Iranian Journal of Pharmaceutical Sciences
2004:1(1): 53-57

Original Article

Modeling The Capacity Factor of Analytes in Micellar Electrokinetic Chromatography Using Computational Descriptors

Abolghasem Jouyban[a,*], Samira Yeghanli[b], Maryam Khoubnasabjafari[c]

[a]*School of Pharmacy,* [b]*Pharmaceutical Research Center, Tabriz University of Medical Sciences, Tabriz, Iran, 51664*
[c]*Kimia Research Institute, Tabriz, Iran, P.O. Box 51665-171*

Abstract

A multiple linear regression model is proposed to calculate capacity factor of analytes using structural features computed using HyperChem software. The chemical descriptors of analytes were computed using HyperChem software and regressed against the experimental capacity factors of analytes collected from the literature. The absolute average percentage deviation (AARD) and individual percentage deviation (IPD) were calculated as accuracy criteria. The accuracy of the proposed method was compared with that of a previously reported linear solvation energy relationship (LSER). The proposed method was tested on ten experimental data sets and mean ± standard deviation of AARDs were 48.5 ± 20.4 and 130.1 ± 79.7, respectively, for the proposed and LSER models in which the mean difference was statistically significant ($p<0.01$). The distribution of IPDs sorted in three subgroups, i.e. ≤ 45%, 45%-90%, and >90 %, shows the superiority of the proposed model over the LSER. A significant improvement in capacity factor modeling was achieved and the improvement factor is about 2.7. The descriptors could be easily computed and the calculations are straightforward. Therefore, the model is suggested to be employed in practice, however, the efforts should be continued until providing more accurate models.

Keywords: Capacity factor; Micellar electrokinetic chromatography; Modeling.
Received: August 2004; *Accepted:* November 2004

1. Introduction

Electrophoresis is a sepration technique in which the analytes are separated based on their mobilities in a conductive medium, usually an aqueous buffer, in response to an applied electric field. Capillary electrophoresis (CE) is a relatively new technique in analytical sciences, which separates charged species using high voltage (i.e. up to 30 KV and even more). In this method, when an electric field is applied to a capillary tube, the sample's ions migrate as a result of two

*Corresponding author: Abolghasem Jouyban, Department of Medicinal Chemistry, School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Easthern Azarbayjan, Iran, 51664.
Tel. (+98) 411-3341315, Fax (+98) 411-3344798
Email: ajouyban@hotmail.com

types of movements, i.e. electrophoretic and electroosmotic mobilities. Electrophoretic mobility is the ion's response to the applied electric field. Cations move toward the negetively charged cathod, anions move toward the positively charged anode, and neutral species, which do not respond to the electric field; migrate with the electroosmothic flow. Consequently, the charged analytes in CE could be separetd from each other and the uncharged analytes move altogether with electroosmotic flow. By adding a surface active agent above its critical micelle concentration to the running buffer, this limitation of CE could be overcome. The resulted method is called micellar electrokinetic chromatography (MEKC). Uncharged molecules in MEKC are separated according to their partitioning between the aqueous phase and the micellar phase (pesudostationary phase). In the case of charged solutes, a combination of partitioning between the phases and electrophoretic mobility is the mechanisms of their separations.

The capacity factor (k') or retention factor (k) in MEKC is defined as the ratio of total amount of the analyte in the micellar phase ($n_{mic.}$) to the amount in the running buffer ($n_{buf.}$), and can be calculated by:

where $K_{mic.-buf.}$ is the micelle-buffer partition coefficient, $V_{mic.}$ and $V_{buf.}$ are volumes of miecellar and buffer phases, t_r, t_0 and $t_{mic.}$ are migration times of the analyte, electroosmotic flow and micelles, respectively [(1]). The k' is the main characteristics of an analyte in MEKC and in order to predict the possibility of successful separation of an analyte in a given analytical condition, its numerical value should be estimated. Despite the experimental determination of k' values, it could be calculated using quanitative structure property relationships (QSPRs). The QSPRs are able to predict k' values after training by a minimum number of experimental data points which is called training set, and a pure predictive method (a method without any curve-fitting parameter) is not available so far. It should be also noted that the QSPR models are able to predict k' values at the analytical conditions which the training data sets have been collected and by changing the analytical conditions, the models should be re-trained using new data sets. Although a number of equations were presented for modeling of the electrophoretic mobility data in CE [(2-6]), the only available model for calculating capacity factor of analytes in MEKC is the linear solvation relationship (LSER) model, which is used in the literature ([7-10]). In addition to the lowess accuracy and relatively the high percentage errors of LSER, the numerical values of its descriptors are not easily available for most of the pharmaceutical compounds. The aim of this work is to present a simple least squares model for calculating k' of analytes using computational descriptors. The accuracy of the proposed model is checked employing experimental k' data collected from the literature and, it is compared with that of a previously published model.

2. Experimental data and computational methods

2.1. Experimental data

The eExperimental data of capacity factors of pharmaceutically/chemically-interested compounds was collected from the literature. The details of the data were given in Tables 1 to -4. In Table 2, for a number of analytes, using different surfactants, the numerical values of k' arewas reported equal to zero which means that the analyte moves with electroosmotic flow, and there is no tendency to the micelles. These data points were excluded from the calculations, since ln (0) is not a defined value.

