Quantitative structure–retention relationship modelling of esters on stationary phases of different polarity

Érica Silva Souza [a], Carlos Alberto Kuhnen [b], Berenice da Silva Junkes [c],
Rosendo Augusto Yunes [a,*], Vilma Edite Fonseca Heinzen [a,*]

[a] Departamento de Química, Universidade Federal de Santa Catarina, Campus Universitário, Trindade, Florianópolis, Santa Catarina 88040-900, Brazil
[b] Departamento de Física, Universidade Federal de Santa Catarina, Campus Universitário, Trindade, Florianópolis, Santa Catarina 88010-970, Brazil
[c] Centro Federal de Educação Tecnológica de Santa Catarina, Avenida Mauro Ramos, n° 950, Centro, Florianópolis, Santa Catarina 88020-300, Brazil

ARTICLE INFO

ABSTRACT

The semi-empirical electrotopological index, I_{SET}, used for QSRR models, was developed and optimized to describe the chromatographic retention of saturated esters on the five different stationary phases (SE-30, OV-7, DC-710, OV-225 and XE-60). The simple linear regressions between the retention indices and the proposed index were of good statistical quality, high internal stability and good predictive ability for external groups, especially for the stationary phase with low polarity, showing that the specific molecular interactions occur on highly polar phases. For the esters, the interactions between the molecules and the stationary phase are slowly increased relative to hydrocarbons due to the charge redistribution that occurs in the presence of the heteroatom. These facts were included in the calculation of I_{SET} through a small increase in the SET_i values for heteroatoms and the carbon atoms attached to them. The increase in the SET_i values originates from the dipole moment of the whole molecule and an equivalent local dipole moment related to the net charges of the atoms belonging to the functional group and the carbon atoms attached to them. The polarity of the stationary phases, indicated by the retention polarity (P_R) given by Tarján, is reflected in the intercept of the equations obtained for each stationary phase. Thus, a single combined QSRR model was generated with a satisfactory predictive quality, including a parameter that represents the polarity retention of all stationary phases studied.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Several researchers have used multivariate regression methods and recognition patterns with a variety of molecular descriptors to study the relationships between chromatographic retention and molecular structure for a large class of organic compounds. Recently, Kaliszan and Héberger published interesting reviews on publications in the QSRR field [1–4]. As was indicated by Kaliszan's review [1], the importance of the QSRR in chromatography is due to the fact that it is a unique method, which can yield a great amount of quantitatively comparable, precise, and reproducible retention data for large sets of structurally diversified compounds (analytes). Therefore, quantitative structure– (chromatographic) retention relationships (QSRR) have been considered as a model to establish strategy and methods of property predictions. According to Héberger's review [2], if the QSRR models are not sufficient for

identification purposes; still the models are useful to elucidate retention mechanisms, to optimize the separation of complex mixtures or to prepare experimental designs.

For the purpose of prediction and to obtain insights regarding the mechanism of chromatographic retention, correlations between retention indices on stationary phases having different polarities and molecular descriptors have been investigated [5–11].

The use of appropriate molecular descriptors is fundamental to obtaining good correlations between the retention data and the molecular structure and several topological, geometric, electronic and quantum-chemical descriptors have been already used [1,2,12–22].

Our research group has previously developed a semi-empirical topological index (I_{ET}), based on the hypothesis that the interaction between the molecules and the stationary phase is determined by the number and electronic properties of the carbon atoms and by the steric hindrance to these interactions resulting from other carbon atoms attached to it [23]. The index (I_{ET}) was initially developed to predict the chromatographic retention for alkanes and alkenes and, subsequently, was extended to other classes of

organic compounds [5,6,23–30]. The equation obtained to calculate the I_{ET} was generated using a molecular graph, the values attributed to each type of carbon atom (primary, secondary, tertiary and quaternary) and the functional groups based on the results of the experimental chromatographic behavior, and further supported by theoretical considerations. It can be concluded that this topological index is strongly related to the chromatographic retention.

In order to obtain a molecular descriptor not directly related to the retention phenomena, a new molecular descriptor was proposed which can, in principle, express all of the electronic and geometric properties of molecules and their interactions. Since the interactions between the solute and the stationary phase are dispersive for alkanes and electrostatic for alkenes, the chromatographic retention is dependent on the electronic charge distribution of each carbon atom of these molecules. In this regard, our group recently investigated, in a previous study [31], the relation between the values attributed by the semi-empirical topological index (for primary, secondary, tertiary and quaternary carbon atoms) and the net atomic charges of the carbon atoms in the molecules. For alkanes and alkenes it was possible to develop a new semi-empirical electrotopological index (I_{SET}) for QSRR models [31] based on the fact that the interactions between the solute and the stationary phase are due to electrostatic and dispersive forces. The I_{SET} was generalized to include polar molecules such as ketones and aldehydes [32]. The presence of heteroatoms like oxygen changes considerably the charge distribution of the corresponding hydrocarbons, resulting in a partial increasing in the interactions between the solute and the stationary phase. Therefore, an appropriate way to calculate the I_{SET} was developed taking into account the dipole moment of the whole molecule and also an equivalent local dipole moment related to the net charges of the atoms belonging to functional group and the carbon atoms attached to them.

The main goal of this study is to extend the new index, I_{SET}, to a series of saturated esters to predict the retention indices of these compounds and to verify its predictive ability when applied to stationary phases of different polarity. Another aim of this study is to apply this index in an attempt to interpret the molecular mechanism of interactions between eluents and stationary phases with different polarities.

2. Materials and methods

2.1. Calculation of semi-empirical electrotopological index for esters

The interaction between polar molecules like esters and a stationary phase is strongly dependent on the overall electronic charge distribution of the molecule and clearly the interaction of each carbon atom is driven also by the electronic charges of the neighboring carbon atoms. This means that the chromatographic retention has a strong dependence on the electronic charge distribution of each carbon atom and oxygen atom of these molecules. For alkanes and alkenes [31] the inclusion of the steric effects of the neighboring carbon atoms leads to the semi-empirical electrotopological index that was calculated according to Eq. (1),

$$I_{SET} = \sum_{i,j} (SET_i + \log SET_j). \qquad (1)$$

In the above expression the i sum is over the all the atoms of the molecule (excluding the H atoms) and the j is an inner sum of atoms attached to the i atom. A simple linear regression equation was obtained between the values of the carbon atoms, the SET_i values, and the net atomic charges (δ_i) of these atoms, as can be seen in Eq. (2).

$$SET_i = -1.77125\delta_i + 0.62417. \qquad (2)$$

From Eq. (2) it is clear that knowledge of the atomic charges is sufficient to calculate the SET_i value for all kinds of carbon atoms and not only the values given by the carbon models (that is 1.0, 0.9, 0.8 and 0.7). From the above equations and from a semi-empirical quantum-mechanics calculation, using the AM1 method where the atomic charges of the carbon atoms were obtained by Mulliken population analysis, it was possible to obtain a good correlation between the I_{SET} and the experimental retention index for alkanes and alkenes [31].

