Asian Journal of Chemistry Vol. 17, No. 2 (2005), 681–688

Prediction of Microemulsion Electrokinetic Chromatographic Migration Indices from Molecular Structure

MOHAMMAD R. HADJMOHAMMADI and FARIBA SAFA*
Institute of Chemistry, Mazandaran University, P.O. Box 453, Babolsar, Iran
E-mail: safa_f@yahoo.com; Tel: (98)11252-42025; Fax: (98)11252-42002

A quantitative structure-migration relationship study was performed to develop a predictive model correlating 33 molecular structures with their migration indices (MI) in microemulsion electrokinetic chromatography. A high quality model (R = 0.992, standard error of estimate = 0.199) was generated using only calculated descriptors and multiple linear regression techniques. The cross-validation procedure results showed good predictive ability of the model (R^2_{cv} = 0.905).The results of the study indicated that a five-parameter equation can be utilized for prediction of migration indices of compounds for which there are no empirical MI values.

Key Words: Migration indices, Microemulsion electrokinetic chromatography, Quantitative migration-activity relationship, Mathematical modelling.

INTRODUCTION

Retention indices in gas and liquid chromatography have found widespread application in identification, characterization of separation systems and in investigation of solute-retention relationships[1–5]. In recent years, Muijselaar *et al.*[6] and Ahuja and Foley[7] described the possibilities of a migration index scale in micellar electrokinetic chromatography (MEKC). The migration index scale was applied to microemulsion electrokinetic chromatography (MEKKC) by Ishihama *et al.*[8] Migration index scales can be applied for different purposes. They provide a reproducible parameter for identification[6, 8], characterization and classification of pseudo-stationary phase[9] and determination of solute lipophilicity[8, 10].

Similar to MEKC, the capacity factor k′ of a neutral solute in MEEKC can be calculated as follows:

$$K' = \frac{t_s - t_{eo}}{t_{eo}\left(1 - \frac{t_s}{t_m}\right)} \quad (1)$$

where t_{eo}, t_m and t_s are the migration times of the electroosmotic flow, the microemulsion and the solute, respectively. The MI value of a solute is defined as

$$MI = c \log k' + d \quad (2)$$

where c and d are the slope and the intercept of a calibration line between log k′ values of the solutes in a reference scale and their carbon number, respectively.

Preliminary quantitative migration-activity relationship (QMAR) studies suggest the usefulness of the migration index values as biological hydrophobic parameters. The MI values obtained by MEEKC could be used as hydrophobic parameters instead of log P_{ow}. This parameter can be used not only to predict bioactivity of drugs but also to predict selectivity in MEEKC.

The aim of the present research was to develop a model capable of predicting the microemulsion electrokinetic chromatographic migration indices of 33 aromatic compounds possessing different functionalities. Calculations generate numerical descriptors that encode structural information about compounds in the data set. Then multiple linear regression statistical analysis relates the descriptors to migration index. The validity of the model for predicting migration indices is also studied.

EXPERIMENTAL

This quantitative structure-migration study was performed in four phases: (1) selection of data set, (2) generation of molecular descriptors, (3) multiple linear regression analysis, and (4) model validation. The SPSS/PC software[11] was used for regression analysis. The HYPERCHEM[12] and MOPAC (version 6.0)[13] softwares were employed for optimization of the molecules and quantum-chemical descriptor calculations.

Data set

The experimental values of the migration indices for aromatic compounds in the data set was taken from the values reported by Ishihama *et al.*[8] They obtained migration indices for 53 aromatic compounds. Some compounds were not modelled in this study as their structures prohibited accurate molecular modelling. Migration indices were obtained under the same conditions using a Beckman (Palo Alto, CA) P/ACE system 2100 equipped with an uncoated fused silica capillary (GL Sciences, Tokyo, Japan) of 50 μm I.D. and 27 cm length (20 cm to the detector). Separation solution was containing 1.44% SDS, 6.49% 1-butanol, 0.82% heptane (wt. %) in 0.1 M borate-0.05 M phosphate buffer (pH = 7.0). Applied voltage and UV detector wavelength were set at 7.5 kV and 214 nm. respectively. Migration index values were calculated using *n*-alkyl benzenes as reference scale at temperature 25°C. Table-1 shows the compounds studied in this work.