2.2. Computational methods

2.2.1. Calculation of the descriptors

More detailed structural information of the studied compounds was calculated using molecular and quantum mechanical calculation methods. The theoretical descriptors (containing hydrophobic, electronic, theoretical and steric descriptors) for each compound were computed using AM1 semi-empirical quantum mechanical method by employing molecular descriptors, properties and orbital programs of HyperChem 7.0 [(11]). The structure of each compound was drawn in 2D and was converted to 3D using HyperChem 7.0 [(11]), and pre-minimized by Polak-Ribiere (12) geometry optimization using MM+ [(12). The structures were found by MM+, used as the starting point for re-minimization by Polak-Ribiere optimization using AM1 [(13]) semi-empirical quantum mechanical method. The calculated descriptors were octanol-water partition coefficient, positive and negative partial charges on the analytes, accessible surface area, molar volume, hydration energy, molar refractivity, polarizability, dipole moment, total energy and heat of formation.

2.2.2. Statistical analysis

The calculated descriptors and the collected capacity factors were stored in SPSS data files and were analyzed using SPSS (version 10.0) software. To select the descriptors for including in the model, the Pearson correlation of log k' and the descriptors were studied and the overall mean of correlation coefficients of 10 studied sets was considered along with the correlation matrix of the descriptors. The selection criteria were the highest mean of the correlation coefficients between log k' and the descriptors and the lowest mean of inter-corelations between the descriptors. To provide a single multi-linear model, the selected descriptors were included in the model building process using ENTER subcommand of SPSS.

A number of diagnostic criteria are available to check the performance of a model. The most widely used criterion is the coefficient of determination (R^2) or the correlation coefficient (R). It is calculated from the regression sum of squares and the total sum of squares. The significance of R^2 and R values could be checked using F value and its associated probability. These criteria give useful information regarding the performance of the model. However, one could not compare the calculated k' values with the experimental values. In addition, the numerical values of R and F could be affected by logarithmic or other mathematical transformations on the dependent variable, the number of independent variables included in the model and also the number of data points in each set. Therefore, to test the correlation ability of the proposed model, all data points in each set were fitted to the model and the back-calculated capacity factors were compared with the corresponding experimental values and the absolute average relative deviation (AARD) was computed as an accuracy criterion, by:

where N is the number of data points in each set. The main advantage of AARD is that, it is a comparable quanitity with the experimentally obtained relative standard deviation (RSD) of repeated experiments. To study the individual percentage deviations (IPD) of each data, the IPDs were computed using:

To test the stability of the correlation ability of the proposed model, the cross-validation by one-leave-out method was employed. To do this, one datum was excluded from the set and the remaining (N-1) data points were correlated. Then the back-calculated mobilities for each run of

Table 1. Details of analytes, experimental logk' values (8) and the chemical descriptors

Analyte	logk'[a]	LP	CH^+	TE	PZ
p-Xylene	0.78	2.98	0.13	-26798.7	14.10
4-Bromotoluene	1.07	3.31	0.15	-31036.2	14.90
4-Chlorotoluene	0.92	3.03	0.15	-31508.9	14.20
4-Nitrotoluene	0.46	0.53	0.57	-42365.5	14.11
1,4-Diethylbenzene	1.55	3.77	0.13	-33983.1	17.77
Diphenylamine	1.27	3.30	0.23	-43668.3	21.44
4-Amionophenol	-0.76	0.98	0.22	-32098.2	12.42
4-Propylphenol	0.88	3.02	0.22	-37783.1	16.58
3-Bromophenol	0.41	2.55	0.22	-34834.4	13.70
4-Bromophenol	0.44	2.55	0.22	-34835.0	13.70
4-Iodophenol	0.68	3.02	0.22	-34696.9	16.10
4-Ethylphenol	0.43	2.63	0.22	-34190.2	14.74
4-Butoxyphenol	0.98	2.72	0.22	-48755.4	19.05
4-Fluorophenol	-0.22	1.90	0.22	-37874.1	10.98
4-Isopropylphenol	0.77	2.96	0.22	-37781.0	16.58
3,5-Dimethylphenol	0.40	2.70	0.22	-34192.0	14.74
3-Chlorophenol	0.27	2.28	0.22	-35307.5	13.00
4-Chlorophenol	0.29	2.28	0.22	-35307.6	13.00
4-Propoxyphenol	0.54	2.32	0.22	-45161.7	17.21
3-Methylphenol	0.01	2.23	0.22	-30597.6	12.91
4-Methylphenol	0.04	2.23	0.22	-30597.6	12.91
4-Methoxyphenol-	0.41	1.51	0.22	-37976.3	13.54
4-Acetylphenol	-0.07	1.07	0.27	-40933.9	14.83
4-Ethoxyphenol	0.06	1.85	0.22	-41568.6	15.38
Phenethyl alcohol	-0.06	1.76	0.20	-34192.8	14.74
4-Chlorobenzyl alcohol	0.27	2.03	0.20	-38899.8	14.83
4-Methylbenzyl alcohol	0.09	1.98	0.20	-34189.8	14.74
3-Methylbenzyl alcohol	0.07	1.98	0.20	-34193.4	14.74
4-Hdroxybenzyl alcohol	-0.67	1.23	0.22	-37988.8	13.54
1,4-Benzenedimethanol	-0.58	0.98	0.20	-41580.9	15.38
4-Bromoacetophenone	0.82	2.15	0.27	-41371.0	16.82
Propiophenone	0.47	4.16	0.27	-37132.0	16.03
Butyrophenone	0.84	4.55	0.26	-40725.0	17.86
Hexanophenone	1.69	3.17	0.27	-47911.8	21.53
Valerophenone	1.24	2.78	0.27	-44318.7	19.69
1,4-Acetylbenzene	0.26	0.66	0.26	-47469.2	17.95
4-Chloroacetophenone	0.67	1.87	0.27	-41844.0	16.12
4-Iodacetophenone	1.04	2.61	0.27	-41232.7	19.22
4`-Fluoroacetophenone	0.21	1.49	0.27	-44411.1	14.10
p-Aminobenzophenone	1.06	2.48	0.32	-54009.0	23.37
1,3,5-Trichlorobenzene	1.55	3.60	0.17	-44521.6	16.22
o-Dichlorobenzene	0.98	3.08	0.15	-36217.2	14.29
1,4-Dibromobenzene	1.21	3.63	0.15	-35272.8	15.69
4-Bromochlorobenzene	1.07	3.36	0.15	-35745.7	14.99
1-Fluoro-4-iodobenzene	0.88	3.96	0.16	-46475.9	17.30
p-Dichlorobenzene	0.92	3.08	0.15	-36218.5	14.29
1-Ethyl-4- iodobenzene	1.72	4.17	0.14	-34490.5	19.13
1-Bromo-4- iodobenzene	1.48	4.10	0.15	-35134.6	18.09
1-Chloro-4- iodobenzene	1.33	3.82	0.15	-35607.5	17.39
1-Chloro-4-fluorobenzene	0.47	2.70	0.15	-38785.3	12.27
2-Chloronitrobenzene	0.42	1.78	0.23	-33011.0	13.71
1-Bromo-4-fluorobenzene	0.63	2.98	0.16	-38312.6	12.97

