Asian Journal of Chemistry Vol. 22, No. 6 (2010), 4727-4734

Prediction of Electrophoretic Mobility of Analytes in Capillary Electrophoresis Using Molecular Connectivity Indices

SOMAIEH SOLTANI†, SIAVOUSH DASTMALCHI‡ and ABOLGHASEM JOUYBAN*
Department of Pharmaceutical and Food Control, Faculty of Pharmacy,
Tabriz University of Medical Sciences, Tabriz 51664, Iran
Fax: (98)(411)3344798; Tel: (98)(411)3363234
E-mail: ajouyban@hotmail.com

Molecular connectivity indices are the most widely used topological indices in QSAR and QSPR studies. In this paper, the ability of these indices for prediction of electrophoretic mobility of a diverse set of analytes is investigated. The proposed single parameter models were able to predict the mobilities well. Main advantages of the proposed method are: (1) its simplicity, (2) applicability to a wide range of acidic, basic and neutral analytes and (3) applicability to represent the migration behaviours for different electrophoresis methods such as MEKC and non-aqueous capillary electrophoresis.

Key Words: Electrophoretic mobility, Capillary electrophoresis, Quantitative structure property relationship, Modeling, Connectivity indices.

INTRODUCTION

Capillary electrophoresis (CE) is a separation technique which provides many advantageous including high resolution, good efficiency, availability of several separation modes, rapid analyses and small consumption of both sample and solvent in comparison with other analytical methods like HPLC. The electrophoretic mobility is the most important parameter, governing the separation of analytes in capillary electrophoresis and any attempt to predict the mobility of an analyte could provide useful information for the analyst to develop a new separation method using capillary electrophoresis. A number of papers has been published dealing with mobility modeling in capillary electrophoresis using different independent variables[1-8]. The aim of this communication is to present a simple quantitative relationship between the electrophoretic mobility of analytes and connectivity indices computed by Dragon software, as independent variables. The accuracy of the proposed model is evaluated using collected data sets by computing mean percentage deviation (MPD) of the calculated mobilities from experimental values.

†Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz 51664, Iran.
‡Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz 51664, Iran.

The most widely used topological indices are molecular connectivity indices introduced by Randic and developed extensively by Kier and Hall[9]. The molecular connectivity indices can encode the structural features such as size, branching, un saturation, heteroatom content and molecular cyclicity. These indices have been used successfully to correlate the structures of molecules with their physical, chemical or biological properties. Kier and Hall in a series of papers[10-12] showed that there is a significant linear correlation between a Randic connectivity index and physicochemical properties such as solvent cavity surface area, molecular polarizability, water solubility, boiling point and partition coefficient of a number of hydrocarbons. After two decades they published another paper and analyzed the molecular connectivity indices for the information in bond terms and concluded that the connectivity indices are indeed non-empirical structure descriptors which are rich in information. The authors discussed these indices as bimolecular encounter accessibility in a milieu[9]. These findings beside the applications of these simple indices for about 30 years in different physicochemical and biological areas[13-16] encouraged us to use them in mobility prediction of organic molecules, which is the best example of interaction of a molecule in a milieu.

A review of papers showed that there is just one publication which used these parameters in mobility prediction. Liang and co-workers[17] developed 23 different three constant MLR models using connectivity indices in order to represent the mobility of 13 flavonoids with correlation coefficients ranging from 0.93-0.99 and relative standard error of about 10 %. In addition to a relatively high error value, there is a dilemma of choosing a model among 23 presented models for practical uses.

Other QSPR models are developed until now are relayed on other calculated structural parameters. A summary of these models are included in Table-1. A linear model representing the electrophoretic mobility of different data sets using a single algorithm based on structural features of the analytes was presented as[18]:

$$\ln \mu = K_0 + K_1PQ + K_2V^{2/3} + K_3TE + K_4\Delta H_f + K_5MR \quad (1)$$

where μ is the effective electrophoretic mobility, PQ is partial charge, $V^{2/3}$ denotes surface area, TE stands for total energy, ΔH_f represents heat of formation, MR is molecular refractivity and K_0-K_5 are the model constants which are calculated using a least squares analyses. A set of 115 carboxylic acids' mobilities were predicted using different linear MLR and various non-linear (ANN, RBFNN, CART-ANFIS) methods and the researchers concluded that non-linear methods are better than linear ones[7]. The mechanistic methods are also reported[19]. The details of the summarized methods and utilized descriptors could be found in literature. A brief comparison between the accuracy of the proposed model with previous models is presented in this work.