Descriptor generation

The structure of a compound can be represented by a set of calculated numerical descriptors. A total of 39 separate descriptors were calculated for each compound in the data set. These descriptors were either topological or electronic in nature.

Topological descriptors include fragment and molecular connectivity indices. Fragment descriptors account for the number of substructures in the molecule. They included molecular weight and the number of atoms and bonds in the compound. Some of the molecular connectivity descriptors calculated were Wiener index[14], Balaban index[15] and simple path connectivity indices ($^0\chi$, $^1\chi$ and $^2\chi$) developed by Kier and Hall[16].

The electronic descriptors consisted of heat of formation, ionization potential, partial charge of the most negative and most positive atom in the molecule, etc. These descriptors were calculated by semiempirical quantum chemical modelling of the molecules using AM1 methodology[17] in the framework of the MOPAC program.

Statistical analysis and model validation techniques

The linear relationship between experimental MI values and molecular descriptors was modelled using stepwise multiple regression technique[18]. In order to reduce the number of descriptors prior to modelling, the concept of nonredundant descriptors[19] was applied. Based on this concept, when two descriptors are highly correlated ($R > 0.9$), the one which presents the best correlation with the dependent variable is used for further analysis, leaving out the descriptor showing lower correlation. Criteria for selection of the best model were the multiple correlation coefficient (R), the standard error (SE), F ratio value and the number of descriptors in the model.

The models obtained were tested for their statistical validity and robustness by using cross-validation procedure[20]. This methodology was conducted leaving out one compound from the data set and regenerating the model coefficients which then generate a predicted value for the missing compound. This is repeated for the entire data set and cross-validated R^2 (R^2_{cv}) is calculated. Higher R^2_{cv} value indicates the stability of the model and their prediction ability for compounds of nature similar to those in the data set.

TABLE-1
DATA SET

No.	Compound	No.	Compound
1.	Pyrazine	18.	Anisole
2.	4-Methylpyrimidine	19.	Methyl benzoate
3.	Methylpyrazine	20.	Benzene
4.	4.6-Dimethylpyrimidine	21.	Propiophenone
5.	Ethyl pyrazine	22.	*p*-Chlorophenol
6.	Resorcinol	23.	*p*-Ethylphenol
7.	N-methylbenzamide	24.	Butyrophenone
8.	Benzyl alcohol	25.	Toluene
9.	Acetanilide	26.	2-Naphthol
10.	Quinoxaline	27.	Chlorobenzene
11.	*p*-Methoxyphenol	28.	*p*-Propylphenol
12.	Phenol	29.	Ethylbenzene
13.	Benzaldehyde	30.	Naphthalene
14.	Acetophenone	31.	Propylbenzene
15.	*p*-Cresol	32.	Butylbenzene
16.	*o*-Cresol	33.	Anthracene
17.	*m*-Cresol		—

RESULTS AND DISCUSSION

Molecular descriptors and regression coefficients of the best model for the prediction of the $(\log MI)^{-1}$ values of data set compounds are given in Table-2. Mean effects of the model parameters are also given in the table. As shown, the best model was obtained when the inverse form of the dependent variable was used. The high value of R and the F statistics indicates that the model is quite successful in calculating the microemulsion electrokinetic chromatographic migration indices. The standard error is 0.199 and almost 98% of the variance is accounted for by the equation. Calculated descriptors employed in the prediction model are listed in Table-3 for all molecules used in this study. Fig. 1 shows a plot of predicted *vs.* experimental 1/log MI values. The line with $R > 0.99$ indicates theoretically perfect fit. Residual plot (Fig. 2) illustrates that the residuals randomly distributed about a

