Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1002 (2003) 221–229
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Quantitative structure–property relationship studies of migration index in microemulsion electrokinetic chromatography using artificial neural networks

M.H. Fatemi
Department of Chemistry, Faculty of Basic Science, Mazandaran University, P.O. Box 453, Babolsar, Iran

Received 25 February 2003; received in revised form 16 April 2003; accepted 16 April 2003

Abstract

Artificial neural networks (ANNs) were successfully developed for the modeling and prediction of migration indices of the 53 benzene derivatives and heterocyclic compounds in microemulsion electrokinetic chromatography. The selected descriptors that appear in multiple linear regression models are: 3D-MoRSE signal 25 unweighted, 3D-MoRSE signal 19 weighted by atomic Sanderson electronegativity, *R* maximal autocorrelation index lag 1 weighted by atomic mass (R_1M^+), *R* maximal autocorrelation index lag 2 weighted by polarizability (R_2P^+) and average atomic composition index. These descriptors were used as inputs for generated 5-4-1 networks. After training and optimization of the ANN parameters it was used to prediction of migration index of the test set compounds. The results obtained using ANNs were compared with the experimental values as well as with those obtained using regression models and showed the superiority of ANNs over regression models.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Structure–property relationships; Migration index; Microemulsion electrokinetic chromatography; Neural networks, artificial; Regression analysis

1. Introduction

Microemulsions are transparent liquids that generally consist of a surfactant, a cosurfactant such as a medium-chain alcohol, oil and water. These dispersion systems are thermodynamically stable [1]. Watarai used oil-in-water (O/W) microemulsions of water–sodium dodecyl sulfate (SDS)–1-butanol–heptane as a pseudo stationary phase in electrokinetic chromatography [2] and called it microemulsion electrokinetic chromatography (MEEKC). Analogous to micellar electrokinetic chromatography, the retention factor, *k*, of a neutral solute in MEEKC can be calculated as follows:

$$K = (t_s - t_0)/t_0 \cdot (1 - t_s/t_m) \quad (1)$$

where t_0, t_s and t_m are the migration times of the electroosmotic flow, the solute and the microemulsion, respectively.

The migration index (MI) scale that was introduced by Muijselear et al. in micellar electrokinetic chromatography [3] applied to MEEKC with some modifications by the Ishihama et al. [4]. They defined the MI scale for a solute in MEEKC as follows:

$$\mathrm{MI} = c\log k + d \quad (2)$$

E-mail address: mhfatemi@umz.ac.ir (M.H. Fatemi).

0021-9673/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0021-9673(03)00687-3

where c and d are the slope and the intercept of a calibration line between log k values of reference solutes such as alkyl benzenes and their migration indices, respectively. The Migration index scale can be applied for all neutral solutes that migrate in the range from t_0 to t_m and this might be independent of the volume of the microemulsion. The values of migration index can be used as a hydrophobic parameter and were used in quantitative structure–property relationship (QSPR) studies.

Although the experimental determination of MI is time consuming and requires high-purity samples and skilled operators, so the development of alternative methods such as QSPRs would be useful for the theoretical calculation of MI values. In addition these investigations can be help to better understand the migration behavior of molecules in MEEKC. QSPRs have been used extensively to explain separation mechanisms, predict retention behavior and characterize the physicochemical properties of solutes in thin-layer chromatography [5], gas chromatography [6] and high-performance liquid chromatography [7,8]. Also there are some reports on QSPR studies in capillary electrophoresis [9–11]. Fu and Lucy developed empirical expressions for the prediction of electrophoretic mobility of monoamines and carboxylic acids [12,13]. They correlated the mobility of analytes with the molecular mass, molar volume and dissociation constant using non-linear equations. Also, Liang et al. studied the correlation between the electrophoretic mobility of 13 flavonoids and their topological indices [14]. Artificial neural networks (ANNs) have been applied to a wide variety of chemical problems such as prediction of ^{13}C nuclear magnetic resonance (NMR) chemical shift [15], selectivity coefficients of ion selective electrodes [16], simulation of MS spectra [17] and QSPR investigations [18–22]. Also Jalali-Heravi and Garakani-Nejad used artificial neural networks for the prediction of electrophoretic mobilities of sulfonamides in capillary zone electrophoresis [23]. They used heat of formation, most positive partial charge and molecular surface area in their QSPR models.