TABLE-1
SUMMARY OF PUBLISHED ELECTROPHORETIC MOBILITY PREDICTION METHODS

Data set	No. of data points (N^a)	Modeling method	MPD	Reference
Sulfonamides (cationic, anionic)	13	MLR ANN	9.3-1.9 1.3-0.5	1
Flavonoides	13	MLR	10.0	17
Pyridine derivatives	31	ANN MLR	0.8 1.4	20
Ammonium derivatives	56	ANN	3.7	21
β-Blockers	10		5.8	
Benzoates	26		7.0	
NSAIDs	11	MLR	1.5	18
Sulfonamides	13		2.7	
Amines	18		7.2	
Carboxylic acid	115	HM RBFNN	11.7 5.0	4
Carboxylic acid Aromatic sulfonic acid	115 21	Mechanistic equation	7.5 4.0	19
Carboxylic acid	115	CART-ANFIS Variable Ranking	4.8 28.4	7
Monoamines	34	Non linear MLR	4.1	22
Flavonoids	13	MLR RBFNN	7.2 2.2	3

EXPERIMENTAL

Electrophoretic mobilities of different sets of analytes collected from the literature (Table-2), were used to check the applicability of the proposed model. The molecular connectivity indices used in this work were computed using Dragon 5.4 software. The molecular structures of the analytes were drawn using HyperChem software and the structural data files were transformed into the Dragon 5.4 software. The software computes 30 different connectivity indices. Stepwise regression analyses were used as a parameter selection method and the selected parameters were used to build single parameter models.

TABLE-2
DETAILS OF COLLECTED DATA SETS FROM LITERATURE

Data set	No. of data points (N^a)	Reference
β-Blockers	10	23, 24
NSAIDs	11	25
Sulfonamides	13	26
Amines	18	27
Aromatic sulfonic acid	21	19
Benzoates	26	28
Pyridine derivatives	31	20
Carboxylic acid	115	29

The calculated mobilities were compared to the experimental values and mean percentage deviation (MPD) was computed as an accuracy criterion by:

$$MPD = \frac{100}{N}\sum_{1}^{N}\left|\frac{\mu_{calculated} - \mu_{observed}}{\mu_{observed}}\right| \quad (2)$$

where N is the number of experimental data points. All the calculations were performed using SPSS software.

In order to compare the proposed method with previously approved empirical well-known Offord method, the $Q/M^{2/3}$ parameter was calculated for all data sets. The Q (charge) was calculated for each molecule according to the buffer pH and the pK_a of each molecule. The molecules were considered full ionized if the pK_a at least 2 units differs from the pH of the background electrolyte.

RESULTS AND DISCUSSION

The stepwise regression methods was performed for each set (which were divided to 2/3 training and 1/3 prediction data points) in order to select the best predictors. The descriptors with the highest correlation with the experimental mobility data and the lowest inter-correlation with each other were selected for each set. The selected variables were compared with each other and the most frequently selected ones were chosen. The selected variables (x0v, x1v, x2sol) were correlated with the dependent variable (ln μ which we found that results to better predictions) and also with each other. As the highest Pearson correlation coefficient was observed between electrophoretic mobility and predictors were different for each data set, we developed three single parameter models and compared the models properties with each other. To test the applicability of the selected descriptors, the constants of the considered models, *i.e.*, eqn. 3 were computed and their statistical significance were evaluated using *t*-test.

Average of correlation coefficients, along with F values and MPDs were considered as the final variable selection criteria. The general form of the best MLR model is:

$$\ln \mu = J_0 + J_1X_1 \quad (3)$$

where J_0 and J_1 are the model constants and X_1 denotes one of the selected descriptors.

