ELSEVIER

Journal of Chromatography A, 798 (1998) 233–242

JOURNAL OF CHROMATOGRAPHY A

Prediction of migration behaviour of flavonoids in capillary zone electrophoresis by means of topological indices

Hai-Rui Liang[a,*], Heikki Vuorela[a], Pia Vuorela[a], Marja-Liisa Riekkola[b], Raimo Hiltunen[a]

[a]*Pharmacognosy Division, Department of Pharmacy, P.O. Box 56, University of Helsinki, FIN-00014 Helsinki, Finland*

[b]*Laboratory of Analytical Chemistry, Department of Chemistry, P.O. Box 55, University of Helsinki, FIN-00014 Helsinki, Finland*

Abstract

Topological indices of 13 flavonoids were calculated to represent the molecular structures. The electrophoretic mobilities (μ_e) of the flavonoids were obtained to represent the migration in capillary zone electrophoresis (CZE). Then multiple linear regression techniques were used to establish the models for predicting the mobilities by means of the topological indices. 23 models of two-indices were generated with similar statistical results ($R \geq 0.93$, $n=13$, relative standard error$\leq$10%). Each model developed was based on a molecular connectivity index ($^m\chi_t^v$) and the other electrotopological state index (S_i). The $^m\chi_t^v$ and S_i involved in models can reasonably explain the μ_e. Thus, the results demonstrate that $^m\chi_t^v$ and S_i can successfully be used to develop models of structure–mobility of flavonoids in CZE. The structural factors affecting observed and predicted μ_e were discussed.

Keywords: Topological indices; Electrotopological indices; Connectivity indices; Structure–mobility correlation; Flavonoids

1. Introduction

Topological indices are numerical descriptors of molecule structures based on certain topological features of their hydrogen-suppressed graphs [1–3]. The most widely used topological indices are molecular connectivity indices introduced by Randić [4] and developed extensively by Kier and Hall [5]. The molecular connectivity indices can encode the structural features such as size, branching, unsaturation, heteroatom content and cyclicity. Recently Kier and Hall [6] also introduced an atom-level index called the electrotopological state index (S_i), which encodes the combination of electronic, topological and valence state information. These topological indices have been used widely and successfully to correlate the structures of molecules with their physical, chemical or biological properties [7–9].

Quantitative structure–retention relationships (QSRRs) have extensively been studied to explain separation mechanisms, predict retention behaviours and characterise the physicochemical properties of solutes in thin-layer chromatography (TLC) [10], gas chromatography (GC) [11] and high-performance liquid chromatography (HPLC) [7,12]. Capillary electrophoresis (CE) including capillary zone electrophoresis (CZE) and micellar electrokinetic capillary chromatography (MEKC) has potential in structure–property studies because they offer some similar advantages to chromatography for physico-chemical characterisation of molecules [13]. In addition, CE as a complementary technique to HPLC has demonstrated some advantages over HPLC, e.g., high separation efficiency and high power of resolution. Therefore, it is important to carry out QSRR

*Corresponding author.

0021-9673/98/$19.00
PII S0021-9673(97)00958-8

studies in CE. There have been some reports on QSRR studies in CE [13–15] but the QSRR study of flavonoids has not yet been undertaken.

CE has relatively widely been applied to separate flavonoids [16–22]. The effects of molecular structures of some flavonoids on electrophoretic mobility have been investigated [23]. In CZE, the migration behaviour of monohydroxyl substituted flavonoid isomers can be predicted as a function of their pK_a values [24], but the pK_a values of most flavonoids are unknown [18]. In addition, the investigated flavonoids differ in the degree and pattern of hydroxylation and glycosylation. Thus it is difficult to predict the migration of flavonoids. Topological indices of flavonoids can be calculated from two-dimensional molecular structures. If the relationships between the topological indices and migration of flavonoids are reliably and reasonably established by multiple linear regression techniques, the migration behaviour in CZE can be predicted by means of the topological indices.

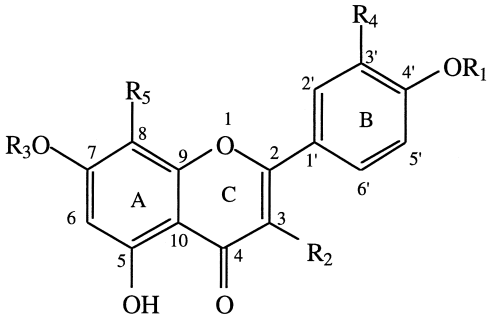
The aim of the present study was to develop correlations between the migration in CZE and topological indices of flavonoids; and try to predict the migration of flavonoids.

2. Experimental

2.1. Chemicals, apparatus and CE procedures

Chemicals and most of the investigated flavonoids are as described earlier [17,18]. In addition, avicularin and vitexin were from Sigma (St. Louis, MO, USA). Structures of the flavonoids are shown in Fig. 1.