In the present study an ANN was employed to generate a QSPR model between the molecular based structural parameters (descriptors) and observed migration indices of some benzene derivatives and heterocyclic compounds in MEEKC.

2. Methods

2.1. Data set

The data set of migration indexes in MEEKC was taken from the values reported by Ishihama et al. [4]. The molecules in the data set including benzene derivatives and heterocyclic compounds are shown in Table 1. The migration indices of all the molecules included in the data set were obtained under the same conditions. The migration indices fall in the range of 1 to 9.9 for pyrimidine and anthracene, respectively. The data set was randomly divided in two groups; a training set and a prediction set consisting of 43 and 10 molecules, respectively. The training set was used for the model generation and the prediction set was used for the evaluation of the generated model.

2.2. Descriptors

The MI value is related to molecular structure in a complex way. It depends on the size of molecule and the strength of interactions between the solute and solvent molecules. The molecular descriptors used to search for the best model of the migration index were calculated by the Dragon program [24] on the basis of the minimum energy molecular geometries optimized by the MOPAC package [25] (AM1 semiempirical method). Dragon is a new, freely available software (by Milano Chemometrics and the QSAR Research Group) for the calculation of more than 800 molecular descriptors. Subsequently, the method of stepwise multiple linear regression (MLR) was used to select the most important descriptors and to calculate the coefficients relating the descriptors to migration index. The descriptors that appear in the best MLR equation are shown in Table 2. These descriptors are: 3D-MoRSE signal 25 unweighted, 3D-MoRSE signal 19 weighted by atomic Sanderson electronegativity, R maximal autocorrelation index lag 1 weighted by atomic mass (R_1M^+), R maximal autocorrelation index lag 2 weighted by polarizability (R_2P^+) and average atomic composition index. These descriptors were used as inputs for generated ANNs.

A detailed description of the theory behind of these descriptors has been adequately described elsewhere [26]. Many physical, chemical and bio-

Table 1
Data set and corresponding observed and predicted values of the MI[a]