Whole training data points of each set were fitted to eqn. 3 and the back-calculated mobilities were used to compute the MPD values which were listed in Table-3. These analyses were called correlative analyses and showed the fitness ability of a model. The overall MPDs (OMPDs) for these analyses were 4.7, 4.7 and 7.1 % for x0v, x1v and x2sol as dependent variable, respectively. Similar results were reported in an earlier paper for a six constant MLR model (eqn. 1) using chemical descriptors computed by HyperChem for these 5 data sets[18]. The OMPD for correlative analyses of eqn. 1 was 1.4 % and the results of paired *t*-test showed that there is no significant difference between OMPDs of eqns. 1 and 3 ($p > 0.05$). It should be noted that eqn. 1

TABLE-3
SINGLE PARAMETER CONNECTIVITY INDICES
MODEL DETAILS FOR EACH DATA SET

Data set	Predictor	R	MPD (SD)		F
			Training	Prediction	
Sulfonamides	x1v	0.6	3.0 (3.1)	4.7 (3.1)	76.2
	x2sol	0.8	2.6 (1.6)	4.0 (3.6)	70.9
	x0v	0.9	2.2 (1.8)	1.9 (1.4)	23.6
NSAIDs	x1v	0.9	2.3 (2.0)	1.6 (0.7)	41.4
	x2sol	0.9	2.2 (1.6)	3.8 (2.6)	27.1
	x0v	1.0	1.1 (0.5)	1.6 (0.6)	259.2
β-Blockers	x1v	1.0	5.0 (2.7)	6.7 (6.2)	28.7
	x2sol	0.9	8.2 (6.4)	9.0 (1.6)	23.4
	x0v	0.9	5.6 (3.2)	5.9 (5.6)	41.3
Benzoates	x1v	0.6	3.7 (2.6)	6.0 (3.6)	2.6
	x2sol	0.4	4.1 (2.6)	6.3 (3.9)	2.1
	x0v	0.4	3.7 (2.5)	5.9 (3.7)	3.0
Amines	x1v	0.9	8.5 (7.5)	12.3 (6.4)	42.3
	x2sol	0.8	11.8 (7.3)	14.3 (6.4)	26.4
	x0v	0.9	9.8 (6.4)	11.3 (7.0)	32.1
	OMPD for x1v	4.7	5.9		
	OMPD for x2sol	7.1	7.1		
	OMPD for x0v	4.7	5.1		

possesses six constant terms whereas eqn. 3 employs only two constant terms and as a general rule in least square models, the more constant terms in a MLR model, the more accurate the correlation. However, in this case, no significant difference between MPDs of the models meaning that the descriptors of eqn. 3 are able to better correlate the dependent variable in comparison with those of eqn. 1.

In order to check the predictive ability of the developed models, the mobilities of prediction data points of each set were calculated using developed models and reported in Table-3. The OMPDs for predictive analyses were 5.1, 5.9 and 7.1 % for x0v, x1v and x2sol contained models, respectively. The corresponding value for eqn. 1 was 4.8 %. As correlative analyses the difference is not significant. It should be noted that all parameter selection and model development steps for eqn. 3 was performed in the absence of prediction data points whereas the parameter selection step for eqn. 1 was done in the presence of all data points.

Using OMPD as selection criteria between the developed models, both x0v and x1v containing models were selected as suitable methods. In order to check the capability of the developed models for using in electrophoretic method development of external data sets, we obtained 3 different data sets from the literature and tested present developed method without any parameter selection step. It is found that the selected parameters (*i.e.*, x1v or x0v) have significant correlation with these mobility data. The correlative and predictive analyses were carried out for

these data sets and the results are summarized in Table-4. These results reveal, the model is able to predict the mobility with the OMPD of 9.1, 10.5 and 10.4 % for x1v, x2sol and x0v as dependent variable, respectively. In order to evaluate the models, leave group out method was used for each data set. About 30 % of data were removed randomly from each data set and the remained were regressed to x0v as training set. The MPDs were calculated for the predicted set (the removed data) and summarized in Table-5. The largest deviation from the proposed methods appeared through aromatic sulfonic acids data set which can be resulted from the experimental data inaccuracy. The mean RSD of the reported mobilities from different laboratories is about 8 %[19]. By considering 8 % for repeadibility of the mobility data, the preduced MPDs could be considered acceptable.

TABLE-4
ONE PARAMETER CONNECTIVITY INDICES
MODEL DETAILS FOR EXTERNAL DATA SETS

Data set	Predictor	R	MPD (SD)		F
		Training	Training	Prediction	
Carboxylic acids	x1v	0.4	18.6 (11.2)	18.1 (12.3)	16.4
	x2sol	0.3	19.6 (12.7)	19.2 (14.1)	4.9
	x0v	0.4	17.3 (11.9)	20.0 (11.1)	14.4
Sulfonic acids	x1v	0.8	13.3 (14.0)	15.2 (10.9)	20.0
	x2sol	0.8	12.7 (13.2)	16.8 (10.5)	23.3
	x0v	0.7	13.9 (15.8)	15.3 (11.4)	15.7
Pyridine derivatives	x1v	1.0	3.2 (2.1)	3.4 (2.7)	352.4
	x2sol	0.9	6.3 (5.5)	6.2 (5.1)	64.5
	x0v	1.0	3.9 (3.3)	3.5 (3.2)	185.2
	OMPD for x1v	11.7	9.1		
	OMPD for x2sol	12.9	10.5		
	OMPD for x0v	11.7	10.4		

Conclusion

The proposed model showed reasonably accurate calculations for the mobility of analytes in the studied electrophoretic conditions. The advantages of the proposed models are: (i) There is no need to special software to calculate the descriptor and one can draw a molecule on paper and calculate the descriptors. (ii) It's a general model and there is no need to descriptor selection for each new data set. (iii) There is no need for complicated nonlinear numerical methods. (iv) It is an ideal method for structural isomers, homologue series and small molecules. (v) It is applicable for both CZE and MEKC methods.

But there are some limitations: (i) The proposed models are not able to distinguish between E and Z isomers. (ii) It is not so useful for completely different molecular structures. (iii) It is not so useful for larger molecules (such as peptides), possibly because of many internal hydrogen binding and other interactions.

TABLE-5
LEAVE GROUP OUT RESULTS FOR x0v CONTAINING MODEL

Data set	Group	R	MPD (SD)	Reported MPD
Amines	1	0.8	13.1 (4.6)	12.3
	2	0.9	8.2 (9.0)	
	3	0.9	12.8 (10.1)	
Benzoates	1	0.7	5.2 (4.7)	6.0
	2	0.4	4.5 (3.3)	
	3	0.5	2.8 (15.0)	
β-Blockers	1	0.9	5.6 (3.1)	6.7
	2	0.9	7.4 (10.0)	
	3	1.0	6.7 (6.2)	
NSAIDs	1	0.9	1.5 (1.0)	1.6
	2	1.0	3.7 (3.0)	
	3	1.0	1.6 (2.5)	
Sulfonamides	1	0.9	4.0 (4.8)	4.7
	2	0.7	4.4 (1.8)	
	3	0.7	2.2 (1.7)	
Pyridines	1	1.0	2.8 (1.6)	3.4
	2	1.0	3.8 (3.2)	
	3	1.0	2.9 (1.8)	
Aromatiic sulfonic acids	1	0.8	19.3 (24.2)	15.2
	2	0.8	12.7 (12.5)	
	3	0.7	9.8 (6.3)	
Carboxylic acids	1	0.4	19.6 (10.3)	18.1
	2	0.5	19.9 (11.9)	
	3	0.4	18.1 (11.3)	

ACKNOWLEDGEMENT

The authors would like to thank the Drug Applied Research Center, Tabriz University of Medical Sciences for financial support of this work (Grant No. 83-42).

REFERENCES

1. M. Jalali-Heravi and Z. Garkani-Nejad, *J. Chromatogr. A*, **927**, 211 (2001).
2. D. Barron, E. Jimenez-Lozano and J. Barbosa, *J. Chromatogr. A*, **919**, 395 (2001).
3. S. Wang, C. Xue, X. Chen, M. Liu and Z. Hu, *J. Chromatogr. A*, **1033**, 153 (2004).
4. C. Xue, H. Liu, X. Yao, M. Liu, Z. Hu and B. Fan, *J. Chromatogr. A*, **1048**, 233 (2004).
5. T. Yokoyama, H. Tsuji, T. Akamatsu and M. Zenki, *Anal. Sci.*, **20**, 657 (2004).
6. M.H. Fatemi and N. Goudarzi, *Electrophoresis*, **26**, 2968 (2005).
7. M. Jalali-Heravi and P. Shahbazikhah, *Electrophoresis*, **29**, 363 (2008).
8. A. Jouyban, H. Ahmadi and S. Soltani, *Asian J. Chem*., **20**, 1148 (2008).
9. L.B. Kier and L.H. Hall, *Croat. Chem. Acta*, **75**, 371 (2002).
10. L.B. Kier, L.H. Hall, W.J. Murray and M. Randic, *J. Pharm. Sci*., **64**, 1971 (1975).
11. L.H. Hall, L.B. Kier and W.J. Murray, *J. Pharm. Sci*., **64**, 1974 (1975).
12. W.J. Murray, L.H. Hall and L.B. Kier, *J. Pharm. Sci*., **64**, 1978 (1975).
13. H. Dureja and A.K. Madan, *Chem. Biol. Drug Des*., **73**, 258 (2009).

14. Q. Chen, Y.W. Kou, Q. Wang, H. Chen and J. Yuan, *J. Environ. Sci. Health A*, **44**, 288 (2009).
15. M. Shamsipur, R. Ghavami, H. Sharghi and B. Hemmateenejad, *J. Mol. Graph Model.*, **27**, 506 (2008).
16. X. Du, Y. Chen and C. Feng, *J. Huazhong Univ. Sci. Tech. Med. Sci., (Natural Science Edition)*, **36**, 124 (2008).
17. H.R. Liang, H. Vuorela, P. Vuorela, M.L. Riekkola and R. Hiltunen, *J. Chromatogr. A*, **798**, 233 (1998).
18. A. Jouyban and B.H. Yousefi, *Comput. Biol. Chem*., **27**, 297 (2003).
19. Y. Cheng and H. Yuan, *Anal. Chim. Acta*, **565**, 112 (2006).
20. M. Jalali-Heravi and Z. Garkani-Nejad, *J. Chromatogr. A*, **971**, 207 (2002).
21. Q. Li, L. Dong, R. Jia, X. Chen, Z. Hu and B.T. Fan, *Comput. Chem*., **26**, 245 (2002).
22. S. Fu and C.A. Lucy, *Anal. Chem*., **70**, 173 (1998).
23. A. Jouyban, H.K. Chan, M. Khoubnasabjafari, N. Jouyban and B.J. Clark, *Daru*, **9**, 1 (2001).
24. A. Jouyban, H.K. Chan, B.J. Clark and E. Kenndler, *J. Microcolumn Sep*., **13**, 346 (2001).
25. M. Fillet, I. Bechet, V. Piette and J. Crommen, *Electrophoresis*, **20**, 1907 (1999).
26. C.E. Lin, W.C. Lin, Y.C. Chen and S.W. Wang, *J. Chromatogr. A*, **792**, 37 (1997).
27. W.H. Matchett and W.C. Brumley, *J. Liq. Chromatogr. Rel. Tech*., **20**, 79 (1997).
28. K. Sarmini and E. Kenndler, *J. Chromatogr. A*, **806**, 325 (1998).
29. M. Wronski, *J. Chromatogr. A*, **657**, 165 (1993).

(*Received*: 19 October 2009; *Accepted*: 20 February 2010) AJC-8468