The separation was performed on a Hewlett-Pac-



No	Name	R_1	R_2	R_3	R_4	R_5
1	epimedin B	CH_3	O-rha-xyl	glc	H	prenyl
2	epimedin C	CH_3	O-rha-rha	glc	H	prenyl
3	icariin	CH_3	O-rha	glc	H	prenyl
4	icariin I	CH_3	OH	glc	H	prenyl
5	luteolin-3′,7-O-diglucoside	H	H	glc	O-glc	H
6	icariin II	CH_3	O-rha	H	H	prenyl
7	luteolin-7-O-glucoside	H	H	glc	OH	H
8	tricin	H	H	H	OCH_3	H(5′-OCH_3)
9	8-OCH_3-isoquercitrin	H	O-glc	H	OH	OCH_3
10	vitexin	H	H	H	H	C-glc
11	quercitrin	H	O-rha	H	OH	H
12	avicularin	H	O-ara	H	OH	H
13	luteolin	H	H	H	OH	H

rha: rhamnose; xyl: xylose; glc: glucose; ara: arabinose

Fig. 1. Structures of the investigated flavonoids.

kard Chemstation 3D CE equipped with a HP diode array detection (DAD) system and an air cooling system for the capillary cassette (Hewlett-Packard, Avondale, PA, USA). The dimensions of fused-silica capillary (Composite Metal Services, Worcestershire, UK) were 58 cm (50 cm effective length)×0.050 mm I.D.×0.360 mm O.D. The running voltage was kept at 20 kV, detection was at 254 nm and all experiments were carried out at 25°C. The buffer solutions containing 20, 25, 30, 40 or 50 m*M* disodium tetraborate decahydrate (borate) were prepared and adjusted to pH 10.5 each at 25°C. The pH values of the buffers containing 20 m*M* borate were adjusted to 10.5, 10.0, 9.5, 9.0 and 8.5 at 25°C to obtain solutions of different pH. The pH of the buffers was adjusted with a Jenway 3030 pH meter connected to a Jenway electrode (Jenway, Felsted, UK) containing 4 *M* KCl in saturated AgCl. The electrode system was calibrated with potassium hydrogenphthalate (pH 4.00) and borate–sodium hydroxide solutions (pH 11.00). A water system from Gelman Sciences (Ann Arbor, MI, USA) was used for ion-exchange of the distilled water.

The relative standard errors of migration times (t_m) of flavonoids from four replicates under each condition were less 1%. In CZE, the electrophoretic

Table 1
Partial topological indices of 13 flavonoids

No[1]	Topological indices[2]							
	$^{0}\chi$	$^{1}\chi$	$^{2}\chi$	$^{3}\chi$	$^{4}\chi$	$^{5}\chi$	$^{6}\chi$	$^{4}\chi_{p/c}$
1	41.455	27.051	25.298	22.686	17.732	14.514	11.533	12.357
2	42.325	27.462	25.815	23.198	18.093	14.768	11.675	12.912
3	35.024	22.763	21.224	18.940	14.861	12.131	9.560	10.329
4	27.723	18.065	16.633	14.670	11.663	9.529	7.414	7.730
5	31.162	20.425	19.126	17.342	13.840	10.844	8.236	9.426
6	27.016	17.527	16.464	14.254	11.426	9.227	7.370	7.661
7	23.154	15.189	14.366	12.697	10.255	7.979	6.036	6.817
8	17.430	11.439	10.527	9.031	7.960	5.907	4.568	4.523
9	25.602	16.582	15.431	14.119	11.421	9.007	7.058	7.967
10	23.154	15.206	14.301	12.660	10.347	8.284	6.318	6.721
11	23.317	15.078	14.628	12.882	10.330	8.145	6.269	7.346
12	22.447	14.706	13.960	12.354	10.212	7.921	6.065	6.700
13	15.146	9.952	9.628	7.923	6.827	5.166	3.972	4.057
	$^{6}\chi_{ch}$	$^{0}\chi^{v}$	$^{1}\chi^{v}_{p}$	$^{5}\chi^{v}_{p}$	$^{4}\chi^{v}_{p/c}$	$S_{C1'}$	$S_{C6'}$	S_{O1}
1	0.320	31.315	18.158	4.675	4.002	0.239	1.537	6.458
2	0.310	32.185	18.608	4.818	4.260	0.238	1.539	6.472
3	0.265	26.587	15.238	3.828	3.350	0.302	1.569	6.349
4	0.219	20.990	11.865	2.923	2.455	0.371	1.590	6.022
5	0.270	22.073	13.017	3.196	2.787	0.163	1.217	5.822
6	0.219	20.835	11.718	2.868	2.497	0.393	1.619	6.146
7	0.225	16.321	9.497	2.271	1.931	0.448	1.513	5.839
8	0.167	12.861	6.895	1.595	1.256	0.371	1.590	6.022
9	0.204	18.022	10.178	2.495	2.188	0.066	1.078	5.681
10	0.230	16.620	9.825	2.719	2.173	0.065	1.238	5.473
11	0.214	16.537	9.496	2.315	2.068	0.064	1.078	5.681
12	0.169	15.667	9.052	2.112	1.846	0.068	1.076	5.655
13	0.179	10.569	5.977	1.359	1.065	0.060	1.238	5.473

[1] For the explanation of Nos. 1–13, see Fig. 1.