In this study, the question concerning the chromatographic retention of organic functions such as esters, which are polar molecules, is addressed. By considering the stationary phase as non-polar material, the interactions between these molecules and the stationary phase are electrostatic, with a contribution from dispersive forces. In this regard, we generated the I_{SET} for polar molecules considering the fact that the interactions between these molecules and the stationary phase are slowly increased relative to the corresponding hydrocarbons. This increase in the interactions is due to the charge redistribution that occurs in the presence of the heteroatoms. As in the case of ketones and aldehydes [32] for esters we have included the above-mentioned facts by calculating the retention index as in Eq. (1), also using Eq. (2) for heteroatoms, but including a slight alteration, i.e. incorporating the effects of the dispersive dipolar interactions. For esters the major effects related to the charge distribution are due to the presence of the two oxygen atoms and they occur on these atoms and in their neighborhood (their adjacent carbon atoms). The excess charge on these atoms leads to electrostatic interactions that are stronger than the weak dispersive dipolar interactions. For esters, all these facts were included in the calculation of the retention index through a small increase in the SET_i values for heteroatoms and the carbon atoms attached to them. One way to obtain a small increase in the SET_i values is to consider the molecular dipole moment, which is used as a quantum-chemical descriptor in QSPR/QSAR studies [2,7,33]. In this study, this physical property was included in the calculation of the retention index. As in the case of ketones and aldehydes, it was verified that the introduction of the dipole moment of the molecule is not sufficient to explain the chromatographic behavior of these molecules. Thus, it was necessary to introduce an equivalent local dipole moment of the (–COOC–) group that participates in the increase in the retention value. This was carried out by multiplying the SET_i values of the atoms belonging to the O=C–O–C group by a function A_μ which is dependent on the dipole moment of molecule and the net charge on the oxygen and carbon atoms (to include both the electrostatic and dispersive interactions). The same approach used for ketones and aldehydes was applied for esters [32], that is, considering that A_μ has a weaker dependence on the dipole moment than the linear one. As in other chromatographic behaviors, this dependence was found to be consistently logarithmic, leading to a definition of A_μ that differs slightly from unity and which is logarithmically dependent on the dipole moment of the molecule as in Eq. (3),

$$A_\mu = 1 + \log\left(1 + \frac{\mu}{\mu_F}\right) \qquad (3)$$

where μ is the calculated molecular dipole moment and μ_F is an equivalent local dipole moment (in units of dipole moment) which is dependent on the charges of the atoms belonging to the O=C–O–C group. Clearly μ_F must be directly related to the net charge on the oxygen atoms since it must reflect some contribution of the electrostatic interaction between these atoms and the stationary phase. In this regard, μ_F may also be related to the atomic charge

on the carbon atom of the functional group C=O or related to the difference between the atomic charges on these atoms. Hence, μ_F can be defined in different ways and some definitions of μ_F can be used in preliminary calculations. For ketones and aldehydes, good definitions of μ_F have been achieved [32], relating it to the bond length of the C=O bond and net charges of the C and O atoms. Esters have two oxygen atoms and thus it is possible to define two local functions, one being dependent on the charges and bond length of the C=O_1 bond and another on the charges and bond length of the C–O_2 bond. Therefore, it was necessary to perform some calculations with different definitions for the equivalent local dipole moment. After the preliminary calculations it was found that for esters the charge difference, $Q_O - Q_C$, does not give reasonable results because the charges on the oxygen atoms mask the charge on the carbonyl carbon. As expected, our best choice was for the esters $\mu_{F1} = d_1|Q_{O_1}|$ and $\mu_{F2} = d_2|Q_{O_2}|$, where d_1 and d_2 are the calculated C_1=O_1 and C_1–O_2 bond lengths and $|Q_{O_1}|$ and $|Q_{O_2}|$ are the absolute values of the atomic charges on oxygen atoms (O_1 and O_2). The equivalent local dipole moment is then calculated as the magnitude of the vectorial sum of two dipole moments, that is, $\mu_F = (\mu_{F1}^2 + \mu_{F2}^2 + 2\mu_{F1}\mu_{F2}\cos\theta)^{1/2}$, where θ is the angle between the C=O_1 and C–O_2 bonds. For formates, a specific charge distribution occurs in the polar region of the molecules and the best mathematical model for the local moment was taking into account the contribution to the electrostatic interactions that originate from the atomic charges of the oxygen atoms, the carbon atoms and the H atom belonging to the $C_1O_1O_2C_{Al}$ group of the formate molecules (C_{Al}, represents the carbon of alcoholic side). Thus, the equivalent dipoles were built from the net charges of HC_1O_1, HC_1O_2 and O_2C_{Al} groups of atoms. The equivalent dipoles associated with these net charges are: $\mu_{F1} = d_1|Q_H + Q_{C_1} - Q_{O_1}|$, $\mu_{F2} = d_2|Q_H + Q_{C_1} - Q_{O_2}|$ and $\mu_{F3} = d_3|Q_{CA1} + Q_{O_2}|$ where d_1 and d_2 are the calculated C_1=O_1 and C_1–O_2 bond lengths and d_3 is the calculated C_{Al}–O_2 bond length. In a first approach, the local moments μ_{F2} and μ_{F3} are taken as being collinear and another equivalent dipole is obtained with the difference between μ_{F2} and μ_{F3} that is, $\mu_{F4} = \mu_{F2} - \mu_{F3}$ and the final equivalent local moment is calculated as above, that is, $\mu_F = (\mu_{F1}^2 + \mu_{F4}^2 + 2\mu_{F1}\mu_{F4}\cos\theta)^{1/2}$ where θ is the angle between the C=O_1 and C–O_2 bonds. Hence, for formates the charge of the hydrogen atom attached to the carbon atom of the COO functional group is also considered, as in the case of aldehydes [32], because the charge of the H atom contributes explicitly to the positive charge of the local polar region of the molecule. The above-mentioned best definitions for μ_F imply that the present approach to calculating the retention index considers important polar features of the organic functions, such as ketones, aldehydes and esters, through the information carried by the local moment μ_F. In other words, according to Eq. (3) there is an increase in the interaction between the molecules and the stationary phase due to the presence of a dipole moment and this contribution may be screened by the charge located on the heteroatoms and the carbon atom of the C=O group if $\mu_F > \mu$, or may be increased if $\mu_F > \mu$. In the case of esters, the local function μ_F is less than the dipole moment showing that A_μ has an appreciable contribution from the atomic charges on those atoms. This verifies the contribution of the oxygen atom to the electrostatic interaction between the solute and the stationary phase.