No.	Name	MI exp	MI ann	MI MLR	E_{rel} (%)
Training set					
1	Methylpyrazine	2.15	2.49	2.88	1.49
2	*N*-Methylbenzamide	3.67	4.04	4.05	3.04
3	1-Methylpyrrole	3.95	4.03	4.43	3.03
4	*p*-Methoxyphenol	4.02	4.53	4.58	3.53
5	Benzonitrile	4.51	4.28	4.36	3.28
6	*o*-Creosol	5.09	4.90	5.12	3.90
7	Benzene	5.93	6.19	5.92	5.19
8	*p*-Chlorophenol	6.26	6.38	5.94	5.38
9	3-Methylindole	6.69	6.45	6.32	5.45
10	Propylbenzene	9.10	9.11	8.49	8.11
11	Propiophenone	5.83	5.52	5.24	4.52
12	*p*-Propylphenol	7.51	7.32	7.20	6.32
13	4,6-Dimethylpyrimidine	2.65	2.27	2.74	1.27
14	Benzaldehyde	4.53	4.25	4.34	3.25
15	*m*-Cresol	5.09	4.91	5.00	3.91
16	Butylbenzene	9.89	9.78	9.78	8.78
17	Acetophenone	4.87	4.72	5.11	3.72
18	Benzyl alcohol	3.92	4.19	4.32	3.19
19	Butyrophenone	6.75	6.36	6.68	5.36
20	Methyl benzoate	5.82	5.71	4.99	4.71
21	Ethylpyrazine	3.17	2.74	3.17	1.74
22	1-Methylindole	6.68	6.32	6.53	5.32
23	2,6-Dimethylpyrrole	4.20	4.63	5.29	3.63
24	Pyrimidine	1.00	0.87	1.34	−0.13
25	Ethyl 2-furoate	4.67	5.15	5.65	4.15
26	Anthracene	9.90	9.69	9.98	8.69
27	Thiophene	5.23	5.17	5.02	4.17
28	2-Naphthol	6.77	6.63	6.55	5.63
29	Quinoxaline	4.21	3.84	3.33	2.84
30	2-Methylindole	6.24	6.32	7.17	5.32
31	Pyrazine	1.28	1.37	1.18	0.37
32	Indole	5.69	5.71	5.88	4.71
33	Naphthalene	8.19	8.34	7.95	7.34
34	Nitrobenzene	5.15	5.63	5.09	4.63
35	Resorcinol	3.12	3.27	3.31	2.27
36	*p*-Nitrotoluene	6.03	2.49	2.88	4.84
37	2-Naphthol	7.26	5.84	5.49	6.38
38	Methyl 2-furoate	3.74	7.38	8.36	2.74
39	Furan	4.17	3.74	3.05	3.01
40	Toluene	7.03	4.01	3.64	5.94
41	*p*-Nitroanisol	5.54	6.94	6.61	4.50
42	Acetanilide	3.93	5.50	5.04	3.24
43	*p*-Cresol	5.17	4.24	4.54	3.92
Prediction set					
1	2-Ethylfuran	6.63	6.92	7.11	5.92
2	*p*-Nitroaniline	4.37	3.89	3.23	2.89
3	4-Methylpyrimidine	1.90	2.18	2.25	1.18
4	Pyrrole	3.03	3.02	1.89	2.02
5	Phenol	4.29	4.89	4.90	3.89
6	2-Methylfuran	5.40	6.10	5.75	5.10
7	Anisol	5.79	5.48	5.67	4.48
8	*p*-Ethylphenol	6.36	6.39	6.50	5.39
9	Benzofuran	6.85	6.54	6.40	5.54
10	Ethylbenzene	8.05	8.52	7.94	7.52

[a] exp refers to experimental; ann refers to artificial neural network; MLR refers to multiple linear regression determined values of migration index; E_{rel} (%) represents the relative error between ann predicted and experimental determined values of migration index.

Table 2
Specification of multiple linear regression models

Descriptor	Notation	Coefficient
3D-MoRSE signal 25 unweighted	MoRSE-25	2.972 (±0.288)
3D-MoRSE signal 19 weighted by atomic Sanderson electronegativity	MoRSE-19	3.418 (±0.301)
R maximal autocorrelation index lag 1 weighted by atomic mass	R_1M^+	−18.120 (±02.035)
R maximal autocorrelation index lag 2 weighted by polarizability	R_2P^+	−12.329 (±3.179)
Average atomic composition index	AAC	−4.192 (±0.689)
Constant		6.359 (±1.123)

logical properties of compounds are dependent on the three-dimensional (3D) arrangement of the atoms in a molecule. Consideration of the 3D structure of organic compounds in qualitative structure–activity relationship (QSAR) studies has been hindered by the lack of data on the three-dimensional structure of the compounds to be considered. The three-dimensional structure of molecule can experimentally be derived from electron or X-ray diffraction studies or from NMR measurements of nuclear Overhouser effects (NOEs). Schuur and Gasteiger recently developed a mathematical transformation on the molecular 3D structure that gives a fixed number of variables for 3D structure representation built on equations used in the analysis of the intensity distributions obtained in electron diffraction experiments [26–28]. The new 3D structure code was therefore, named 3D molecular representation of structures based on electron diffraction code (3D-MoRSE code). They can be calculated by summing atomic properties viewed by different angular scattering functions. After some modifications and simplification to the main equation they reported the following equation for the calculation of these codes:

$$I(s) = \sum_{i=2}^{n}\sum_{j=1}^{i-1} A_i A_j \cdot \frac{\mathrm{Sin}(sr_{ij})}{sr_{ij}}$$