[2] The definitions of topological indices: molecular connectivity indices: $^{0-6}\chi$: the simple connectivity indices from zero- to the sixth-order. $^{0,1,5}\chi^{v}_{p}$: the zero-, first- and fifth-order valence level connectivity indices. $^{6}\chi_{ch}$: the sixth-order chain connectivity index. $^{4}\chi_{p/c}$, $^{4}\chi^{v}_{p/c}$: the fourth-order path/cluster simple or valence level connectivity indices.

Electrotopological state indices S_i: $S_{C1',\ 6'}$: S_i for the C1′, 6′ positions. S_{O1}: S_i for the O1 position.

mobility μ_e of a solute is calculated from the experimental result as follows:

$$\mu_e = [(1/t_m) - (1/t_0)](L_{det}L_{tot})/V \quad (1)$$

where L_{det} is effective capillary length (to the detector), L_{tot} is total capillary length (cm), V is applied voltage and t_0 is migration time of a neutral marker. The t_0 was measured by injecting methanol. The other procedures of CE are as described earlier [17,18].

2.2. Structure entry, storage and topological index generation

The molecular structures of the flavonoids were sketched as hydrogen-suppressed diagrams and stored as connection tables. Topological indices of the flavonoids were calculated by the Molconn-X 1.0 software (Lowell H. Hall, Hall Associates Consulting, Eastern Nazarene College, Quincy, MA, USA). The molecular connectivity indices ${}^{m}\chi_t^{\nu}$ include simple or valence connectivity indices from zero- to tenth-order ${}^{0-10}\chi$, ${}^{0-10}\chi^{\nu}$; simple or valence cluster/path connectivity indexes ${}^{3,4}\chi_{c/pc}$, ${}^{3,4}\chi^{\nu}_{c/pc}$; the sixth-order chain simple or valence connectivity index ${}^{6}\chi_{ch}$, ${}^{6}\chi^{\nu}_{ch}$ [5]. The electrotopological state indices (S_i) of flavonoid skeletons include the S_i of carbons 2–10, 1′–6′ ($S_{C2-10,\ C1'-6'}$), S_i of oxygen 1 (S_{O1}) and S_i of 4-O, 5-O, 7-O and 4′-O (Fig. 1) [6]. The topological indices involved in following regression models were listed in Table 1. The indices found to have an inordinate number of zero values were withdrawn from consideration.

2.3. Statistical analysis

Systat 6.0 for Windows (Systat, Chicago, IL, USA) was used for statistical analysis. The statistical analysis included correlation and multivariate regression analyses. Correlation analysis is a multivariate procedure for measuring the strength of association between variables. The Pearson correlation for the correlation coefficients was applied. The regression analysis involved model validation. For this purpose, in a first step, several criteria were taken into consideration: (1) the multiple correlation coefficient R, an indicator of the amount of variation in the covariates accounted for by the model, (2) standard errors and relative standard errors, (3) the F ratio, an indicator of statistical credibility and fit of the models to the calculated values, (4) the number of indices included in the model and (5) multicollinearity problems. The next step involved testing the candidate models for outliers. Outliers can exert undue influence on the regression function, skewing the actual equation as it attempts to model the data. In a final step the stability and robustness of the model were evaluated by internal validation. The method omits one compound from the data set at a

Table 2
Pearson correlation matrix of topological indices

	${}^{0}\chi$	${}^{1}\chi$	${}^{2}\chi$	${}^{3}\chi$	${}^{4}\chi$
${}^{0}\chi$	1.000				
${}^{1}\chi$	1.000	1.000			
${}^{2}\chi$	0.999	0.999	1.000		
${}^{3}\chi$	0.998	0.998	0.999	1.000	
${}^{4}\chi$	0.998	0.998	0.998	0.999	1.000
${}^{5}\chi$	0.999	0.999	0.999	0.999	0.999
${}^{6}\chi$	0.999	0.998	0.998	0.997	0.996
${}^{4}\chi_{p/c}$	0.993	0.993	0.995	0.997	0.996
${}^{0}\chi^{\nu}$	0.996	0.995	0.994	0.989	0.989
${}^{1}\chi^{\nu}_{p}$	0.998	0.998	0.997	0.994	0.993
${}^{5}\chi^{\nu}_{p}$	0.992	0.992	0.991	0.988	0.987
${}^{4}\chi^{\nu}_{p/c}$	0.996	0.995	0.996	0.994	0.992
${}^{6}\chi_{ch}$	0.938	0.940	0.944	0.941	0.938
S_{O1}	0.813	0.812	0.802	0.786	0.790
$S_{C6'}$	0.379	0.380	0.361	0.334	0.334
$S_{C1'}$	0.188	0.189	0.169	0.152	0.150

	${}^{5}\chi$	${}^{6}\chi$	${}^{4}\chi_{p/c}$	${}^{0}\chi^{\nu}$	${}^{1}\chi^{\nu}_{p}$
${}^{5}\chi$	1.000				
${}^{6}\chi$	0.999	1.000			
${}^{4}\chi_{p/c}$	0.995	0.993	1.000		
${}^{0}\chi^{\nu}$	0.993	0.996	0.981	1.000	
${}^{1}\chi^{\nu}_{p}$	0.996	0.997	0.986	0.998	1.000
${}^{5}\chi^{\nu}_{p}$	0.992	0.993	0.980	0.992	0.996
${}^{4}\chi^{\nu}_{p/c}$	0.996	0.996	0.990	0.994	0.997
${}^{6}\chi_{ch}$	0.939	0.932	0.933	0.927	0.940
S_{O1}	0.794	0.810	0.763	0.850	0.826
$S_{C6'}$	0.350	0.368	0.284	0.444	0.415
$S_{C1'}$	0.157	0.166	0.112	0.238	0.211