Therefore, in the case of esters the I_{SET} value is here calculated as in Eq. (4),

$$I_{SET} = \sum (A_\mu SET_i + \log A_\mu SET_j) \tag{4}$$

where the SET_i values are obtained using Eq. (2) through AM1 calculations of the net atomic charges. As mentioned above, Eq. (4) is calculated by multiplying the SET_i values of the atoms

belonging to the $C_1O_1O_2C_{Al}$ group by the dipolar function A_μ which is taken as unity for the remaining carbon atoms of the molecules. The above expression for I_{SET} reduces to Eq. (1) when the dipole moment of the molecule is zero (or near zero), as is the case for alkanes and alkenes since $A_\mu = 1$ for $\mu = 0$. Hence, Eq. (4) is a general definition for the electrotopological index that can be applied to different organic functions, which are specified through appropriate definitions for the equivalent local moment μ_F. The preliminary applications of I_{SET} as given by Eq. (4) showed that this expression overestimates the calculated retention index for branched esters and underestimates the results for methyl esters. This finding reveals the need to consider other definitions for the local moment μ_F for branched esters and methyl esters. However, another easy choice is to take into account the steric effects for the branched esters and methyl esters. The simplest way to do this is to consider the steric hindrance of the C_{Al} carbon atom of the $C_1O_1O_2C_{Al}$ group and the carbon atom attached to the acid side of the COO functional group (here named the C_{Ac} carbon). As seen in Eq. (1), the $\log SET_j$ factor gives, precisely, the steric effect of atom j. Thus, to include a steric correction (sc) in Eq. (4) for branched esters the term sc = $n\log SET(C_{Ac}) + n\log SET(C_{Al})$ was added, where n is the number of branches of the ester. On the other hand, for methyl esters the C_{Al} carbon is bound to three H atoms and it is necessary to remove the overestimated steric effects of the $\log A_\mu$ SET_j terms in Eq. (4). For methyl esters this is easily achieved by including a second steric correction (ssc) by adding to Eq. (4) the term ssc = $- \log SET(C_{Al})$. Very good results were obtained using this approach, which reveals that in this model the complex steric effects in branched esters can be included simply by considering the steric hindrance using the net charge (through the SET values) of the two carbon atoms bound to the alcoholic and acid sides of the COO functional group. The calculation of I_{SET} for a large amount of molecules is easily carried out by means of a FORTRAN code developed in our lab that calculates I_{SET} by reading the output data (calculated net charges, dipole moment and atomic positions) from AM1 semi-empirical calculations.

2.2. Data set

The Kováts indices for aliphatic esters on five stationary phases, SE-30, OV-7, DC-710, XE-60 and OV-225, at 150 °C, were taken from the literature [10,11,34,35]. The esters and their gas chromatographic retention indices are shown in Table 1. The five stationary phases containing non-, low-, medium- and high-polarity all comprised polysiloxanes. The characteristics of the stationary phases, as well as the polarity of the McReynolds constants [10], are summarized in Table 2.

The semi-empirical quantum-mechanics calculations were performed using the AM1 method included in the Hyperchem program package [36]. The calculations were carried out with optimized initial geometries obtained from molecular mechanics (MM) calculations. These were subsequently optimized using the AM1 semi-empirical method [36,37], with the Polak-Ribiere algorithm and gradient minimization techniques where the convergence limit was 0.0001 and the RMS gradient 0.0001 kcal $(A mol)^{-1}$. The atomic charges of the carbon atoms and oxygen atoms are calculated using Mulliken population analysis. Mulliken population analysis allows a net charge to be assigned to each atom in the molecule. The value of this net charge (δ_A) is the difference between the electronic charge of the isolated atom (Z) and the calculated charge of the bound atom (q_A), that is, $\delta_A = Z - q_A$ [37,38]. It is important to note that the present calculations use Eq. (2), which was obtained from the AM1 calculations for alkanes and alkenes [31]. Therefore, it is necessary to calculate the net

Table 1
Values of experimental retention indices for five stationary phases SE-30, OV-7, DC-710, XE-60, OV-225 and semi-empirical electrotopological indices (I_{SET}).