$$S = 0, \ldots, 31.0\ \text{Å}^{-1} \qquad (3)$$

In this equation $I(s)$ represent the 3D-MoRSE code, A_i and A_j are any atomic properties of atom i and j, r_{ij} is interatomic distance between atom i and j, n is the number of atoms in molecule and S is a reciprocal distance. The value of S was considered only at discrete positions within a certain range. In many applications 32 equidistant values were chosen from 0 to 31 Å^{-1}. So the function $I(s)$ is discrete, reporting its values only at equally spaced values of S within a certain range. Thus the entire 32 3D-MoRSE values span a 32-dimensional space where each structure corresponds to a point in this space. Also we can use some atomic properties A_i, like atomic number, atomic mass, partial atomic charge, atomic electronegativites and atomic polarizabilities in Eq. (3). The possibility for choosing an appropriate atomic property give great flexibility to the 3D-MoRSE code for adapting it to the problem under investigation. The 3D-MoRSE codes have great potential for representation of molecular structure. It is worth noting that they reflects the three-dimensional arrangement of the atoms of a molecule and do not care about chemical bonds.

Other newly developed 3D descriptors are geometry topology and atomic weight assembly (GETAWAY) descriptors that were presented by Consonni et al. [29,30]. They encode geometrical information that given from influence matrix, topological information given by molecular graph and chemical information from selected atomic properties. One type of these molecular descriptors is R-GETAWAY and represented by $R_k(w)$ that calculated as follows. The molecular influence matrix, denoted by **H** and resembles the leverage (or influence) matrix defined in regression diagnostics [29]. The value of **H** was calculated from the molecular matrix **M** (**M** has A rows corresponding to the atoms in a molecule and three column corresponding to the Cartesian coordinates x, y, z of each atom in optimized molecular structure) as follows:

$$\mathbf{H} = \mathbf{M}(\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1} \cdot \mathbf{M}^{\mathrm{T}} \qquad (4)$$

where the superscript T refers to the transposed matrix. The diagonal elements h_{ii} of the **H** matrix, called leverage, encode atomic information and

represent the ‘‘influence’’ of each molecule atom in determining the hole shape of molecule; for example mantle atoms always have higher h_{ii} values than atoms near the molecule center. Moreover, the magnitude of the maximum leverage in the molecule depends on the size and shape of the molecule itself. Lower leverage can be found for atoms in molecules of spherical shape, while higher leverage for atoms in more linear molecules.

The autocorrelation R-GETAWAY descriptors have been defined based on this matrix as follows:

$$R_k(w) = \sum_{i=1}^{A-1} \sum_{j>i} \frac{\sqrt{h_{ii}.h_{jj}}}{r_{ij}} \cdot W_i W_j \delta(k,d_{ij}) \quad k = 1,2,d \quad (5)$$

where $R_k(w)$ is the w-weighted kth order of autocorrelation index, r_{ij} is the 3D geometric distances between each pair of atoms i and j, d_{ij} is the topological distance between atoms i and j, d is the topological diameter, h_{ii} and h_{jj} are diagonal terms of the **H** matrix and δ, is a Dirac-delta function defined as:

$$\delta(Kd_{ij}) = \begin{cases} 1 \text{ if } d_{ij} = k \\ 0 \text{ if } d_{ij} \neq k \end{cases} \quad (6)$$

To take into account local; aspect of the molecule, from Eq. (5) the maximal contribution to the autocorrelation at each lag (i.e., topological distance) has also been proposed as a molecular descriptor:

$$R_k^+(w) = \max_{ij} \left[\frac{\sqrt{h_{ii}.h_{jj}}}{r_{ij}} \cdot W_i W_j \delta(k,d_{ij}) \right]$$
$$i \neq j \text{ and } k = 1,2,\ldots,d \quad (7)$$

where $R_k^+(w)$ is the w-weighted kth order maximal R index. These descriptors can match 3D-molecular geometry, atom relatedness and chemical information.

The average atomic compositional (AACs) indices are molecular zero dimensional descriptors that are derived from the chemical formula of a molecule and represent the mean information on the elemental composition of the molecule and are calculated as:

$$\text{AAC} = -\sum \frac{A_g}{A^h} \cdot \log_2 \frac{A_g}{A^h} \quad (8)$$

where A^h is the total number of atoms (hydrogen included) and A_g is the number of equal-type atoms in the gth equivalence class [26].