	${}^{5}\chi^{\nu}_{p}$	${}^{4}\chi^{\nu}_{p/c}$	${}^{6}\chi_{ch}$	S_{O1}	$S_{C6'}$
${}^{5}\chi^{\nu}_{p}$	1.000				
${}^{4}\chi^{\nu}_{p/c}$	0.997	1.000			
${}^{6}\chi_{ch}$	0.949	0.941	1.000		
S_{O1}	0.798	0.800	0.700	1.000	
$S_{C6'}$	0.402	0.376	0.357	0.765	1.000
$S_{C1'}$	0.179	0.169	0.156	0.60	0.896

time and regenerates the model coefficients. The migration parameters of the flavonoids that were left out was then predicted.

3. Results and discussion

3.1. Optimised separation of flavonoids by CZE

After studies on the effects of borate concentrations and buffer pH, the optimised condition for separation of the flavonoids was obtained with 20 m*M* borate solution at pH 10.5. The separated peaks were identified through comparison of migration times, UV spectra as well as by spiking of individual standards. The electrophoretic mobility μ_e was used as the parameter to represent the migration of flavonoids because it does not incorporate the electroosmotic flow μ_{EOF}, which is a major factor to influence the repeatability of analysis in CZE.

3.2. Analysis of topological indices

Pearson correlation analysis indicated that the molecular connectivity indices were highly intercorrelated with one another (correlation between the indices $r_{ij}>0.90$) (Table 2). Thus, only one of the connectivity indices could be chosen in two- or multiple-covariate models in order to avoid multicollinearities, which reduce the predictive capability of models and render them unstable.

3.3. Model establishment and validation

The relationships between the topological indices and μ_e of the flavonoids were analysed by backward stepwise regression procedures. The regression

Table 3
Regression models of two covariates for 13 flavonoids in CZE

	Regression models (1–23) μ_e (response variable)=(regression coefficient 1±s[1])·covariate 1+(regression coefficient 2±s)·covariate 2+(constant±s)	s of estimate[1]	Relative s (%)[2]	R	F ratio	r_{ij}[3]
1	$\mu_e=(24.61\pm2.86)\cdot S_{C1'}+(1.47\pm0.11)\cdot {}^1\chi_p^v-(41.14\pm1.33)$	1.42	7	0.99	162.00	0.21
2	$\mu_e=(25.47\pm2.56)\cdot S_{C1'}+(1.07\pm0.07)\cdot {}^1\chi-(43.42\pm1.32)$	1.28	7	0.99	200.43	0.19
3	$\mu_e=(25.53\pm2.68)\cdot S_{C1'}+(0.69\pm0.05)\cdot {}^0\chi-(43.25\pm1.37)$	1.34	7	0.99	182.93	0.19
4	$\mu_e=(26.76\pm2.73)\cdot S_{C1'}+(1.95\pm0.14)\cdot {}^5\chi-(43.19\pm1.40)$	1.37	7	0.99	174.59	0.16
5	$\mu_e=(28.57\pm3.21)\cdot S_{C1'}+(2.12\pm0.18)\cdot {}^4\chi_{p/c}-(42.10\pm1.59)$	1.62	7	0.98	123.31	0.11
6	$\mu_e=(25.97\pm3.20)\cdot S_{C1'}+(5.34\pm0.45)\cdot {}^5\chi_p^v-(39.79\pm1.41)$	1.60	7	0.98	126.12	0.18
7	$\mu_e=(23.55\pm3.04)\cdot S_{C1'}+(0.87\pm0.07)\cdot {}^0\chi^v-(41.37\pm1.42)$	1.50	7	0.98	144.81	0.24
8	$\mu_e=(15.89\pm3.09)\cdot S_{C6'}+(1.29\pm0.18)\cdot {}^1\chi_p^v-(55.53\pm3.92)$	2.16	10	0.97	67.19	0.42
9	$\mu_e=(17.51\pm2.81)\cdot S_{C6'}+(1.69\pm0.21)\cdot {}^5\chi-(58.99\pm3.72)$	2.02	10	0.97	77.53	0.35
10	$\mu_e=(16.49\pm3.49)\cdot S_{C6'}+(4.59\pm0.73)\cdot {}^5\chi_p^v-(54.65\pm4.45)$	2.45	10	0.95	51.02	0.40
11	$\mu_e=(16.74\pm2.78)\cdot S_{C6'}+(0.61\pm0.07)\cdot {}^0\chi-(58.37\pm3.63)$	1.98	10	0.97	81.34	0.38
12	$\mu_e=(16.69\pm2.74)\cdot S_{C6'}+(0.94\pm0.11)\cdot {}^1\chi-(58.47\pm3.57)$	1.94	10	0.97	84.27	0.38
13	$\mu_e=(15.14\pm3.23)\cdot S_{C6'}+(0.77\pm0.11)\cdot {}^0\chi^v-(55.01\pm4.03)$	2.22	10	0.96	63.28	0.44
14	$\mu_e=(14.04\pm4.14)\cdot S_{O1}+(0.30\pm0.17)\cdot {}^0\chi-(110.17\pm20.88)$	2.90	10	0.94	35.13	0.81
15	$\mu_e=(13.92\pm4.09)\cdot S_{O1}+(0.47\pm0.26)\cdot {}^1\chi-(109.68\pm20.61)$	2.87	10	0.94	35.81	0.81
16	$\mu_e=(14.38\pm4.05)\cdot S_{O1}+(0.47\pm0.28)\cdot {}^2\chi-(111.96\pm20.45)$	2.92	10	0.94	34.68	0.80
17	$\mu_e=(14.60\pm3.90)\cdot S_{O1}+(0.49\pm0.29)\cdot {}^3\chi-(112.71\pm19.92)$	2.90	10	0.94	35.00	0.79
18	$\mu_e=(14.62\pm3.95)\cdot S_{O1}+(0.67\pm0.41)\cdot {}^4\chi-(113.51\pm19.86)$	2.92	10	0.94	34.59	0.79
19	$\mu_e=(14.60\pm4.00)\cdot S_{O1}+(0.78\pm0.48)\cdot {}^5\chi-(112.76\pm20.32)$	2.93	10	0.93	34.39	0.79
20	$\mu_e=(14.71\pm4.23)\cdot S_{O1}+(0.92\pm0.62)\cdot {}^6\chi-(112.88\pm21.53)$	2.98	10	0.93	32.90	0.81
21	$\mu_e=(15.50\pm3.84)\cdot S_{O1}+(0.74\pm0.50)\cdot {}^4\chi_{p/c}-(116.67\pm19.85)$	2.99	10	0.93	32.76	0.76
22	$\mu_e=(14.84\pm4.12)\cdot S_{O1}+(2.25\pm1.50)\cdot {}^4\chi_{p/c}^v(112.34\pm21.59)$	2.97	10	0.93	33.12	0.80
23	$\mu_e=(14.52\pm3.00)\cdot S_{O1}+(53.13\pm21.32)\cdot {}^6\chi_{ch}-(117.16\pm14.83)$	2.58	10	0.95	45.48	0.70