Number	Compound	SE-30	OV-7	DC-710	XE-60	OV-225	I_{SET}
1	Propyl formate	608	662	703	821	834	6.0387
2	Butyl formate	709	766	805	923	934	6.8701
3	Pentyl formate	807	866	903	1021	1031	7.6858
4	Hexyl formate	906	968	1002	1120	1128	8.5112
5	Octyl formate	1104	1171	1201	1318	1324	10.1444
6	Methyl acetate	509	559	609	737	759	5.2056
7	Propyl acetate	695	755	796	920	932	6.8215
8	Butyl acetate	794	855	895	1022	1030	7.6480
9	Pentyl acetate	891	958	996	1122	1132	8.4025
10	Heptyl acetate	1084	1151	1181	1299	1305	10.0882
11	Octyl acetate	1182	1252	1280	1397	1401	10.9107
12	Isopropyl acetate	643	686	722	842	840	6.3210
13	Isopentyl acetate	859	922	955	1086	1087	8.1012
14	Methyl propionate	617	667	710	837	851	5.9612
15	Ethyl propionate	692	743	783	897	905	6.7012
16	Propyl propionate	789	851	884	1003	1013	7.5108
17	Pentyl propionate	980	1047	1082	1202	1211	9.1446
18	Hexyl propionate	1074	1142	1176	1300	1307	9.9675
19	Heptyl propionate	1173	1242	1270	1388	1392	10.7753
20	Isopropyl propionate	733	785	817	920	928	7.0056
21	Isobutyl propionate	848	906	938	1057	1062	8.0329
22	Isopentyl propionate	948	1014	1042	1156	1169	8.7925
23	Methyl butyrate	702	767	803	923	938	6.7703
24	Ethyl butyrate	778	840	876	986	999	7.5262
25	Butyl butyrate	969	1036	1069	1187	1193	9.1349
26	Pentyl butyrate	1062	1131	1164	1284	1288	9.9412
27	Hexyl butyrate	1156	1225	1258	1380	1388	10.7642
28	Isopropyl butyrate	820	878	905	1010	1019	7.8011
29	Isobutyl butyrate	933	993	1023	1143	1146	8.7154
30	Isopentyl butyrate	1029	1093	1124	1243	1247	9.5886
31	Ethyl pentanoate	876	941	975	1085	1105	8.3094
32	Propyl pentanoate	971	1036	1069	1179	1196	9.0964
33	Butyl pentanoate	1063	1130	1163	1279	1289	9.9220
34	Pentyl pentanoate	1155	1222	1257	1375	1382	10.7282
35	Hexyl pentanoate	1247	1316	1349	1470	1476	11.5513
36	Isopropyl pentanoate	915	971	1000	1105	1116	8.5874
37	Isobutyl pentanoate	1027	1087	1117	1235	1240	9.6203
38	Isopentyl pentanoate	1132	1185	1217	1333	1340	10.3762
39	Methyl hexanoate	902	974	1006	1133	1140	8.3867
40	Ethyl hexanoate	976	1045	1073	1196	1203	9.1100
41	Butyl hexanoate	1156	1231	1256	1377	1383	10.7431
42	Pentyl hexanoate	1246	1325	1349	1471	1508	11.5493
43	Hexyl hexanoate	1337	1417	1440	1565	1575	12.3724
44	Isopropyl hexanoate	1008	1074	1115	1207	1206	9.4076
45	Isobutyl hexanoate	1119	1189	1211	1332	1333	10.4412
46	Isopentyl hexanoate	1212	1287	1308	1430	1432	11.1961
47	Methyl heptanoate	998	1062	1094	1212	1219	9.1928
48	Ethyl heptanoate	1068	1135	1165	1283	1289	9.9322
49	Propyl heptanoate	1155	1224	1253	1370	1375	10.7395
50	Isopropyl heptanoate	1105	1172	1202	1319	1325	10.2120
51	Isobutyl heptanoate	1213	1284	1311	1428	1432	11.2465
52	Methyl octanoate	1094	1162	1192	1309	1315	10.0329
53	Isopropyl octanoate	1200	1270	1298	1415	1419	11.0326
54	Isobutyl octanoate	1306	1379	1404	1521	1524	12.0669
55	Ethyl decanoate	1354	1428	1452	1569	1571	12.3802
56	Ethyl isobutyrate	732	793	831	968	935	7.0167
57	Butyl isobutyrate	931	987	1019	1130	1136	8.6499
58	Pentyl isobutyrate	1024	1082	1115	1225	1230	9.4558
59	Hexyl isobutyrate	1117	1177	1208	1323	1323	10.2790
60	Isopropyl isobutyrate	780	829	849	953	947	7.5115
61	Isobutyl isobutyrate	899	946	975	1096	1086	8.5664
62	Isopentyl isobutyrate	994	1045	1075	1196	1188	9.2980
63	Methylpropyl isobutyrate	892	954	989	1106	1115	8.3204
64	Methyl isopentanoate	763	823	854	969	982	7.2346
65	Butyl isopentanoate	1021	1081	1111	1224	1231	9.5371
66	Pentyl isopentanoate	1112	1174	1204	1318	1325	10.3430
67	Hexyl isopentanoate	1204	1267	1296	1414	1417	11.1662
68	Isopropyl isopentanoate	874	923	950	1049	1064	8.1431
69	Isobutyl isopentanoate	985	1039	1066	1177	1184	9.2330
70	Isopentyl isopentanoate	1081	1136	1164	1278	1281	9.9907
71	Methyl isohexanoate	875	936	970	1094	1104	8.1146
72	Ethyl isohexanoate	943	1004	1036	1158	1163	8.7801
73	Propyl isohexanoate	1035	1094	1128	1249	1253	9.5870
74	Butyl isohexanoate	1125	1189	1220	1340	1345	10.4121
75	Pentyl isohexanoate	1215	1281	1313	1433	1438	11.2181
76	Isopropyl isohexanoate	979	1028	1059	1169	1172	9.0156

Table 1 (*Continued*)

Number	Compound	SE-30	OV-7	DC-710	XE-60	OV-225	I_{SET}
77	Isobutyl isohexanoate	1089	1147	1175	1296	1309	10.1081
78	Isopentyl isohexanoate	1181	1244	1272	1393	1394	10.8636
79	Methyl 2-methylbutyrate	762	820	858	976	987	7.2145
80	Ethyl 2-methylbutyrate	832	892	928	1045	1055	7.8835
81	Propyl 2-methylbutyrate	924	987	1021	1138	1146	8.6905
82	Butyl 2-methylbutyrate	1016	1082	1113	1231	1238	9.5156
83	Pentyl 2-methylbutyrate	1110	1178	1207	1325	1330	10.3217
84	Hexyl 2-methylbutyrate	1204	1275	1302	1419	1423	11.1447
85	Isobutyl 2-metyllbutyrate	980	1044	1077	1194	1202	9.1567
86	Isopentyl 2-metyllbutyrate	1072	1139	1169	1287	1293	9.9145
87	Methyl 2-methylpentanoate	853	908	945	1054	1065	8.0240
88	Ethyl 2-methylpentanoate	917	973	1006	1118	1121	8.6899
89	Propyl 2-methylpentanoate	1009	1064	1096	1204	1213	9.4969
90	Butyl 2-methylpentanoate	1097	1157	1189	1296	1303	10.3210
91	Isopropyl 2-methylpentanoate	952	999	1021	1121	1128	8.8739
92	Isobutyl 2-methylpentanoate	1064	1116	1144	1254	1257	9.9628
93	Isopentyl 2-methylpentanoate	1153	1208	1238	1352	1350	10.7184
94	Methyl 2-ethylbutyrate	845	899	934	1049	1053	7.8968
95	Ethyl 2-ethylbutyrate	914	969	1000	1122	1113	8.5635
96	Propyl 2-ethylbutyrate	1005	1060	1092	1204	1206	9.3699
97	Pentyl 2-ethylbutyrate	1183	1241	1273	1390	1387	11.0014
98	Hexyl 2-ethylbutyrate	1273	1332	1364	1482	1478	11.8248
99	Isobutyl 2-ethylbutyrate	1060	1112	1139	1253	1249	9.8351
100	Isopentyl 2-ethylbutyrate	1149	1202	1232	1352	1343	10.5429

Table 2
Statistical results of the final models for *RI* and the McReynolds constants for five stationary phases using 100 esters.