2.3. ANN generation

A detailed description of the theory behind a neural network has been adequately described in the literature [31–33]. In addition we report some relevant principles of ANNs in previous papers [17,18,21,22,34]. The program for the feed-forward neural network that was trained by a back-propagation algorithm was written in FORTRAN 90. This network has five nodes in the input layer and one node in the output layer. Descriptors that appeared in the selected MLR model were used as inputs for the generated ANN and its output was the migration index for molecule of interest. The number of nodes in the hidden layer would be optimized. The initial weights were randomly selected from a uniform distribution that ranged between -0.3 and $+0.3$. The initial bias values were set to be one. These values were optimized during the network training. The value of each input was divided into its mean value to bring the values of the input variables into the dynamic range of the sigmoid transfer function in the ANN. Before training, the network was optimized for the number of nodes in the hidden layer, learning rates and momentum. Then the network was trained using the training set to optimize the values of weights and biases. Finally in order to evaluate the prediction power of the ANN, trained network was employed to calculate the migration indexes for the prediction set.

3. Results and discussion

The data set and corresponding observed and ANN predicted values of the migration indexes of all molecules studied in this work are shown in Table 1. Table 2 shows the best MLR models. It can be seen from this table that five descriptors are used in MLR model. These descriptors are: 3D-MoRSE signals 25 unweighted, 3D-MoRSE signals 19 weighted by atomic Sanderson electronegativity, R maximal autocorrelation index lag 1 weighted by atomic mass (R_1M^+), R maximal autocorrelation index lag 2 weighted by polarizability (R_2P^+) and average

atomic composition index. Each of these variables encodes different aspects of the molecular structure. Among the different factors affecting the migration behavior of molecules in MEEKC, mass, size, bulkiness and electronic parameters are most important. The appearance of 3D-MoRSE and GETAWAY descriptors that encode the 3D structure of a molecule reveals the role of bulkiness and steric interactions of a solute in MEEKC. Although the fact that some of these descriptors weighted by polarizability and electronegativity represent the role of electronic interactions in migration behavior of a molecule in MEEKC. The calculated values of these descriptors are shown in Table 3 for all the molecules included in the data set.

The next step was the generation of the artificial neural network. Before training the network, the parameters of the number of nodes in the hidden layer, weights and biases learning rates and momentum values were optimized. The procedure for the optimization of these parameters is reported in our previous papers [17,18]. Table 4 shows the architecture and specifications of the optimized ANNs. After the optimization of the ANN parameters, the network was trained using a training set for the adjustment of weights and biases values. To control the overfitting of the network during the training procedure, the values of standard error of calibration (SEC) and standard error of prediction (SEP) [35] were calculated and recorded to monitor the extent of the learning after each 500 iterations. Obtained results showed that after 38 500 iterations, the SEP values started to increase and overfitting began. To maintain the predictive power of the network at a desirable level training was stopped at this point. Based upon the high values of iterations two points may arise. First, the architecture of the generated ANN was correctly designed and second the descriptors appeared in the MLR models have been adequately chosen.

For the evaluation of the predictive power of the network, a trained ANN was used to predict the migration indexes of the molecules included in the prediction set. Table 1 represents the experimental and ANN predicted values of migration indexes for the training and prediction set compounds. Table 5 compares the statistical parameters for ANN and MLR models. The correlation coefficients and standard error values of these models shows the superiority of the ANN over the MLR model. The standard errors of calibration and prediction for the MLR model are 0.510 and 0.602 should be compared with the values of 0.260 and 0.421, respectively, for the ANN model. In the case of the ANN, the

Table 3
The values of the descriptors that were used in this work[a]