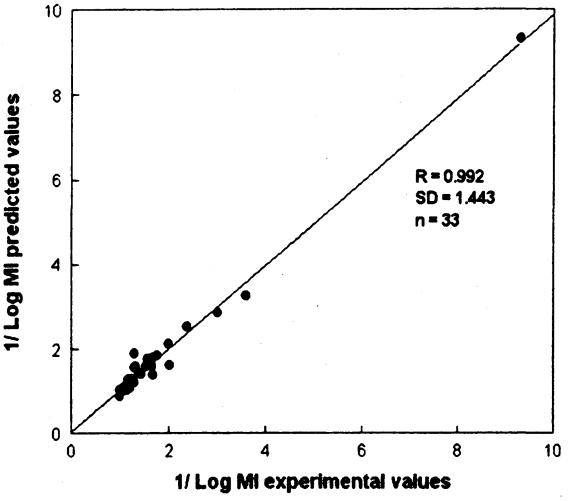

Fig. 1. Plot of predicted *vs.* experimental microemulsion electrokinetic chromatographic $(\log MI)^{-1}$ for 33 aromatic compounds using the prediction model.

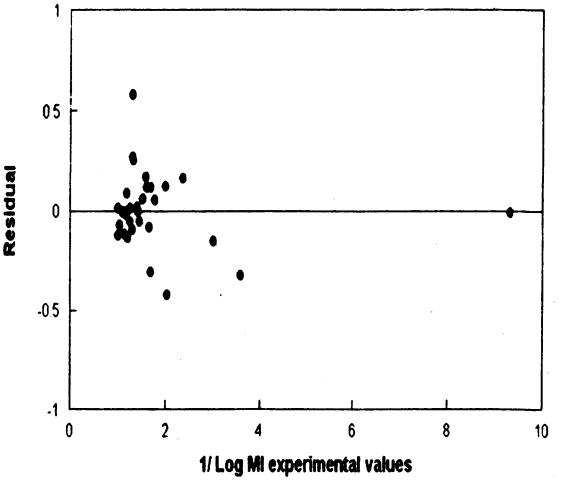

Fig.2. Plot of residual *vs.* experimental values of $(\log MI)^{-1}$

mean of zero and the error variance is nearly constant. Thus it can be concluded that no correlation exists between the chemical structure of the compounds studied and the residuals.

TABLE-2
SPECIFICATION OF THE SELECTED MULTIPLE LINEAR REGRESSION MODEL FOR THE PREDICTION OF $(\log MI)^{-1}$ VALUES

Descriptor[a]	Coefficient[b]	Mean effect
CXREL	13.896 (±1.418)	1.245
1/CCREP	2589.394 (±161.569)	0.733
CEFF	0.338 (±0.055)	2.037
1/QPOS	–0.108 (±0.030)	–0.558
MW	–0.025 (±0.005)	–2.910
Constant	1.223 (±0.386)	
Statistics:		
R^{2}[c]	0.985	
SE	0.199	
F	343.14	
R^{2}_{cv}[d]	0.905	

[a] Definitions of the descriptors are given in the text.
[b] The standard error (SE) of coefficients is given in parentheses.
[c] Square of the correlation coefficient.
[d] Square of the cross-validated correlation coefficient

As previously mentioned, cross-validation experiment was performed to validate the model. The results showed that the five-parameter correlation (Table-2) is quite valid and stable as judged from its R^{2}_{cv} (0.905).