4-Chloronitrobenzene	0.43	0.59	0.57	-47074.0	14.20
4-Bromonitrobenzene	0.58	0.86	0.57	-46600.8	14.90
4-Bromobenzaldehyde	0.63	2.51	0.22	-37778.0	14.98
4-Iodobenzaldehyde	0.86	2.98	0.22	-37639.6	17.38
1-Methylnaphthalene	1.44	3.52	0.13	-35641.4	19.54
2-Methylnaphthalene	1.48	3.52	0.13	-35643.6	19.54
1-Nitronaphthalene	1.20	1.07	0.57	-51206.6	19.54
4-Bromoanisole	0.87	2.59	0.16	-38415.6	15.53
4-Chloroanisole	0.72	2.31	0.15	-38888.1	14.83
Methyl-2-methylbenzoate	0.79	1.82	0.27	-37126.6	16.03
Ethyl benzoate	0.83	1.98	0.27	-37132.1	16.03
2-Amino-m-cresol	-0.23	1.45	0.24	-35692.7	14.26
2-Amino-p-cresol-	0.14	1.45	0.24	-35692.9	14.26
3-Amino-p-cresol-	0.41	1.45	0.23	-35694.0	14.26
4-Amino-o-cresol-	0.28	1.45	0.22	-35691.7	14.26
2,3,5,6-Tetrachloroaniline	1.90	3.34	0.24	-57920.2	19.50
4-Bromoaniline	0.36	2.06	0.23	-32538.7	14.41
4-Chloroaniline	0.20	1.78	0.23	-33010.9	13.71
Ethoxybenzene	0.45	2.14	0.15	-34176.0	14.74
1,4-Dimethoxybenzene	0.31	1.54	0.15	-41556.8	15.38
1,4-Diacetoxybenzene	0.36	1.04	0.32	-62239.1	19.22
2,7-Dimethylquinoline	1.19	3.25	0.15	-40733.7	20.66
Quinaldine	0.92	2.79	0.15	-37139.4	18.83
Phenazine	1.18	3.02	0.15	-47472.3	23.56
Biphenyl	1.55	3.73	0.13	-38584.0	20.09
Carbazole	1.43	2.94	0.25	-43031.6	20.67
p-Anisidine	-0.09	1.01	0.22	-35678.6	14.26
p-Tolunitrile	0.42	2.70	0.16	-33469.8	15.40

[a] k' values determined using 40 mM SDS in 20 mM NaH_2PO4 at pH=7.0 or 12.0 [(8]).

the analysis were used to compute the AARD term. Less variation in AARD values for the runs means that the model is robust and taking out one datum and/or addition of one more datum could not affect its accuracy.

2.2.3. The proposed model

The correlation of the calculated descriptors with the logarithm of the capacity factor (logk') were studied considering above mentioned statistical criteria, and the following equation is proposed to provide better accuracy for modeling capacity factor of analytes in MEKC. The model is:

wWhere J_0-J_4 are the model constants computed using a least squares analysis and their numerical values represent the interactions between the set of analytes and the buffer and/or micelles, LP is the logarithm of partition coefficient, TE represents the total energy, PZ stands for polarizability and CH^+ is the partial positive charge on the analyte.

The previously presented LSER model [(7-10]) is:

wWhere c, v, r, s, a, and b are the model constants, V_x is the McGowan's characteristics volume, R_2 is the excess molar refractivity, is the analyte's dipolarity/polarizability, and are the analyte's hydrogen bond acidity and basicity, respectively. The basic mechanism of the analyte's retention in MEKC is the partitioning between the buffer and the micelles. Both models represents the possible interactions in the solution using different descriptors of the analytes. The LSER model has been employed to represent other

Table 2. Details of analytes, experimental logk' [10] values (10) and the chemical descriptors