Equation number	Stationary phase	Mp	b_0	b_1	r	r^2	r^2_{CV}	s	f
5	SE-30	217	−74.7563	115.0361	0.99905	0.9981	0.9980	7.65	51777
6	OV-7	592	−39.5558	117.8831	0.9986	0.9973	0.9972	9.42	35905
7	DC-710	827	11.1427	115.8248	0.9984	0.9968	0.9967	10.01	30690
8	XE-60	1785	128.1568	115.8136	0.9971	0.9942	0.9940	13.51	16852
9	OV-225	1813	141.4299	115.0095	0.9959	0.9918	0.9914	16.02	11814

Mp: McReynolds constants [10]; b_0: intercept; b_1: slope; b_1 average = 115.9134; r: correlation coefficient; r^2: coefficient of determination; r^2_{CV}: cross-validation coefficient; s: standard deviation; f: null hypothesis test (*F*-test).

charge for esters using the same AM1 method. The previous [31,32] and the present calculations were performed using the AM1 method since this gives reliable semi-empirical charges, dipole moments and bond lengths that are more easily obtained for molecules of reasonable size than those obtained employing a minimal base set in *ab initio* calculations [39]. Thus, even though the quantitative accuracy of semi-empirical methods is usually limited, the computational efficiency available nowadays enables the study of a series of molecules in a reasonable amount of time, which it is an important feature for developing quantitative structure-activity relationships (QSAR). Despite the fact that the calculated partial atomic charges may be less reliable than other molecular properties, and that *ab initio* or different semi-empirical methods give values for the net charges with poor numerical agreement, it is important to recognize that their calculation is easy and that the values at least indicate the trends of the charge density distributions in the molecules. Since many chemical reactions or physico-chemical properties are strongly dependent on local electron densities, net atomic charges and other charge-based descriptors are currently used as chemical reactivity indices. For example, net charges on atoms may be considered as non-directional reactivity indices [40].

2.3. Regression analysis

All calculations in the regression analysis were carried out using the Origin [41] and TSAR [42] programs. To test the quality of the regression models obtained, the correlation coefficient (*r*), standard deviation (*s*), coefficient of determination (r^2) and null

hypothesis test (*F*-test) (*f*) were used as statistical parameters. The validity of the model was tested with the cross-validation coefficient (r^2_{CV}) using 'leave-one-out' in the software program Tsar 3.3 for Windows [42]. A group of 10 compounds, not included in the original QSRR models, was employed as an external validation.

3. Results and discussion

The isopropyl isobutyrate molecule represented in the graph below is taken as an example of the I_{SET} calculation using the present approach. All calculations below were performed with six decimal places. The net atomic charges and SET_i values are in Table 3.

Table 3
The net atomic charge (δ_i) and the SET_i values for each carbon and oxygen atom of isopropyl isobutyrate molecule.

Atoms	Isopropyl isobutyrate		
	δ_i	SET_i	$AuSET_i$
O_1	−0.359	1.2605	2.1994
O_2	−0.278	1.1168	1.9487
C_1	+0.305	0.0832	0.1452
C_2	+0.036	0.5601	0.9773
C_3	−0.096	0.7941	–
C_4	−0.210	0.9963	–
C_5	−0.210	0.9963	–
C_6	−0.237	1.0441	–
C_7	−0.215	1.0051	–

3.1. Isopropyl isobutyrate
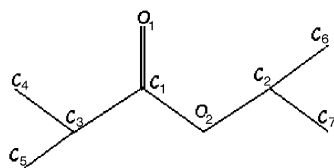
$$\mu_{F1} = d_{C_1-O_1}|Q_{O_1}|$$

$$\mu_{F1} = 1.232948| - 0.358939| = 0.442553$$

$$\mu_{F2} = d_{C_1-O_2}|Q_{O_2}|$$

$$\mu_{F2} = 1.364974| - 0.278490| = 0.380131$$

$$\mu_F = (\mu_{F1}^2 + \mu_{F2}^2 + 2\mu_{F1}\mu_{F2}\cos\theta)^{1/2}$$

$$\mu_F = [(0.442553)^2 + (0.380131)^2 + (2 \times 0.442553^{1/2}$$
$$\times 0.380131 \times \cos 119.023)]$$
$$= 0.420853$$

$$A_\mu = 1 + \log\left(1 + \frac{\mu}{\mu_F}\right) = 1 + \log\left(1 + \frac{1.9180}{0.4208}\right) = 1.7449$$

This molecule is a branched ester with $n = 2$ thus,

$$I_{SET} = \left(\sum I_{SET_i}\right) + sc$$
$$= \left(\sum A_\mu SET_i + \log A_\mu SET_j\right) + n\log SET_{C_{Ac}} + n\log SET_{C_{Al}}$$

$$I_{SET_{O_1}} = A_\mu SET_{O_1} + \log A_\mu SET_{C_1} = 2.1994 + \log 0.1452 = 1.3626$$

$$I_{SET_{O_2}} = A_\mu SET_{O_2} + \log A_\mu SET_{C_1} + \log A_\mu SET_{C_2}$$
$$= 1.9487 + \log 0.1452 + \log 0.9773 = 1.1021$$

$$I_{SET_{C_1}} = A_\mu SET_{C_1} + \log A_\mu SET_{O_1} + \log A_\mu SET_{O_2} + \log SET_{C_3}$$
$$= 0.1452 + \log 2.1994 + \log 1.9487 + \log 0.7941 = 0.6788$$