[1] s: Standard error; s of estimate: square root of the residual mean square.
[2] Relative s: (s of estimate)/(square root of the estimate mean square).
[3] r_{ij}: correlation between covariates (topological indices).
[4] Definitions of the covariates see Table 1.

models containing two indices were found to be optimum since addition of more indices resulted in only minor increase in R, coupled with little decrease in the standard error. This also was within reasonable limits to prevent overfitting of the data set due to the relative small numbers of observance.

In order to avoid the possibility of chance correlation, 23 two-covariate models with similar statistical results were selected (Table 3). The coefficients are reported with their 95% confidence intervals. In the internal model validation, the results showed that no one compound exerted undue influence on any of the models. Each of these two-covariate models included one molecule connectivity index ${}^{m}\chi_{t}^{v}$ and the other electrotopological state index S_i. The predictions for models 1 and 2 are shown in Figs. 2 and 3. For the two models, in no case was there a residual greater than twice the standard error (Table 4). The plots of residuals versus predicted μ_e are shown in Figs. 4 and 5.

3.4. Explanation of the correlation

Each of the two-covariate models related μ_e of a flavonoid to a linear combination of one ${}^{m}\chi_{t}^{v}$ and the other S_i. In CZE, μ_e of a given analyte can most simply be expressed in terms of its radius r, the charge on the molecule q and the viscosity of the separation medium η, as follows:

$$\mu_e = q/6\pi r\eta \quad (2)$$

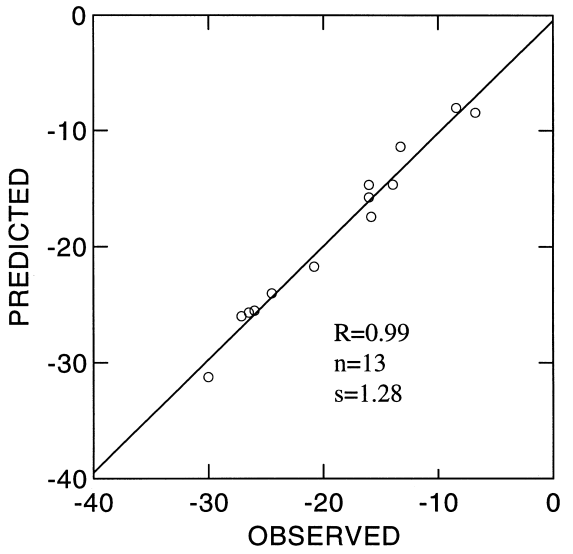


Fig. 3. Plot of predicted electrophoretic mobility μ_e versus observed μ_e for model 2 (see Table 3) (s: standard error of estimate).