Analyte	logk'							LP	CH+	TE	PZ
	LPFOSa	LDSa	SDSa	SCa	SDCa	TTABa	HTABa				
Butan-1-ol	-0.276	-0.310	-0.301	-0.357	-0.553	-b	-b	0.94	0.20	-22402.8	8.75
Pentan-1-o1	-0.086	-0.086	-0.086	-0.174	-0.409	-0.585	-b	1.34	0.20	-25996.5	10.60
Pentan-3-o1	-0.222	-0.252	-0.244	-0.337	-0.523	-b	-b	1.43	0.20	-25994.1	10.60
Butan-1,4-diol	-b	-b	-0.699	-b	-b	-b	-b	-0.20	0.198	-29796.5	9.39
Pentan-1,5-diol	-0.456	-0.569	-0.569	-0.585	-0.638	-b	-b	0.19	0.20	-33389.6	11.20
Thiourea	-b	-0.620	-0.620	-0.481	-0.602	-0.658	-0.620	-0.31	0.26	-18256.3	8.56
Benzene	-0.222	-0.022	-0.046	-0.027	-0.167	-0.071	-0.013	1.60	0.13	-19609.7	10.40
Toluene	-0.081	0.170	0.149	0.185	0.076	0.117	0.182	1.75	0.13	-23204.2	12.30
Ethylbenzene	0.041	0.342	0.322	0.367	0.288	0.281	0.350	2.15	0.13	-26796.9	14.10
Propylbenzene	0.182	0.547	0.534	0.558	0.487	0.474	0.534	2.54	0.13	-30389.8	15.90
Butylbenzene	0.324	0.760	0.750	0.696	0.643	0.660	0.661	2.94	0.13	-33983.5	17.80
p-Xylene	0.053	0.358	0.348	0.371	0.301	0.294	0.367	1.90	0.13	-26798.7	14.10
Naphthalene	0.049	0.483	0.462	0.439	0.393	0.483	0.558	1.67	0.13	-32051.0	16.60
Chlorobenzene	-0.086	0.220	0.204	0.241	0.127	0.190	0.253	1.37	0.15	-27914.4	12.40
Bromobenzene	-0.076	0.281	0.274	0.310	0.212	0.262	0.330	1.65	0.15	-27441.6	13.10
Anisole	-0.060	0.090	0.057	0.000	-0.114	0.004	0.053	1.79	0.15	-30583.6	12.90
Benzaldehyde	-0.027	0.009	-0.032	-0.201	-0.260	-0.167	-0.119	0.57	0.22	-29946.9	12.40
Acetophenone	0.093	0.090	0.049	-0.143	-0.194	-0.114	-0.071	1.53	0.27	-33539.8	14.20
Propiophenone	0.204	0.230	0.190	0.000	-0.060	0.049	0.100	2.16	0.27	-37132.1	16.00
Butyrophenone	0.326	0.386	0.348	0.155	0.093	0.210	0.265	2.55	0.27	-40725.0	17.90
Valerophenone	0.455	0.558	0.525	0.328	0.286	0.387	0.455	2.95	0.27	-44318.7	19.70
Heptanophenone	0.750	0.957	0.953	0.651	-c	-c	-c	3.74	0.27	-51506.2	23.40
Benzophenone	0.455	0.591	0.558	0.360	0.305	0.415	0.473	2.68	0.31	-48910.9	22.00
Methyl benzoate	0.170	0.225	0.188	0.025	-0.060	0.037	0.086	1.01	0.35	-40935.2	14.80
Benzyl benzoate	0.524	0.838	0.866	0.582	0.566	0.702	-c	2.03	0.34	-59901.3	24.50
Benzonitrile	-0.009	0.004	-0.027	-0.194	-0.268	-0.149	-0.102	1.32	0.15	-27001.2	12.30
Aniline	-0.276	-0.137	-0.167	-0.319	-0.398	-0.208	-0.155	1.26	0.22	-24705.9	11.80
o-Toluidine	-0.167	-0.004	-0.032	-0.208	-0.292	-0.066	-0.018	0.03	0.22	-28298.5	13.60
3-Chloroaniline	-0.244	0.033	0.079	-0.009	-0.161	0.173	0.233	-0.34	0.23	-33010.7	13.70
4-Chloroaniline	-0.229	0.111	0.093	0.017	-0.149	0.158	0.215	1.78	0.23	-33010.9	13.70
2-Nitroaniline	-0.071	0.114	0.083	-0.027	-0.137	0.149	0.210	-0.94	0.58	-43870.8	13.60
3-Nitroaniline	-0.194	-0.004	-0.027	-0.119	-0.229	0.045	0.104	-0.94	0.56	-43867.1	13.60
4-Nitroaniline	-0.222	0.009	-0.013	-0.056	-0.174	0.079	0.134	-0.94	0.58	-43869.9	13.60