No.[b]	Morse-25	MoRSE-19	R_1M^+	R_2P^+	AAC
Training set					
1	0.067	0.387	0.14	0.103	1.46
2	0.849	0.312	0.071	0.098	1.483
3	0.543	0.415	0.11	0.128	1.296
4	0.8	0.185	0.111	0.097	1.402
5	0.635	0.299	0.1	0.117	1.296
6	0.588	0.48	0.1	0.094	1.272
7	0.64	0.535	0.077	0.115	1
8	0.578	0.082	0.329	0.149	1.614
9	0.827	0.741	0.077	0.101	1.245
10	0.723	1.255	0.072	0.101	0.985
11	0.521	0.737	0.072	0.099	1.234
12	0.717	0.901	0.11	0.099	1.207
13	−0.094	0.437	0.128	0.083	1.406
14	0.585	0.368	0.089	0.105	1.296
15	0.613	0.361	0.109	0.095	1.272
16	0.771	1.654	0.072	0.108	0.98
17	0.617	0.6	0.088	0.117	1.264
18	0.538	0.439	0.079	0.104	1.272
19	0.626	1.06	0.072	0.098	1.209
20	0.787	0.488	0.086	0.1	1.392
21	0.078	0.485	0.148	0.137	1.406
22	0.812	0.809	0.073	0.09	1.245
23	0.458	0.711	0.089	0.092	1.248
24	0.193	−0.144	0.146	0.123	1.522
25	0.854	0.669	0.07	0.097	1.357
26	1.835	0.426	0.065	0.08	0.98
27	0.337	0.398	0.351	0.341	1.392
28	1.184	0.175	0.103	0.086	1.236
29	0.43	0.169	0.115	0.097	1.406
30	0.756	0.995	0.092	0.096	1.245
31	−0.004	0.019	0.145	0.119	1.522
32	0.793	0.503	0.103	0.099	1.272
33	1.209	0.459	0.073	0.087	0.991
34	0.505	0.504	0.221	0.102	1.727
35	0.647	−0.042	0.117	0.096	1.449
36	0.47	0.654	0.203	0.093	1.658
37	0.611	0.39	0.345	0.158	1.325
38	0.371	0.182	0.139	0.102	1.522
39	0.322	0.172	0.168	0.125	1.392
40	0.709	0.6	0.085	0.107	0.997
41	0.637	0.417	0.198	0.09	1.722
42	0.716	0.533	0.09	0.103	1.483
43	0.552	0.407	0.116	0.099	1.272

Table 3. Continued

No.[b]	Morse-25	MoRSE-19	R_1M^+	R_2P^+	AAC
Prediction set					
1	0.424	0.895	0.197	0.129	1.273
2	0.348	0.203	0.206	0.086	1.811
3	−0.09	0.266	0.149	0.095	1.46
4	0.284	0.014	0.104	0.135	1.361
5	0.617	0.314	0.129	0.108	1.314
6	0.424	0.527	0.19	0.123	1.325
7	0.802	0.444	0.102	0.103	1.272
8	0.759	0.647	0.111	0.098	1.236
9	0.783	0.443	0.148	0.105	1.273
10	0.786	0.978	0.077	0.102	0.991

[a] The definitions of the descriptors are given in Table 2.
[b] The numbers refers to the numbers of the molecules given in Table 1.

Table 4
Architecture and specification of the generated ANNs

No. of nodes in the input layer	5
No. of nodes in the hidden layer	4
No. of nodes in the output layer	1
Weights learning rate	0.7
Biases learning rate	0.02
Momentum	0.45
Transfer function	Sigmoid

maximum and minimum relative errors for the predicted migration indices are 7.25 and 1.18% for ethyl benzene and pyrrole, respectively. However, it is worth noting that these values are in agreement with the results obtained by experiments.

In order to enforce the credibility of the obtained results different training and prediction sets were chosen and the network was trained using these training sets. In this procedure a set of 11 compounds out of 53 molecules was chosen randomly as a prediction set each time. Then an ANN model was generated and trained with the remaining molecules and the MIs of removed molecules were predicted using this model. This procedure was repeated four times. The results obtained are included in Table 6

Table 5
Statistical parameters obtained using the ANN and MLR models[a]

Model	SEC (%)	SEP (%)	R_t	R_p	F_t	F_p
ANN	0.260	0.421	0.991	0.977	2351	170
MLR	0.510	0.602	0.968	0.953	609	79

[a] t refers to the training set; p refers to the prediction set; R is the correlation coefficient; F is the statistical F value.