$$I_{SET_{C_2}} = A_\mu SET_{C_2} + \log A_\mu SET_{O_2} + \log SET_{C_6} + \log SET_{C_7}$$
$$= 0.9773 + \log 1.9487 + \log 1.0441 + \log 1.0051 = 1.2896$$

$$I_{SET_{C_3}} = SET_{C_3} + \log A_\mu SET_{C_1} + \log SET_{C_4} + \log SET_{C_5}$$
$$= 0.7941 + \log 0.1452 + \log 0.9963 + \log 0.9963 = -0.0462$$

$$I_{SET_{C_4}} = SET_{C_4} + \log SET_{C_3} = 0.9963 + \log 0.7941 = 0.8975$$

$$I_{SET_{C_5}} = SET_{C_5} + \log SET_{C_3} = 0.9963 + \log 0.7941 = 0.8975$$

$$I_{SET_{C_6}} = SET_{C_6} + \log A_\mu SET_{C_2} = 1.0441 + \log 0.9773 = 1.0353$$

$$I_{SET_{C_7}} = SET_{C_7} + \log A_\mu SET_{C_2} = 1.0051 + \log 0.9773 = 0.9963$$

$$sc = 2\log SET_{C_2} + 2\log SET_{C_3} = 2\log 0.5601 + 2\log 0.7941$$
$$= -0.7020$$

$$I_{SET} = 1.3626 + 1.1021 + 0.6788 + 1.2896 - 0.0462 + 0.8975 + 0.8975$$
$$+ 1.0353 + 0.9963 - 0.7020 = 7.5115$$

In order to verify the applicability of I_{SET} in predicting the RI on different stationary phases and to investigate the influence of the stationary phase polarity on chromatographic retention, simple linear regression equations between the retention indices (RI_{Exp}) and the semi-empirical electrotopological indices ($RI_{Calc} = b_0 + b_{1_{ave}}I_{SET}$) using 100 esters were established for each stationary phase separately, as shown in Fig. 1 and Table 2. The statistical results for the final QSRR models, in all cases, show that the correlation coefficients (r) are better than 0.9958, and are highest for stationary phases SE-30, OV-7 and DC-710 (between 0.9990 and 0.9988), according to the criteria of Mihalic and Trinajstic [43]
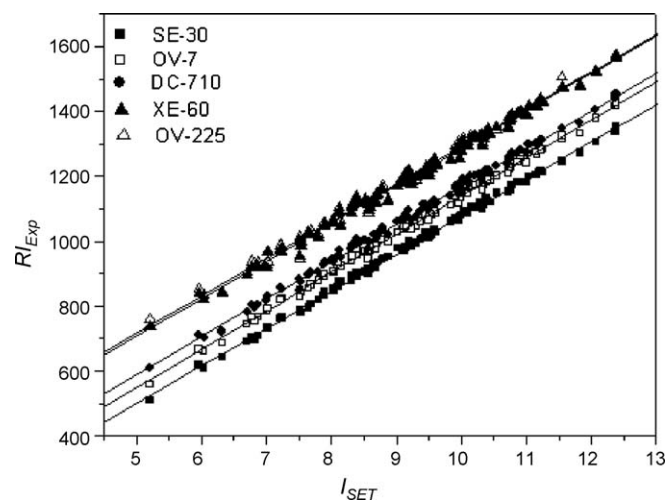
Fig. 1. Simple linear regressions between the retention indices (RI_{Exp}) and the semi-empirical electrotopological indices (I_{SET}) of esters for the stationary phases SE-30, OV-7, DC-710, XE-60 and OV-225.

these statistical parameters indicate that the models are excellent. Clearly, the cross-validation (r_{CV}^2) demonstrates that the final models are statistically significant, with values between 0.9980 and 0.9914, which are very close to the corresponding r^2 values.

The results obtained by single linear regressions (SLR) in this study using the semi-empirical electrotopological descriptor, I_{SET}, are better than those reported by Lu et al. [10]. For SE-30 and OV-7 stationary phases the present results are also better than those found by Liu et al. [11] and for more polar stationary phases the statistical parameters differ only slightly. Both of these studies use multiple linear regression (MLR) between RI and the topological indices for 90 saturated esters on stationary phases with different polarities. Liu et al. [11] applying novel topological indices, polarizability effect index (PEI), odd–even index (OEI) and steric effect index (SV_{ij}), previously developed by them [11], found for each stationary phase correlation coefficients between 0.9989 and 0.9977 and leave-one out cross-validation correlation coefficients between 0.9988 and 0.9975. Through statistical analysis they showed that the polarizability effect, molecular size and branching make dominant contributions to RI of ester compounds, and that the steric effect also provided some contribution. Their models were compared with other recently proposed QSRR models of similar data by Lu et al. [10] in which topological indices such as the Lu index and novel distance-based atom-type DAI topological indices were applied in the MLR method. It was found that the Liu et al. [11] models produce better correlations with Kováts retention indices than the previous topological indices introduced by Lu et al. [10]. Furthermore, the statistical parameters of the present approach have a good agreement with those obtained for saturated esters using the semi-empirical topological index, I_{ET} ($n = 83$; $r = 0.9992$; $s = 8.32$) by SLR previously developed in our group [26].

From Table 3 and Fig. 1, it can be observed that the coefficient b_0 (intercept) increases linearly with increasing polarity of the stationary phases, whereas b_1 (slope) does not show significant variations.