Nitrobenzene	-0.013	0.100	0.009	-0.081	-0.194	-0.046	0.009	0.78	0.57	-38770.6	12.30
2-Nitroanisole	0.124	0.121	0.086	-0.046	-0.155	0.033	0.079	1.75	0.58	-49741.6	14.80
Benzamide	-0.201	-0.125	-0.161	-0.260	-0.301	-0.252	-0.208	0.12	0.35	-35049.2	13.70
4-Aminobenzamide	-0.409	-0.268	-0.301	-0.292	-0.328	-0.553	-0.469	-1.60	0.35	-40147.0	15.10
Acetanilide	-0.180	-0.036	-0.081	-0.174	-0.244	-0.137	-0.092	-0.29	0.31	-38633.9	15.50
4-Chloroacetanilide	-0.056	0.248	0.201	0.161	0.009	0.212	0.267	-0.52	0.31	-46938.4	17.50
Phenol	-0.387	-0.149	-0.167	-0.194	-0.284	0.017	0.072	0.57	0.22	-27003.2	11.10
3-Methylphenol	-0.237	0.025	0.000	-0.056	-0.174	0.188	0.255	2.23	0.22	-30597.6	12.90
2,3-Dimethylphenol	-0.137	0.170	0.149	0.107	-0.041	0.338	0.418	0.88	0.22	-34189.9	14.70
2,4-Dimethylphenol	-0.108	0.201	0.176	0.121	-0.022	0.350	0.427	2.70	0.22	-34190.8	14.70
Thymol	0.072	0.425	0.418	0.322	0.158	0.550	0.667	1.61	0.22	-41374.5	18.40
4-Chlorophenol	-0.292	0.127	0.121	0.149	-d	0.375	0.455	2.28	0.22	-35307.6	13.00
Catechol	-0.481	-0.260	-0.229	-0.222	-d	0.037	0.130	-0.45	0.20	-34392.7	11.70
Resorcinol	-0.538	-0.301	-0.310	-0.194	-d	0.064	0.127	-0.45	0.22	-34396.6	11.70
Hydroquinone	-0.770	-0.377	-0.409	-0.244	-d	-0.180	-0.108	-0.45	0.22	-34395.7	11.70
2-Naphthol	-0.102	0.352	0.332	0.288	-d	-c	-c	0.65	0.22	-39444.5	17.20
1,2,3-Trihydroxybenzene	-0.509	-0.387	-0.357	-0.244	-d	0.025	0.143	-1.48	0.25	-41790.5	12.40
Furan	-0.337	-0.292	-0.310	-0.387	-0.469	-0.367	-0.310	-0.23	0.19	-20434.0	7.59
2,3-Benzofuran	-0.066	0.220	0.204	0.204	0.097	0.215	0.274	-0.06	0.19	-32874.6	13.80
Quinoline	0.310	0.267	0.267	0.000	-0.041	0.017	0.072	0.29	0.16	-33547.4	15.90
Pyrrole	-0.538	-0.387	-0.398	-0.420	-0.523	-0.284	-0.237	-0.40	0.24	-18143.6	8.31
Pyrimidine	-0.301	-0.495	-0.523	-0.469	-0.569	-0.770	-b	-0.44	0.19	-22605.4	9.02
Antipyrine	0.348	0.057	0.000	-0.301	-0.310	-0.469	-0.387	0.79	0.31	-53151.7	21.40
Caffeine	-0.051	-0.081	-0.131	-0.237	-0.229	-0.481	-0.377	-1.06	0.41	-60617.7	18.90
Corticosterone	0.725	0.858	0.876	0.127	0.188	0.589	0.651	3.08	0.26	-101122	37.50
Cortisone	0.607	0.633	0.615	0.029	0.072	0.364	0.425	2.74	0.26	-107874	37.60
Hydrocortisone	0.400	0.640	0.637	0.076	0.111	0.470	0.521	2.37	0.26	-108515	38.10
Estradiol	0.377	0.926	0.917	0.486	0.461	-c	-c	2.19	0.22	-75586.9	31.40
Estratriol	0.017	0.519	0.497	0.290	0.281	-c	0.588	1.42	0.22	-82980.2	32.00
Monuron	0.025	0.212	0.161	0.068	-0.071	0.107	0.149	-0.30	0.41	-55606.3	20.60
Myrcene	0.446	0.847	0.862	1.033	0.801	0.622	0.580	3.29	0.12	-34588.2	18.60
?- Pinene	0.535	0.912	0.938	0.869	0.794	0.645	0.698	2.80	0.13	-34592.3	17.40
Geraniol	0.352	0.538	0.517	0.299	0.179	0.307	0.394	2.46	0.20	-42645.6	19.40