Examination of the descriptors included in the model shows that they encode different aspects of the molecular structures. These parameters mainly show topological and electronic characteristics of the molecules indicating that these properties affect the MI values. The most important descriptor in the model is molecular weight (MW) with the largest mean effect. The presence of this topological descriptor in the model clearly indicates the importance of dispersion interactions related to molecular size in the retention mechanism in MEEKC. The correlation of MI with the molecular weight is obvious, as larger molecules have longer migration time. The number of carbon atoms in the molecule which are connected only to other carbon and/or hydrogen atoms (CEFF) is also a topological descriptor. As can be seen, mean effects for the MW and CEFF descriptors are –2.910 and 2.036, respectively. The net effect of these descriptors is increasing of MI value as the number of carbon atoms and thus the molecular mass increases. The third parameter in the model is the relative number of the C—X bonds (CXREL) where X is O, N or Cl atom that possibly accounts for

TABLE-3
EXPERIMENTAL AND CALCULATED VALUES OF $(\log MI)^{-1}$ AND DESCRIPTORS EMPLOYED IN THE SELECTED MODEL

No.[b]	Descriptors[a]					$(\log MI)^{-1}$	
	CXREL	$CCREP^{-1}$	CEFF	$QPOS^{-1}$	MW	Predicted	Experimental
1.	0.3077	0.0022	0	0.0769	80.089	9.32	9.33
2.	0.2500	0.0003	2	5.2274	94.116	3.26	3.59
3.	0.2500	0.0003	1	5.8582	94.116	2.85	3.01
4.	0.2105	0.0002	3	5.2576	108.143	2.52	2.36
5.	0.2105	0.0002	2	5.8617	108.143	2.12	2.00
6.	0.1177	0.0002	4	4.5249	110.112	1.60	2.03
7.	0.1304	0.0002	6	2.9403	135.165	1.83	1.77
8.	0.0526	0.0002	6	4.9628	108.140	1.38	1.69
9.	0.1304	0.0002	6	3.2310	135.165	1.80	1.68
10.	0.1818	0.0002	4	5.8651	130.149	1.72	1.60
11.	0.1500	0.0002	4	4.6296	124.139	1.57	1.66
12.	0.0625	0.0003	5	4.6019	94.113	1.75	1.58
13.	0.0556	0.0003	6	4.4783	106.124	1.58	1.52
14.	0.0476	0.0002	7	3.7467	120.151	1.40	1.45
15.	0.0526	0.0002	6	4.6126	108.140	1.42	1.40
16.	0.0526	0.0002	6	4.5809	108.140	1.41	1.41
17.	0.0526	0.0002	6	4.6104	108.140	1.42	1.41
18.	0.1053	0.0002	5	6.7522	108.140	1.57	1.32
19.	0.1364	0.0002	6	2.8977	136.150	1.89	1.31
20.	0.0000	0.0004	6	7.6864	78.113	1.56	1.29
21.	0.0417	0.0002	8	3.7594	134.177	1.21	1.31
22.	0.1250	0.0002	4	4.5434	128.558	1.27	1.26
23.	0.0455	0.0002	7	4.6168	122.166	1.19	1.24
24.	0.0370	0.0001	9	3.7453	148.204	1.07	1.21
25.	0.0000	0.0003	7	7.6864	92.104	1.27	1.18
26.	0.0435	0.0002	9	4.6019	144.173	1.20	1.20
27.	0.0667	0.0003	5	6.8259	112.559	1.14	1.16
28.	0.0400	0.0002	8	4.6147	136.193	1.02	1.14
29.	0.0000	0.0002	8	7.5358	106.167	1.10	1.10
30.	0.0000	0.0002	10	7.5758	128.173	1.09	1.10
31.	0.0000	0.0002	9	7.5529	120.194	0.97	1.04
32.	0.0000	0.0002	10	7.5358	134.221	0.88	1.00
33..	0.0000	0.0001	14	7.4571	178.233	1.02	1.00

[a]Definitions of the descriptors are given in the text.
[b]The numbers refer to the numbers of molecules given in Table-1.