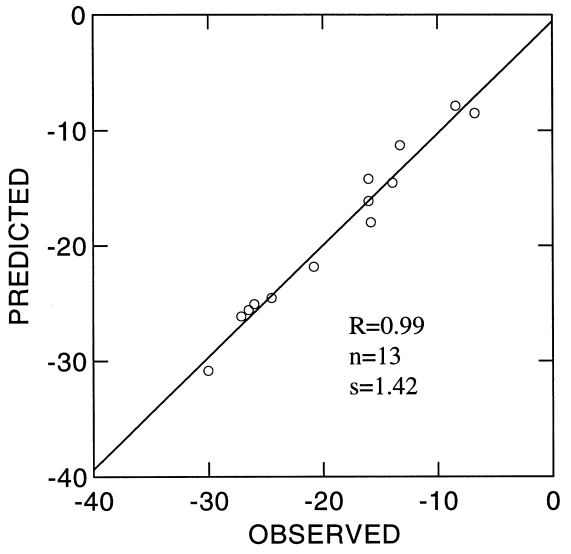


Fig. 2. Plot of predicted electrophoretic mobility μ_e versus observed μ_e for model 1 (see Table 3) (s: standard error of estimate).

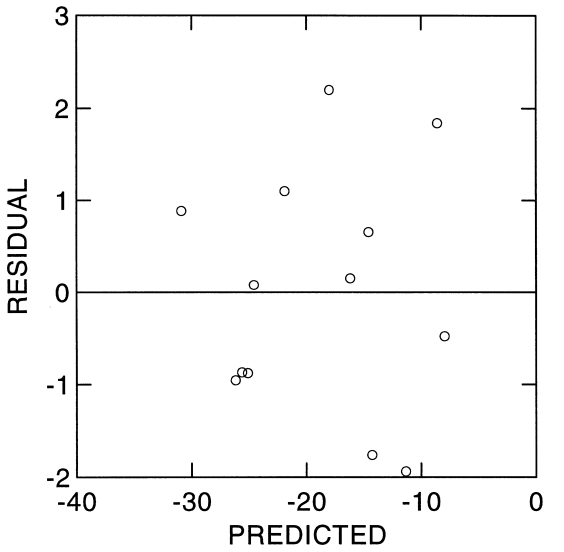


Fig. 4. Plot of residuals against predicted μ_e for model 1 (residuals = observed μ_e − predicted μ_e).

Table 4
Observed electrophoretic mobility μ_e and predicted μ_e

Name	Observed μ_e	Predicted μ_e from model 1	Residuals[1] from model 1	Predicted μ_e from model 2	Residuals from model 2
Epimedin B (1)	−6.72	−8.56	1.84	−8.45	1.73
Epimedin C (2)	−8.40	−7.92	−0.48	−8.04	−0.36
Icariin (3)	−13.24	−11.30	−1.94	−11.43	−1.81
Icariin I (4)	−13.91	−14.56	0.65	−14.69	0.78
Luteolin-3′7-O-diglucoside (5)	−15.79	−17.99	2.20	−17.46	1.67
Icariin II (6)	−16.00	−14.24	−1.76	−14.70	−1.30
Luteolin-7-O-glucoside (7)	−16.00	−16.15	0.15	−15.80	−0.21
Tricin (8)	−20.77	−21.87	1.10	−21.76	0.99
8-OCH_3-isoquercitrin (9)	−24.47	−24.55	0.08	−24.04	−0.43
Vitexin (10)	−25.97	−25.09	−0.88	−25.53	−0.44
Quercitrin (11)	−26.47	−25.60	−0.87	−25.69	−0.78
Avicularin (12)	−27.11	−26.15	−0.96	−25.99	−1.12
Luteolin (13)	−29.99	−30.87	0.88	−31.27	1.28

[1] Residuals = observed μ_e − predicted μ_e.

From this equation it is evident that small, highly charged species have high mobilities whereas large, minimally charged species have low mobilities. We selected model 1 (Table 3) as an example to explain the correlation. The model involved the connectivity index ${}^1\chi_p^v$ and the electrotopological state index $S_{C1'}$. ${}^1\chi_p^v$ can encode molecular size, degree of branching, molecular volume, surface area, topology of unsaturation and heteroatoms [5]. From the model we can deduce that the higher ${}^1\chi_p^v$ values of flavonoids (i.e., the larger the size of the flavonoids), the lower the mobility's of the flavonoids. In addition, ${}^1\chi_p^v$ can also reflect the molecular polarity. According to Kier and Hall [5], the index reflecting polarity may be written:

$$\text{Polarity} = ({}^1\chi - {}^1\chi_p^v) - {}^1\chi = -{}^1\chi_p^v \quad (3)$$

The polarity of molecule is inversely related to ${}^1\chi_p^v$. The higher ${}^1\chi_p^v$, the lower the polarity and the lower the μ_e of flavonoids.