Table 6
Comparison of the SEC and SEP of the selected model with the test models obtained using different molecules

Model	SEC (%)	SEP (%)	R_t	R_p
Selected model	0.260	0.421	0.991	0.977
Test model I	0.251	0.415	0.993	0.981
Test model II	0.273	0.443	0.990	0.972
Test model III	0.267	0.453	0.991	0.967
Test model IV	0.255	0.418	0.992	0.983

for these test models. As can be seen from this table, the results do not depend on the molecules of the training and predictions set and therefore, are not due to chance. Fig. 1 shows the plot of the ANN predicted versus the experimental values for the migration indexes of the training and prediction set. The residuals of the ANN calculated values of the migration indexes are plotted against the experimental values in Fig. 2. The propagation of the residuals in both sides of zero line indicates that no systematic error exists in the development of the ANN.

Abraham et al. have applied a linear free energy relationship (LFER) to the same data set [36]. In this model they correlate experimentally determined solute solvatochromic parameters to retention factors in MEEKC. Comparison between results obtained by Abraham et al. and the present work reveals that the statistical parameters of these models did not have any significant differences. It is worth nothing that

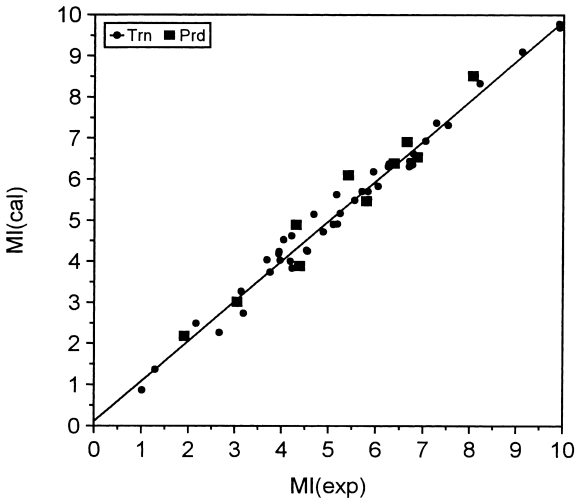

Fig. 1. Plot of the calculated migration indices against the experimental values.

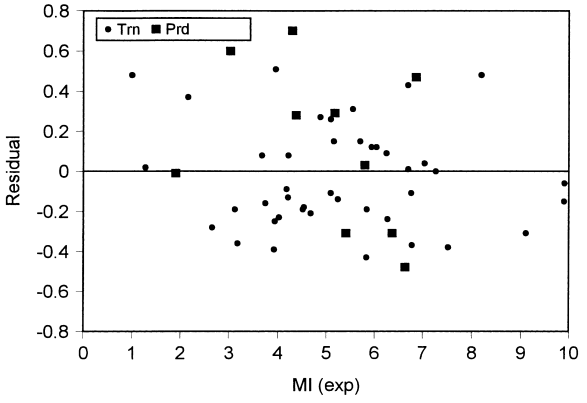

Fig. 2. Plot of the residuals versus experimental values of migration indices.

the model constructed in the presence work used five theoretical calculated descriptors instead of five experimental determined solute solvatochromic parameters proposed in Abraham et al.'s model.

4. Conclusions

The results of this study demonstrate that the QSPR method using ANN techniques can generate suitable models for the prediction of migration indexes in MEEKC. The key strength of the neural networks is their ability to allow for flexible mapping of the selected features by manipulating their functional dependence implicitly, unlike regression analysis. Neural network handles both linear and nonlinear relationships without adding complexity to model. This capability offset the larger computing time required and complexity of the ANN method with respect to MLR.

Descriptors that appear in the MLR models and are included in the ANN provide information related to the different molecular properties that can participate in the physiochemical process that occurs in microemulsion electrokinetic chromatography. From the analysis, we can conclude that the GETAWAY and 3D-MoRSE descriptors have an overall good modeling capability, providing their usefulness in QSPR studies. These descriptors contain local or distributed information molecular structure, so in most case more than one type of these descriptors are needed to reach an acceptable modeling power.