The b_0 value changes according to the extent of interaction between the solute molecules and the molecules of the stationary phase, showing the degree of influence of the stationary phase property on the retention index. Thus, as expected, the intercept value b_0 increases with increasing polarity of the stationary phase. The McReynolds constants were used to evaluate the gas chromatographic polarity.
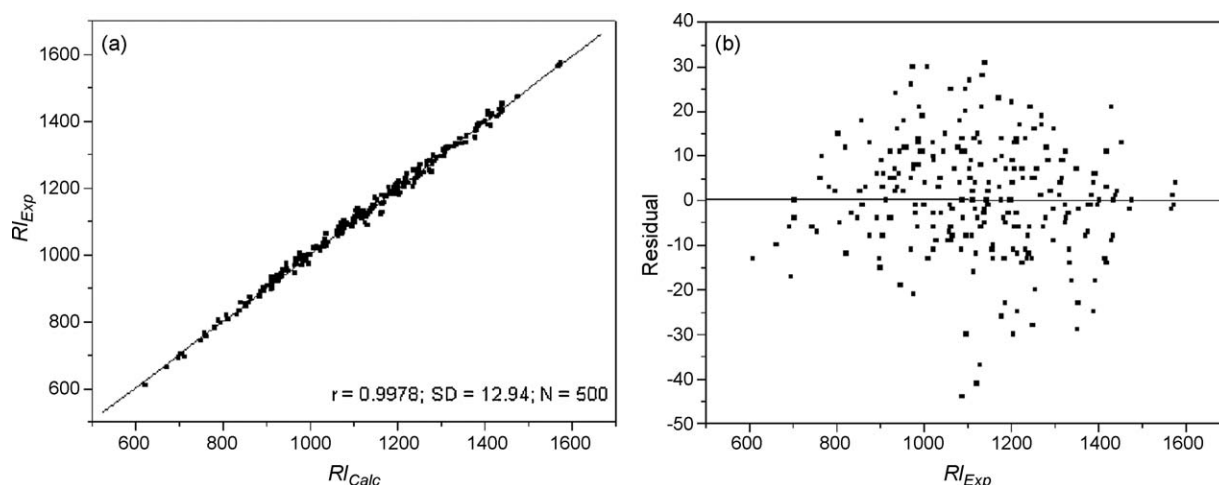
Fig. 2. (a) Experimental (RI_{Exp}) versus calculated retention index (RI_{Calc}) for the 500 data set. (b) Plot of residual versus calculated retention index (RI_{Calc}) of saturated esters for five stationary phases of different polarity.

Table 4
Experimental and predicted retention indices for SE-30 and OV-7 stationary phases for an external group of 10 saturated esters using the Eqs. (5) and (6) (Table 2) and Eq. (11).

Number	Compound	SE-30 Exp.	SE-30 Eq. (5)	ΔRI	SE-30 Eq. (11)	ΔRI	OV-7 7 Exp.	OV-7 Eq. (6)	ΔRI	OV-7 Eq. (11)	ΔRI
1	Heptyl formate	1005	998	7.3	1002	3.3	1070	1059	10.6	1052	17.6
2	Hexyl acetate	988	993	−4.8	997	−8.8	1056	1054	1.6	1047	8.5
3	Butyl propionate	886	889	−2.9	892	−6.1	949	948	1.1	943	6.2
4	Propyl butyrate	875	881	−6.0	884	−9.2	938	940	−1.9	935	3.2
5	Methyl pentanoate	807	798	9.4	800	6.8	871	854	16.6	851	20.2
6	Propyl hexanoate	1064	1066	−2.1	1071	−6.7	1138	1130	8.4	1121	16.6
7	Ethyl octanoate	1165	1162	2.8	1167	−2.5	1234	1228	6.0	1218	15.9
8	Ethyl isopentanoate	839	834	4.5	837	1.7	893	892	0.8	888	5.0
9	Hexyl isohexanoate	1306	1310	−4.3	1317	−10.7	1374	1380	−5.8	1367	6.6
10	Butyl 2-ethylbutyrate	1093	1098	−5.1	1103	−9.9	1150	1162	−12.3	1154	−3.6

The correlation between the b_0 coefficient and McReynolds polarity (Mp) suggests that the solute molecules have different retention behaviors due to the predominance of different molecular mechanisms of interaction between the eluent and the column, with increasing polarity of the column.

The correlation between the McReynolds polarity values (Mp) of stationary phases and the intercept (b_0) of QSRR models obtained using I_{SET} is shown in Eq. (10).

$$b_0 = -108.1958 + 0.1351 Mp$$
$$r = 0.9967; \quad s = 9.10; \quad n = 5 \tag{10}$$

The good linear correlation obtained between the coefficient (b_0) and Mp, as seen through the statistical parameters, indicates similarity between the structures of the phases studied, i.e. SE-30, OV-7, DC-710, XE-60 and OV-225 (polysiloxanes), and similar types of intermolecular interactions contributing to the chromatographic retention of these compounds.

From Eq. (10) and the average angular coefficient, b_1 (Table 2), it is possible to deduce the following combined equation ($RI_{Calc} = b_0 + b_{1ave} I_{SET}$):

$$RI_{Calc} = (-108.1958 + 0.1351 Mp) + 115.9134 I_{SET} \tag{11}$$

Thus, a parameter that represents the property of the stationary phase, Mp, was added to the model, and a model was generated for different stationary phases. A graphical indication of the quality of this combined model (Eq. (11)) can be seen in Fig. 2, through the correlation between the experimental (RI_{Exp}) and calculated (RI_{Calc}) retention indices for the compounds studied.

In order to verify the external validation of the models, two stationary phases were chosen (SE-30 and OV-7). Eqs. (5) and (6) (Table 2) and Eq. (11) were employed to predict the chromatographic retention of 10 compounds not included in the original QSRR models. The experimental and predicted RI values are listed in Table 4.

Theses results (Table 4) indicate that the QSRR models have not only high internal stability, as can be seen observing the values of r^2 and r^2_{CV}, but also good predicting ability for external data.

The excellent results obtained by QSRR using the new semi-empirical electrotopological index are due principally to two facts: (i) that the chromatographic data are very precise and reproducible and (ii) that we are modelling very similar chemical structures. Thus, in the near future the new semi-empirical electrotopological index will be extended to predict others physical–chemical properties and biological activities of different organic compounds.

4. Conclusions

A remodeled SET_i index was developed for esters. This index includes the dipole moment of the whole molecule and an equivalent local dipole moment related to the net charges of the atoms belonging to the functional group, which is dependent on the specific features of the charge distribution in the polar region of the molecule. These electronic and structural effects are demonstrated to participate in the chromatographic retention.

The results suggested that the polar interaction between solute molecules and stationary phases become more important with increasing polarity of the columns.

External validation of the models and the cross-validation, r^2_{CV}, using the leave one-out method, demonstrated that the final models are statistically significant and reliable and, thus, they can be used for the prediction of retention indices for similar compounds in cases where retention values were not readily available.

This new approach to polar molecules allows retention indices with high statistical stability and validity to be obtained through simple linear regression using a descriptor that encodes both charge distribution and structural details, related to the size and branching and the presence of heteroatoms, opening new possibilities to study the QSRR and QSAR of different organic functions.

Acknowledgements

References

[1] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev. 107 (2007) 3212–3246.
[2] K. Héberger, Quantitative structure-(chromatographic) retention relationships, J. Chromatogr. A 1158 (2007) 273–305.
[3] R. Kaliszan, Quantitative Structure-Chromatographic Retention Relationships, Wiley, New York, 1987.
[4] R. Kaliszan, Structure, Retention in Chromatography. A Chemometric Approach, Harwood Academic, Amsterdam, 1997.
[5] B.S. Junkes, R.D.M.C. Amboni, R.A. Yunes, V.E.F. Heinzen, Prediction of chromatographic retention of saturated alcohols on stationary phases of different polarity applying the novel semi-empirical topological index, Anal. Chim. Acta 477 (2003) 29–39.
[6] B.S. Junkes, R.D.M.C. Amboni, R.A. Yunes, V.E.F. Heinzen, Application of novel semi-empirical topological index in the QSRR of aliphatic ketones and aldehydes on stationary phases of different polarity, J. Braz. Chem. Soc. 15 (2004) 183–189.
[7] T. Körtvélyesi, M. Görgényi, K. Héberger, Correlation between retention indices and quantum-chemical descriptors of ketones and aldehydes on stationary phases of different polarity, Anal. Chim. Acta 428 (2001) 73–82.
[8] Y. Wang, A. Li, H. Liu, Q. Zhang, W. Ma, W. Song, G. Jiang, Development of quantitative structure gas chromatographic relative retention time models on seven stationary phases for 209 polybrominated diphenyl ether congeners, J. Chromatogr. A 1103 (2006) 314–328.
[9] B. Ren, Atom-type-based AI topological descriptors for quantitative structure-retention index correlations of aldehydes and ketones, Chemometr. Intell. Lab. Syst. 66 (2003) 29–39.
[10] C. Lu, W. Guo, C. Yin, Quantitative structure-retention relationship study of gas chromatographic retention indices of saturated esters on different stationary phases using novel topological indices, Anal. Chim. Acta 561 (2006) 96–102.
[11] F. Liu, Y. Liang, C. Cao, N. Zhou, QSPR study of GC retention indices for saturated esters on seven stationary phases based on novel topological indices, Talanta 72 (2007) 1307–1315.
[12] A.R. Katritzky, K. Chen, U. Maran, D.A. Carlson, QSRR correlation and predictions of GC retention indexes for methyl-branched hydrocarbons produced by insects, Anal. Chem. 2 (2007) 101–109.
[13] M. Randic, The connectivity index 25 years after, J. Mol. Graph. Model. 20 (2001) 19–35.
[14] M. Randic, J. Zupan, On Interpretation of well-known topological indices, J. Chem. Inf. Comp. Sci. 41 (2001) 550–560.
[15] M. Randic, Wiener-Hosoya Index—a novel graph theoretical molecular descriptor, J. Chem. Inf. Comp. Sci. 44 (2004) 373–377.
[16] M. Randic, On history of the Randic index and emerging hostility toward chemical graph theory, Match-Commun. Math. Comp. Chem. 59 (2008) 5–124.
[17] P. Tulasamma, K.S. Reddy, Quantitative structure and retention relationships for gas chromatographic data: application to alkyl pyridines on apolar and polar phases, J. Mol. Graph. Model. 25 (2006) 507–513.
[18] M. Shamsipur, R. Ghavami, H. Sharghi, B. Hemmateenejad, Highly correlating distance/connectivity-based topological indices 5. Accurate prediction of liquid density of organic molecules using PCR and PC-ANN, J. Mol. Graph. Model. 27 (2008) 506–511.
[19] J. Acevedo-Martínez, J.C. Escalona-Arranz, A. Villar-Rojas, F. Téllez-Palmero, R. Pérez-Rosés, L. González, R. Carrasco-Velar, Quantitative study of structure-retention index relationship in the imine family, J. Chromatogr. A 1102 (2006) 238–244.
[20] P. Tulasamma, K.S. Reddy, Quantitative structure and retention relationship for gas chromatographic data: application to alkyl pyridines on apolar and polar phases, J. Mol. Graphic Modell. 25 (2006) 507–513.
[21] L.B. Kier, L.H. Hall, Molecular Connectivity and Structure-Activity Analysis, Wiley, New York, 1986.
[22] O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, A.T. Balaban, Optimum structural descriptors derived from the Ivanciuc-Balaban Operator Internet Electron, J. Mol. Des. 1 (2002) 319–331.
[23] V.E.F. Heinzen, M.F. Soares, R.A. Yunes, Semi-empirical topological method for prediction of the chromatographic retention of cis- and trans-alkene isomers and alkanes, J. Chromatogr. A 849 (1999) 495–506.
[24] B.S. Junkes, R.D.M.C. Amboni, R.A. Yunes, V.E.F. Heinzen, Use of a semiempirical topological method to predict the chromatographic retention of branched alkenes, Chromatographia 55 (2002) 75–81.
[25] B.S. Junkes, R.D.M.C. Amboni, R.A. Yunes, V.E.F. Heinzen, Quantitative structure-retention relationships (QSRR), using the optimum semiempirical topological index, for methyl-branched alkanes produced by insects, Chromatographia 55 (2002) 707–715.
[26] R.D.M.C. Amboni, B.S. Junkes, R.A. Yunes, V.E.F. Heinzen, Semi-empirical topological method for prediction of the chromatographic retention of esters, J. Mol. Struct. (Theochem.) 579 (2002) 53–62.
[27] R.D.M.C. Amboni, B.S. Junkes, R.A. Yunes, V.E.F. Heinzen, Quantitative structure-property relationship study of chromatographic retention indices and normal boiling points for oxo compounds using the semi-empirical topological method, J. Mol. Struct. (Theochem) 586 (2002) 71–80.
[28] B.S. Junkes, A.C.S. Arruda, R.A. Yunes, L.C. Porto, V.E.F. Heinzen, Semiempirical topological index: application to QSPR/QSAR modeling, J. Mol. Model. 11 (2005) 128–134.
[29] L.C. Porto, E.S. Souza, B.S. Junkes, R.A. Yunes, V.E.F. Heinzen, Semi-Empirical Topological Index: development of QSPR/QSRR and optimization for alkylbenzenes, Talanta (Oxford) 76 (2008) 407–412.
[30] A.C.S. