78	−0.04	0.07	0.11	0.956	0.36	48.16
		(0.19)		(0.13)	(0.08)	(0.10)				
Omitting $a\Sigma\alpha_A$	−1.57	2.86	−0.52	−2.40	–	0.44	0.08	0.979	0.18	7.78
		(0.12)	(0.09)	(0.11)		(0.09)				
Omitting $b\Sigma\beta_A$	−0.89	1.53	−1.65	–	0.02	1.53	0.31	0.643	2.94	604.75
		(0.46)	(0.32)		(0.22)	(0.31)				
Omitting mV_x	0.513	–	−1.03	−1.15	−0.36	1.52	0.32	0.617	3.15	649.68
			(0.38)	(0.40)	(0.21)	(0.33)				
Omitting π_2^*, rR_A, and $a\Sigma\alpha_A$	−1.49	3.00	–	−2.77	–	–	0.15	0.913	0.75	51.40
		(0.20)		(0.17)						

Table 8. Continued

	Parameters used									
	c	m	s	b	a	r	SE	R^2_{adj}	SSE	F_{calc}
(G) 40 m*M* SPN										
Full model	−1.72	3.11 (0.12)	−0.45 (0.09)	−2.58 (0.10)	0.48 (0.05)	0.42 (0.08)	0.07	0.982	0.14	–
Omitting rR_A	−1.83	3.43 (0.14)	−0.19 (0.09)	−2.86 (0.11)	0.50 (0.07)	–	0.10	0.967	0.27	25.63
Omitting $s\pi^*_A$	−1.93	3.24 (0.16)	–	−2.85 (0.12)	0.41 (0.07)	−0.16 (0.09)	0.09	0.966	0.27	27.31
Omitting $a\Sigma\alpha_A$	−1.45	2.66 (0.22)	−0.22 (0.16)	−2.49 (0.20)	–	0.48 (0.16)	0.14	0.591	0.58	91.60
Omitting $b\Sigma\beta_A$	−0.91	1.56 (0.49)	−1.58 (0.34)	–	0.36 (0.23)	1.61 (0.33)	0.33	0.579	3.30	662.61
Omitting mV_x	0.53	–	−0.91 (0.40)	−1.25 (0.41)	−0.04 (0.22)	1.57 (0.34)	0.33	0.579	3.39	682.77
Omitting π^*_2, rR_A, and $a\Sigma\alpha_A$	−1.49	3.00 (0.20)	–	−2.77 (0.17)	–	–	0.15	0.913	3.39	127.31

[a] Numbers in parentheses are the standard errors of the coefficients.

tween MEKC surfactant systems. They also agree with the conclusions of Poole et al. and Abraham et al. in that the revised solute descriptors provide a better fit to MEKC retention data. Statistical analysis suggests that using the revised parameters in the full model (Eq. (2)) is more reliable and can more accurately assess small differences in selectivity between similar surfactant systems. An additional advantage of these parameters is that they allow a smaller solute set to be used without significant loss in model fit which can save a significant amount of analysis time. However, it is important to realize that regardless of the fit and errors in the constants, at this time these models can only be used to obtain qualitative information about the selectivity differences between surfactant systems used in MEKC. The predictive power of these models is extremely limited since the necessary information is unavailable for most solutes. Therefore, although one set of parameters clearly show a better fit to the retention data, both models and sets of parameters typically yield the same conclusions about selectivity differences between surfactants in MEKC. Poole and Poole suggested that the interpretation of the LSER results is dependent on the descriptor values and the model used [6]. Although the model and parameters revised by Abraham et al. (Eq. (2)) provide different and often more sensible coefficient values, the interpretation of the interactions important to solute retention and selectivity does not change. Both Tables 5 and 6 show that the solute size (large m coefficient) has the largest influence on retention. The ability of the micelle to interact with the hydrogen bond accepting moieties of a solute (b coefficient) also plays an important role in solute retention and has the most significant effect on selectivity differences between these systems. Regardless of the model or parameters used, LiPFOS and SDS always have the strongest interactions and TTAB, SC, and SDCV have the weakest interactions with hydrogen bond accepting solutes. In addition, TTAB always has the strongest interaction with hydrogen bond donating solutes. Finally it is interesting to note that TTAB has the weakest interaction with hydrogen bond accepting and one of the strongest with polar/polarizable solutes [1–3,7,26,27]. This is contrary to the general observation that micelles having strong interactions with hydrogen bond accepting solutes typically also have strong interactions with polar solutes.

Acknowledgements

The authors wish to gratefully thank Dr. Sahori Takeda at the Osaka National Research Institute for

donating sodium *N*-parmitoyl sarcosinate and Waters Corporation for donating the (*S*)-dodecoxy-carbonylvaline. We would also like to thank Dr. Conseulo Arellano from the Statistics Department for her helpful discussions during the preparation of this paper. A research grant from the US National Institutes of Health (GM 38738) is also acknowledged.

References

[1] S. Yang, M.G. Khaledi, Anal. Chem. 67 (1995) 499.
[2] S. Yang, M.G. Khaledi, J. Chromatogr. A 692 (1995) 301.
[3] S. Yang, J.G. Bumgarner, M.G. Khaledi, J. Chromatogr. A 738 (1996) 265.
[4] P.G. Muijselaar, H.A. Claessens, C.A. Cramers, Anal. Chem. 69 (1997) 1184.
[5] S.K. Poole, C.F. Poole, Anal. Commun. 34 (1997) 57.
[6] S.K. Poole, C.F. Poole, Analyst 122 (1997) 267.
[7] M.G. Khaledi, J.G. Bumgamer, M. Hadjmohammadi, J. Chromatogr. A 802 (1998) 35.
[8] M.J. Kamlet, R.W. Taft, J Am. Chem. Soc. 98 (1976) 377.
[9] R.W. Taft, M.J. Kamlet, J. Am. Chem. Soc. 98 (1976) 2886.
[10] T. Yokoyama, R.W. Taft, M.J. Kamlet, J. Am. Chem. Soc. 98 (1976) 3323.
[11] M.J. Kamlet, J.-L. Abboud, R.W. Taft, J. Am. Chem. Soc. 98 (1976) 6027.
[12] M.J. Kamlet, M.H. Abraham, R.M. Doherty, R.W. Taft, J. Am Chem. Soc. 106 (1984) 464.
[13] M.J. Kamlet, J.-L. Abboud, R.M. Doherty, M.H. Abraham, R.W. Taft, CHEMTECH (1986) 566.
[14] M.H. Abraham, R.M. Doherty, M.J. Kamlet, R.W. Taft, Chem. Br. 22 (1986) 551.
[15] P.W. Carr, Microchein. J. 48 (1993) 4, and references cited herein.
[16] M.J. Kamlet, R.M. Doherty, M.H. Abraham, Y. Marcus, R.W. Taft, J. Phys. Chem. 92 (1988) 5244.
[17] M.H. Abraham, G.S. Whiting, R.M. Doherty, W.J. Shuely, J. Chem. Soc., Perkin Trans. 2 (1990) 1451.
[18] M.H. Abraham, H.S. Chadha, G.S. Whiting, R.C. Mitchell, J. Pharm. Sci. 83 (1994) 1085.
[19] J.C. McGowan, M.H. Abraham, Chromatographia 23 (1987) 243.
[20] J.D. Weckwerth, P.W. Carr, Anal. Chem. 70 (1998) 4793.
[21] C.F. Poole, S.K. Poole, M.H. Abraham, J. Chromatogr. A 798 (1998) 207.
[22] C.F. Poole, S.K. Poole, J. Chromatogr. A 792 (1997) 89.
[23] M.H. Abraham, H.S. Chadha, J.P. Dixon, C. Rafols, C. Treiner, J. Chem. Soc., Perkin Trans. 2 (1995) 887.
[24] F.H. Quina, E.O. Alonso, J.P.S. Farah, J. Phys. Chem. 99 (1995) 11708.
[25] M.F. Vitha, A.J. Dallas, P.W. Can, J. Colloid Interface Sci. 187 (1997) 79.
[26] M.D. Trone, M.G. Khaledi, Anal. Chem. 71 (1999) 1270.
[27] M.D. Trone, M.G. Khaledi, Electrophoresis, submitted for publication.
[28] S. Chatterjee, P. Bertram, Regression Analysis by Example, Wiley, New York, 1977.