Acknowledgements

This work was supported by the research deputy of Mazandaran University. The author also acknowledges M. Haghdadi for her helpful discussion.

References

[1] K. Shinoda, B. Lindman, Langmuir 3 (1987) 135.
[2] H. Watarai, Chem. Lett. 20 (1991) 391.
[3] P.G.H. Muijselaar, H.A. Claessens, C.A. Cramers, Anal. Chem. 66 (1994) 635.
[4] Y. Ishihama, Y. Oda, K. Uuchikawa, N. Asakawa, Anal. Chem. 67 (1995) 1588.
[5] F.J. Garcia-March, G.M. Anth-fos, F. Prez-Gimenez, M.T. SalabertiSalvador, R.A. Cercs-del-Pozo, J.V. de Julian-Oritz, J. Chromatogr. A 719 (1996) 45.
[6] T.F. Woloszyn, P.C. Jurs, Anal. Chem. 64 (1992) 3059.
[7] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley-Interscience, New York, 1987.
[8] R. Kaliszan, A. Kaliszan, T.A.G. Noctor, J. Chromatogr. 609 (1992) 69.
[9] S. Yang, J.G. Bumgarner, L.F.R. Kruk, M.G. Khaledi, J. Chromatogr. A 721 (1996) 323.
[10] P. Lukkari, H. Vuorela, M.L. Riekkola, J. Chromatogr. A 652 (1993) 451.
[11] M. Salo, H. Siren, P. Volin, S. Wiedmer, H. Vuorela, J. Chromatogr. A 728 (1996) 83.
[12] S. Fu, C.A. Lucy, Anal. Chem. 70 (1998) 173.
[13] S. Fu, C.A. Lucy, Analyst 123 (1998) 1487.
[14] H. Liang, H. Vuorela, M. Riekkola, R.J. Hiltunen, J. Chromatogr. A 798 (1988) 233.
[15] S.L. Anker, P.C. Jurs, Anal. Chem. 64 (1992) 1157.
[16] W.L. Xing, X.W. He, Anal. Chim. Acta 349 (1997) 283.
[17] M. Jalali-Heravi, M.H. Fatemi, Anal. Chim. Acta 415 (2000) 95.
[18] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 915 (2001) 177.
[19] K.L. Peterson, Anal. Chem. 64 (1992) 379.
[20] M. Jalali-Heravi, F. Parastaar, J. Chem. Inf. Comput. Sci. 40 (2000) 147.
[21] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 825 (1998) 283.
[22] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 897 (2000) 227.
[23] M. Jalali-Heravi, Z. Garakani-Nejad, J. Chromatogr. A 927 (2001) 211.
[24] 2001. http://www.disat.unimib.it/chm/

[25] J.J.P. Stewart, MOPAC, Semi Empirical Molecular Orbital Program, QCPE, Vol. 455 (1983), Research Laboratory, United States Air Force Academy, version 6 (1990).
[26] J.H. Schuur, P. Selzer, J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996) 36.
[27] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley–VCH, Weinheim, 2000.
[28] J.H. Schuur, J. Gasteiger, Anal. Chem. 69 (1997) 2398.
[29] V. Consonni, R. Todeschini, M. Pavan, J. Chem. Inf. Comput. Sci. 42 (2002) 682.
[30] V. Consonni, R. Todeschini, M. Pavan, J. Chem. Inf. Comput. Sci. 42 (2002) 693.
[31] J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, Wiley–VCH, Weinheim, 1999.
[32] S. Haykin, Neural Network, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[33] N.K. Bose, P. Liang, Neural Network, Fundamentals, McGraw-Hill, New York, 1996.
[34] M.H. Fatemi, J. Chromatogr. A 955 (2002) 273.
[35] T.B. Blank, S.T. Brown, Anal. Chem. 65 (1993) 3084.
[36] M.H. Abraham, C. Treiner, M. Roses, C. Rafols, Y.Y. Ishihama, J. Chromatogr. A 752 (1996) 243.