Furthermore, the μ_e of the flavonoids decreased with increasing in $S_{C1'}$ according to the model. The S_i analysis considers those atoms common to all molecules in a series of flavonoids to represent the atomic level. The S_i for an atom i encodes information about both the topological environment and the electronic interactions of the atom with all other atoms in the molecule. The topological relationship is based on the graph distance to every other atom, which is related to the molecular size and shape [6,25,26]. The electronic aspect of the atom is based on its intrinsic state plus perturbation due to intrinsic state differences between it and other atoms in the molecule [6,25,26]. Thus, the S_i value also expresses the electronegativity difference between the atom and the other atoms in the molecule, which is related to the molecular polarity. So far, we can conclude that the topological indices ${}^1\chi_p^v$ and S_i could explain μ_e of the flavonoids in CZE.

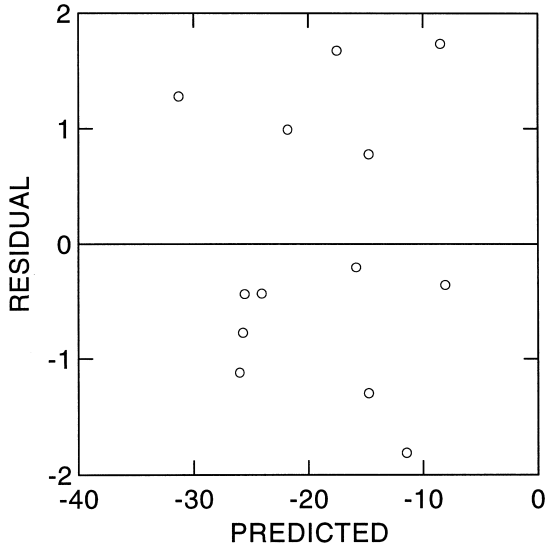


Fig. 5. Plot of residuals against predicted μ_e for model 2 (residuals = observed μ_e − predicted μ_e).

3.5. Structural factors affecting the observed and predicted μ_e

As seen in Table 4, the predicted μ_e for epimedin B (1) and C (2) are inverted by both models compared with the observed μ_e. A similar inversion occurs between luteolin-3′,7-O-diglucoside (5) and icariin II (6). Additionally the predicted μ_e values inverted for luteolin-3′,7-O-diglucoside (5) and luteolin-7-O-glucoside (7). In order to explain the discrepancies generated by the models, the structural factors affecting observed and predicted μ_e were analyzed.

According to the molecular structures of flavonoids, the structural factors affecting observed μ_e are (1) molecular size, (2) ionization degree of hydroxyl groups and (3) the complexation of flavonoids with borate buffer ion [23]. The ionization degree of hydroxyl groups is related to the number and position of the free hydroxyl groups on the flavonoid skeleton. In the alkaline borate buffer, the μ_e of each flavonoid increases with ionization of the hydroxyl group located on the flavonoid skeleton. The pK_a of phenoxyl groups located at different positions of flavonoid skeletons varies from 7.3 to 12.5 [18]. The complexation of flavonoids with borate ion is related to the number, type and position of attached sugar units, and to the presence of vicinal hydroxyl groups on the flavonoid skeletons. The μ_e of each flavonoid increases as it reacts with borate to form a borate–anionic complex. The magnitude of the borate complexation depended on the number of boration sites in sugar and aglycone moieties and consequently on the sugar configuration. The occurrence of *cis*-di hydroxyl vicinal groups in sugars lead to form the stable complex with borate and consequently high μ_e.

The factors affecting observed μ_e were shown in Table 5. From that, we can deduce that the observed μ_e can reasonably be explained by the above structural factors. In addition, the observed μ_e were compared with those of published studies [23,24]. In general, the results are consistent with each other.

Table 6 shows how the two indices involved in model 1 affected the predicted μ_e, respectively. In the case of epimedin B and C, this discrepancy can be explained by the fact the indices do not contain the information about whether the flavonoids can form the stable complex with borate. In the other two cases, the influences of $^1\chi_p^v$ are consistent with the effects of molecular structures. The discrepancies were generated by the influences of $S_{C1'}$. This indicated that molecular connectivity index $^1\chi_p^v$ better represented the structural properties of flavonoids than S_i.

Despite the discrepancies, in no case was there a residual greater than twice the standard error (Table 4). Except the discrepancies, all other predicted μ_e

Table 5
Structural factors affecting the observed μ_e

Compound	Structural factors			Observed μ_e
	Molecular mass (M_r)	Number and position of OH	Occurrence of *cis*-1,2-diols	
Epimedin B (1)	808.8	5-OH	1 rha (+), 1 xyl (−)	1<2
Epimedin C (2)	822.8	5-OH	2 rha (++)	Dual effect
	1 μ_e>2 μ_e^1	1 μ_e=2 μ_e^2	1 μ_e<2 μ_e^3	complexation effect>M_r effect
Luteolin-3′,7-O-diglucoside (5)	610.5	4′-OH (pK_a 8.3–9.5)	2 glc (−)	5<6
Icariin II (6)	514.5	7-OH (pK_a 7.3–8.2)	1 rha (+)	
	5 μ_e<6 μ_e^1	5 μ_e<6 μ_e^2	5 μ_e<6 μ_e^3	
Luteolin-3′,7-O-diglucoside (5)	610.5	4′-OH	glc (−)	5<7
Luteolin-7-O-glucoside (7)	448.4	3′,4′-di-OH	3′,4′-di-OH (+)	
	5 μ_e<7 μ_e^1	5 μ_e<7 μ_e^2	5 μ_e<7 μ_e^3	

[1] The μ_e of compounds (No.) were estimated according to M_r.
[2] The μ_e of compounds (No.) were estimated according to number and position of OH.
[3] The μ_e of compounds (No.) were estimated according to occurrence of *cis*-1,2 diols.

Table 6
Structural descriptors in model 1

Compound	Structural descriptors		Predicted μ_e (model 1)
	$^1\chi_p^v$ influence	$S_{C1'}$ influence	
Epimedin B (1)			1<2
Epimedin C (2)	1 μ_e>2 μ_e^1	1 μ_e<2 μ_e^2	Dual effect $^1\chi_p^v$ influence>$S_{C1'}$ influence
Luteolin-3′,7-O-diglucoside (5)			5>6
Icariin II (6)	5 μ_e<6 μ_e^1	5 μ_e>6 μ_e^2	Dual effect $S_{C1'}$ influence>$^1\chi_p^v$ influence
Luteolin-3′,7-O-diglucoside (5)			5>7
Luteolin-7-O-glucoside (7)	5 μ_e<7 μ_e^1	5 μ_e>7 μ_e^2	Dual effect $S_{C1'}$ influence>$^1\chi_p^v$ influence

[1] The μ_e of compounds (No.) were estimated according to $^1\chi_p^v$ values.
[2] The μ_e of compounds (No.) were estimated according to $S_{C1'}$ values.

are consistent with the observed μ_e. Thus the developed models, to a certain extent, could be used to predict μ_e of flavonoids based on one $^1\chi_p^v$ and the other S_i.

4. Conclusions

The present study has demonstrated that the molecule connectivity index and the electrotopological state index can successfully be used to develop structure–migration models for the flavonoids in CZE. The models derived are statistically stable, fit the data well and show a high degree of correlation between observed and predicted values.

Acknowledgements

This work was supported by a grant (to H.-R.L.) from the Finnish Cultural Foundation, which is greatly appreciated.

References

[1] D. Bonchev, Information Theoretic Indices for Characterisation of Chemical Structures, Wiley, New York, 1983.
[2] D.H. Rouvray, Sci. Am. 255 (1986) 36.
[3] N. Trinajstić, Chemical Graph Theory, Vol. II, CRC Press, Boca Raton, FL, 1983, pp. 105–140.
[4] M. Randić, J. Am. Chem. Soc. 97 (1975) 6609.
[5] L.B. Kier and L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Research Studies Press, Letchworth, 1986.
[6] L.B. Kier, L.H. Hall, Pharm. Res. 7 (1990) 801.
[7] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley-Interscience, New York, 1987.
[8] L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[9] A. Sabljić, in W. Karcher and J. Devillers (Editors), Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, ECSC, EEC, EAEC, Brussels and Luxembourg, 1990, pp. 61–82.
[10] F.J. Garcia-March, G.M. Antín-Fos, F. Prez-Giménez, M.T. Salabert-Salvador, R.A. Cercs-del-Pozo, J.V. de Julián-Ortiz, J. Chromatogr. A 719 (1996) 45.
[11] T.F. Woloszyn, P.C. Jurs, Anal. Chem. 64 (1992) 3059.
[12] R. Kaliszan, A. Kaliszan, T.A.G. Noctor, J. Chromatogr. 609 (1992) 69.
[13] S. Yang, J.G. Bumgarner, L.F.R. Kruk, M.G. Khaledi, J. Chromatogr. A 721 (1996) 323.
[14] P. Lukkari, H. Vuorela, M.-L. Riekkola, J. Chromatogr. A 652 (1993) 451.
[15] M. Salo, H. Sirén, P. Volin, S. Wiedmer, H. Vuorela, J. Chromatogr. A 728 (1996) 83.
[16] P.G. Pietta, P.L. Mauri, A. Rava, G. Sabbatini, J. Chromatogr. 549 (1991) 367.
[17] H.-R. Liang, H. Sirén, M.-L. Riekkola, P. Vuorela, H. Vuorela, R. Hiltunen, J. Chromatogr. A 746 (1996) 123.
[18] H.-R. Liang, H. Sirén, P. Jyske, M.-L. Riekkola, P. Vuorela, H. Vuorela, R. Hiltunen, J. Chromatogr. Sci. 35 (1997) 117.
[19] P.G. Pietta, P.L. Mauri, A. Bruno, C. Gardana, Electrophoresis 15 (1994) 1326.
[20] U. Seitz, P.J. Oefner, S. Nathakarnkitkool, M. Popp, G.K. Bonn, Electrophoresis 13 (1992) 35.
[21] T.K. McGhie, J. Chromatogr. 634 (1993) 107.
[22] Ph. Morin, F. Villard, M. Dreux, J. Chromatogr. 628 (1993) 153.

[23] T.K. McGhie, K.R. Markham, Phytochem. Anal. 5 (1994) 121.
[24] K.R. Markham, T.K. McGhie, Phytochem. Anal. 7 (1996) 300.
[25] L.B. Kier, L.H. Hall, J.W. Frazer, J. Math. Chem. 7 (1991) 229.
[26] L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 35 (1995) 1039.