a k' values determined in H3PO4 solution neutralized with LiOH up to pH=7.0 using 40 mM of LPFOS (lithium perfluorooctanesulfonate) and 40 mM of LDS (lithium dodecyl sulfate), in sodium phosphate buffer at pH=7.0 using 40 mM of SDS (sodium dodecyl sulfate), 80 mM of SC (sodium cholate), 20 mM of TTAB (tetradecyltrimethylammonium bromide) and 20 mM of HTAB (hexadecyltrimethylammonium bromide) and in sodium phosphate-sodium tetraborate buffer at pH=8.0 using 40 mM of SDC (sodium deoxycholate). Dodecanophenone and methanol were used as micellar and electroosmotic flow markers and were excluded from the calculations.
b The roprted k' values are ~0 (10) and were excluded from the calculations.
c The analytes coelute with dodecanophenone and their k' values were not reported roprted in the reference (10).
d Analytes with pKa values between 9 and 10 and therefore, partially ionized at pH=8.0 (10).

phenomena such as retention in reversed phase liquid chromatography [(14]) and to the best of our knowledge, it is the only multiple linear regression model for explaining the capacity factor in MEKC. The calculated k' values for the studied data sets using LSER model were used to compare with the corresponding calculated values using the proposed model.

3. Results and dDiscussion

All data points in each experimental data set were fitted to the equations IV and V, and the back-calculated k' values were used to compute AARD values which arewere listed

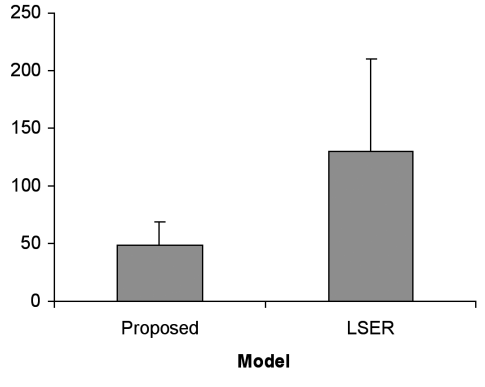

Figure 1. Mean and standard deviation of AARD values for the proposed and LSER models.

in Tables 5 and 6 along with the corresponding model constants. The proposed model shows the the lowest and the highest AARD values of 26.5% and 88.9 % were for the set numbers 2 and 9, respectively, whereas the corresponding values for the LSER model were 16.4% and 252.9 % for the set numbers 11 and 9, respectively. As it is evident from the results shown in Table 5, the proposed model provide better calculated k' values. Figure 1 shows the overall mean and standard deviations of the AARD values for the studied data sets. The difference between the mean of AARDs of the proposed model (48.5 ± 20.4) with the LSER model (130.1 ± 79.7) reveal that the difference is statistically significant (paired t-test, $p<0.01$).

Figure 2 shows the IPD distribution for the proposed and LSER models sorted in three subgroups, i.e. IPD ≤ 45%, 45%<IPD<-90% and IPD>90 %. The proposed model is able to compute k' values with error < 45 in 69 % of the cases, whereas the correponding value for the LSER model is 42 %. The frequency of the IPD >90 % for the proposed and LSER models are 8 and 32 %, respectively. Comparing these results reveal that the proposed model is able to provide better calculations in comparison with the LSER model. In addition, its descriptors could be easily computed by HyperChem using the chemical structure of the analytes of interest. In comparing the proposed model with the LSER, one should keep in mind that the LSER possesses six curve-fitting parameters, whereas those of the proposed model is five, and this could be considered as another advantage of the proposed model.

Once the model is trained using a limited number of experimental data points, the trained model could be used to predict the unmeasured k' values and the possibility of successful separation using the MEKC system under investigation could be predicted-forcasted. The main limitation of this method (and also LSER model) is that the models require a number of experimental

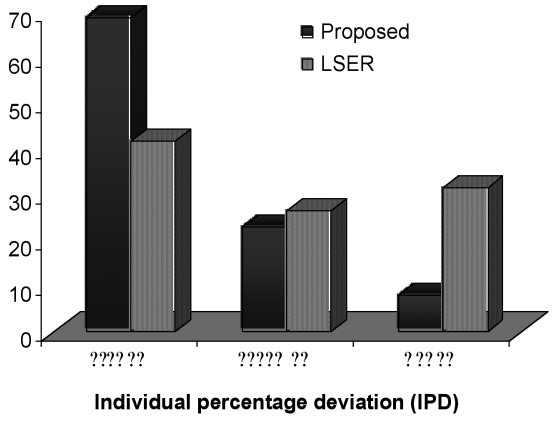

Figure 2. Distribution of individual percent deviation of the models studied.

Table 3. Details of analytes, experimental logk' values [(9]) and the chemical descriptors

Analyte	logk'	LP	CH^+	TE	PZ
CH_3COOCH_3	-1.00	-0.14	0.30	-25555.4	7.00
$CH_3COOC_2H_5$	-0.63	0.21	0.30	-29148.1	8.84
$CH_3COOC_4H_9$	0.23	1.07	0.30	-36335.5	12.50
CH_3-CO-$N(CH_3)_2$	-1.04	-0.54	0.30	-26832.1	9.55
CH_3-CO-$N(C_2H_5)_2$	-0.44	0.15	0.29	-34009.7	13.20
$C_6H_5CH_3$	0.45	1.75	0.13	-23204.2	12.30
C_6H_5-CO-CH_3	0.24	1.53	0.27	-33539.8	14.20
$C_6H_5NO_2$	0.12	0.78	0.57	-38770.6	12.30
$C_6H_5OCH_3$	0.23	0.60	0.15	-30583.6	12.90
$C_6H_5COOC_2H_5$	0.93	1.36	0.35	-44527.8	16.70
C_6H_5-CO-C_2H_5	0.56	2.16	0.27	-37132.5	16.00
$C_6H_5COOCH_2C_6H_5$	2.12	2.03	0.35	-59908.3	24.50
2-Cl-$C_6H_4NO_2$	0.51	0.56	0.57	-47068.5	14.20
$C_6H_5CH_2CN$	0.19	1.57	0.14	-30594.8	14.10
$C_6H_5CH_2$-CO-CH_3	0.17	1.46	0.23	-37133.6	16.03
$C_6H_5(CH_2)_2$-O-CO-CH_3	0.77	1.13	0.30	-48118.5	18.50
Pyridine	-0.52	0.12	0.16	-21108.5	9.73
2-Naphthylamine	0.83	-0.05	0.22	-37147.5	18.00
$C_6H_5NH_2$	-0.29	-0.12	0.22	-24706.0	11.80
$C_6H_5NHC_2H_5$	0.39	0.63	0.23	-31882.5	15.50
2-Cl-C_6H_4-C_6H_5	0.22	-0.34	0.24	-33011.0	13.70
2-NH_2-C_6H_4-C_6H_5	1.13	0.49	0.23	-43669.6	21.40
4,4'-$(NH_2)_2$-Biphenyl	0.69	2.16	0.22	-48776.9	22.80
4-NO_2-C_6H_4-NH_2	0.09	-0.94	0.58	-43869.9	13.60
C_6H_5OH	-0.28	0.57	0.22	-27003.2	11.10
3-Cl-C_6H_4OH	0.37	0.35	0.22	-35307.6	13.00
$C_6H_5CH_2OH$	-0.24	0.75	0.20	-30596.2	12.90
4-Cl-$C_6H_4CH_2OH$	0.36	0.53	0.20	-38903.5	14.80
1,3-$C_6H_4Cl_2$	1.11	1.15	0.16	-36218.4	14.30
Biphenyl	1.69	3.73	0.13	-38584.0	20.10
CH_3SOCH_3	-1.58	-0.68	1.39	-19680.0	5.01
Naphthalene	1.15	1.67	0.13	-32051.0	16.62
1,3,5-$C_6H_3(CH_3)_3$	1.31	2.06	0.13	-30393.2	15.94
1,2,4,5-$C_6H_2Cl_4$	2.01	0.71	0.17	-52823.1	18.15
$C_6H_5(CH_2)_2C_6H_5$	2.16	3.00	0.13	-45771.7	23.76
$C_6H(CH_3)_5$	1.86	2.36	0.13	-37576.3	19.61

data for training process, and the trained models are only valid for the same analytical conditions which training points are collected. and Aanother model should be trained when one or more variables of MEKC system arewere modified. To evaluate the stability of the proposed model, one- leave- out method was employed and the obtained results showed that there are no significant changes in the numerical values of the model constants and also AARD values, and these mean that the model is robust and adding

Table 4. Details of analytes, experimental logk' values [(7]) and the chemical descriptors

	logk'[a]	LP	CH^+	TE	PZ
Pyrrole	-0.73	-0.40	0.24	-18143.6	8.31
Phenol	-0.25	0.57	0.22	-27003.2	11.10
Nitrobenzene	0.15	0.57	0.57	-38770.6	11.10
2-Nitronisole	0.33	-2.14	0.58	-49735.2	14.80
Ethylbenzene	0.87	2.15	0.13	-26796.9	14.10
Furan	0.53	-0.23	0.19	-20434.0	7.59
4-Nitroaniline	0.09	-2.82	0.58	-43869.9	13.60
2,3-Benzofuran	0.60	-0.06	0.19	-32874.6	13.80
2-Naphthol	0.91	0.65	0.22	-39444.5	17.30
Benzaldehyde	0.07	0.57	0.22	-29946.9	12.40
Chlorobenzene	0.60	1.37	0.15	-27914.3	12.40
Resorcinol	-0.52	-0.45	0.22	-34396.6	11.70
3-Nitroaniline	0.07	-0.94	0.56	-43867.1	13.60
2-Nitroaniline	0.33	-0.94	0.58	-43870.8	13.60
4-Chlorophenol	0.41	0.35	0.22	-35305.6	13.00
p-Xylene	0.93	1.90	0.13	-26798.7	14.10
Aniline	-0.34	1.26	0.22	-24705.9	11.80
Acetanilide	-0.06	-0.29	0.31	-38633.9	15.50
Benzonitrile	0.07	1.32	0.15	-27001.2	12.30
Methyl phenyl ether	0.26	0.60	0.15	-30583.7	12.90
Toluene	0.48	1.75	0.13	-23204.2	12.30
Benzophenone	1.48	2.68	0.31	-48910.9	22.00
Benzene	0.03	1.60	0.13	-19609.7	10.40
Naphthalene	1.20	1.67	0.13	-32051.0	16.60
3-Methylphenol	0.14	2.23	0.22	-30597.6	12.90
Bromobenzene	0.75	1.65	0.15	-27441.6	13.10
Propylbenzene	1.36	2.54	0.13	-30390.6	15.90
Pyrimidine	-0.94	-0.44	0.19	-22605.4	9.02
Methyl benzoate	0.57	1.14	0.25	-40947.1	14.80
Pyridine	-0.50	0.12	0.16	-21108.5	9.73

[a] k' values determined in sodium phosphate-sodium tetraborate, pH=8.0 using 50 mM of SDS (sodium dodecyl sulfate).

and/or deleting one datum in a data set could not affect the accuracy of the model. The produced AARD by the proposed model is relatively high, when it is compared to the experimental relative standard deviation (RSD) for repeated experiments. However, it is a fact that the proposed model is able to improve the accucary of k' calculations by a factor of 2.7 and is a step forward in MEKC data modeling. It is obvious that the efforts should be continued until providing a more accurate model with AARD comparable with RSD values.

References

[1] Herbert BJ, Dorsey JG., n-Octanol-water partition coefficient estimation by micellar electrokinetic capillary chromatography., *Anal. Chem.* 1995; 67: 744-749.

[2] McKillop AG, Smith RM, Rowe RC, Wren AC., Modeling and prediction of electrophoretic mobilities in capillary electrophoresis: Separation of alkylpyridines., *Anal. Chem.* 1999; 71: 497-503.

[3] Jalali-Heravi M, Garkani-Nejad Z., Prediction of electrophoretic mobilities of alkyl- and alkenylpyridines in capillary electrophoresis using artificial neural networks., *J. Chromatogr. A* 2002; 971: 207-215.

[4] Liang HR, Vuorela H, Vuorela P, Riekkola MJ, Hiltunen R., Prediction of migration behaviour of flavonoids in capillary zone electrophoresis by means of topological indices., *J. Chromatogr. A* 1998; 798: 233-242.

[5] Li Q, Dong D, Jia R, Chen X, Hu Z, Fan BT.,

Table 5. Absolute average relative deviation (AARD), the model constants and their standard errors (s.e.) of the proposed model for studied sets.

No.	Surfactant[a]	N[b]	AARD	J_0	s.e. of J_0	J_1	s.e. of J_1	J_2 (?10^{-3})	s.e. of J_2 (?10^{-3})	J_3	s.e. of J_3	J_4	s.e. of J_4
1	SDS	80	70.7	-2.366	0.252	0.451	0.051	-0.004[c]	0.008	0.086	0.019	1.592	0.493
2	LPFOS	64	26.5	-0.884	0.078	0.099	0.021	0.022	0.005	0.084	0.014	1.108	0.216
3	LDS	65	29.2	-0.816	0.083	0.094	0.022	0.031	0.005	0.115	0.015	1.018	0.230
4	SDS	66	30.7	-0.839	0.084	0.092	0.023	0.032	0.005	0.118	0.015	0.961	0.236
5	SC	65	41.1	-0.571	0.112	0.072	0.029	0.033	0.007	0.103	0.020	0.743	0.312
6	SDC	58	36.6	-0.933	0.126	0.070	0.030	0.042	0.007	0.135	0.021	1.132	0.332
7	TTAB	58	52.4	-0.431	0.153	0.108	0.038	0.012[c]	0.009	0.050[c]	0.027	0.331[c]	0.392
8	HTAB	56	45.2	-0.158[c]	0.147	0.108	0.034	0.004[c]	0.008	0.024[c]	0.025	-0.083[c]	0.366
9	SDS	36	88.9	-1.869	0.314	0.212	0.093	-0.022[c]	0.014	0.098	0.037	-0.517[c]	0.359
10	SDS	30	63.3	-1.912	0.272	0.156	0.068	0.009[c]	0.010	0.171	0.031	0.473[c]	0.665

[a]Other experimental details are the same as reported in Tables 1-4.

[b] N is the number of data in each set.

[c] The p values for t-test of these constants are >0.05, therefore, these constants are not able to improve the accuracy of the model and could be considered as equal to zero.

Table 6. Absolute average relative deviation (AARD) and the model constants of the LSER model for studied sets

No.	Surfactant[a]	N[b]	AARD	c	v	r	s	a	b
1	SDS	80	24.7	-[c]	-[c]	-[c]	-[c]	-[c]	-[c]
2	LPFOS	64	127.6	-1.410	1.966	-0.113	-0.243	-0.876	-0.455
3	LDS	65	203.4	-1.575	2.609	0.586	-0.595	-0.317	-1.565
4	SDS	66	195.4	-1.680	2.717	0.558	-0.596	-0.266	-1.674
5	SC	65	87.7	-1.408	2.274	0.691	-0.693	0.117	-1.938
6	SDC	58	78.6	-1.833	2.422	0.926	-0.867	0.070	-1.785
7	TTAB	58	121.3	-1.851	2.634	0.902	-0.617	-0.766	-2.410
8	HTAB	56	252.9	-1.833	2.710	1.112	-0.755	0.824	-2.437
9	SDS	36	192.6[d]	-0.610	0.023	-[e]	-0.670	0.100	-2.050
10	SDS	30	16.4	-1.759	2.982	0.348	-0.427	-0.021	-2.024

[a] Other experimental details are the same as reported in Tables 1-4.

[b] N is the number of data in each set.

[c] The model constants were not given in the reference [(8]) and the APD values were computed using predicted reported k' values [(8]).

[d] The APD values calculated using the reported LSER constants in the reference gives APD of 5720.8 %, and it should be due to a typographical error in reporting the model constants. We used the reported LSER constants with opposite signs and obtained 192.6

Development of a quantitative structure-property relationship model for predicting the electrophoretic mobilities., *Comp. Chem.* 2002; 26: 245-251.

[6] Jouyban A, Yousefi BH., A quantitative structure property relationship study of electrophoretic mobility of analytes in capillary zone electrophoresis., *Comput. Biol. Chem*. 2003; 27: 297-303.

[7] Roses M, Rafols C, Bosch E, Martinez AM, Abraham MH., Solute-solvent interactions in micellar electrokinetic chromatography. Characterization of sodium dodecyl sulfate-Brij 35 micellar systems for quantitative structure-activity relationship modeling. *J. Chromatogr. A* 1999; 845: 217-226.

[8] Kelly KA, Burns ST, Khaledi MG., Prediction of retention in micellar electrokinetic chromatography from solute structure. 1. Sodium dodecyl sulfate micelles. *Anal. Chem.* 2001; 73: 6057-6062.

[9] Taillardat-Bertschinger A, Carrupt PA, Testa B., The relative partitioning of neutral and ionized compounds in sodium dodecyl sulfate micelles measured by micellar electrokinetic capillary chromatography. *Eur. J. Pharm. Sci.* 2002; 15: 225-234.

[10] Fuguet E, Rafols C, Bosch E,, Abraham MH, Roses M., Solute-solvent interactions in micellar electrokinetic chromatography. III. Characterization of the selectivotyy of micellar electrokinetic chromatography systems. *J. Chromatogr. A* 2002; 942: 237-248.

[11] HyperChem 7.0., Molecular mechanics and quantum chemical calculations package, 2002, HyperCube Inc., Ontario, Canada.

[12] Fletcher R., *Practical methods of optimization*, 1980, Wiley, New York, USA.

[13] Dewar MJS, Zoebish EG, Healy EF, Stewart JJP., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model., *J. Am. Chem. Soc.* 1985; 107: 3902-3909.

[14] Abraham MH, Chadha HS Leitao RAE, Mitchell RC, Lambert WJ, Kaliszan R, Nasal A, Haber P., Determination of solute lipophilicity, as log P(octanol) and log P(alkane) using poly(styrene–divinylbenzene) and immobilised artificial membrane stationary phases in reversed-phase high-performance liquid chromatography., *J. Chromatogr. A* 1997; 766: 35-47.