TABLE-4
CORRELATION MATRIX FOR THE DESCRIPTORS USED IN THE QSMR MODEL

Descriptor	CXREL	$CCREP^{-1}$	CEFF	$QPOS^{-1}$	MW
CXREL	1.000				
$CCREP^{-1}$	0.479	1.000			
CEFF	–0.856	–0.464	1.000		
$QPOS^{-1}$	–0.495	–0.468	0.400	1.000	
MW	–0.316	–0.469	0.684	–0.008	1.000

the electronic effects of the electronegative atoms (X) on the intermolecular interactions. A relatively large mean effect of this descriptor reveals that intermolecular interactions play some roles in the mechanism of the retention. It should be noted that three topological descriptors used in the model are fragment type and can be generated simply through examination of a structural diagram. The other descriptors including the inverse forms of core-core repulsion energy (CCREP) and charge on most positive atom (QPOS) are electronic in nature. They encode electronic properties of the molecules which presumably relate to the polar interactions between the solute and the MEEKC medium. It can be seen from the equation that MI value is inversely proportional to the charge on the most positive atom in the molecule that encodes the polarity of the molecule. This is in agreement with the experiment that interaction with microemulsion decreases with an increase in solute polarity. On the other hand, CCREP has an opposite effect on MI value. Mean effect of this parameter in the model is positive indicating that a compound with a higher core-core repulsion energy has a higher MI value. The correlation matrix of the descriptors in the proposed model (Table-4) shows the variables to be uncorrelated. Thus it can be concluded that no correlation exists between the chemical structure of the compounds studied and the residuals.

Conclusions

The results of this study demonstrate that the quantitative structure-migration relationship can generate high quality model using only calculated descriptors. The model developed to represent migration behaviour was statistically valid, stable and fits the data well. The good predictive ability of the model should allow estimation of migration indices for similar compounds in cases where migration values are not readily available.

REFERENCES

1. V. Pacakova and L. Feltl, *Chromaiographic Retention Indices,* Ellis Horwood, New York (1992).
2. L. Rohrschneider, *J. Chromatogr. A.*, **77**, 1 (1965).
3. W.O. McReynolds, *J. Chromatogr. Sci.*, **8**, 685 (1970).
4. R.M. Smith, *J.Chromatogr. Lib.*, **57**, 11 (1995).
5. ———, *Adv. Chromatogr.*, **26**, 277 (1987).

6. P.G. Muijselaar. H.A. Claessens and C.A. Cramers, *Anal. Chem.*, **66**, 635 (1994).
7. E.S. Ahuja and J.P. Foley, *Analyst*, **119**, 353 (1994).
8. Y. Ishihama, Y. Oda, K. Uchikawa and N. Asakawa, *Anal. Chem.*, **67**, 1588 (1995).
9. P.G. Muijselaar, H.A. Claessens and C.A. Cramers. *Anal. Chem.*, **69**, 1184 (1997).
10. S. Yang, J.G. Bumgarner, L.F.R. Kruk and M.G. Khaledi, *J. Chromatogr. A.*, **721**, 323 (1996).
11. SPSS/PC: The Statistical Package for IBMPC, Quiad Software, Ontario (1986).
12. Hyperchem, Release 3 for Windows, Molecular Modeling System, HyperCube Inc. and Autodesk Inc., developed by HyperCube Inc. (1993).
13. MOPAC package, version 6.0, U.S. Air Force Academy, Colorado Spring, Co. 80840.
14. H. Wiener, *J. Am. Chem. Soc.*, **69**, 17 (1947).
15. A.T. Balaban, *Chem. Phys. Lett.*, **89**, 399 (1982).
16. L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York (1976).
17. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
18. D. George and P. Mallery, SPSS/PC+ Step by Step: A Simple Guide and Reference, Wadsworth Publishing Company, Belmont, California (1995).
19. A. Katrizky and E. Gordeeva, *J. Chem. Inf. Comput. Sci.*, **33**, 835 (1993).
20. R.D. Cramer, D.E. Patterson and J.D. Bunce, *J. Am. Chem. Soc.*, **110**, 5959 (1988).

(*Received*: 27 January 2004; *Accepted*: 11 October 2004) AJC-3993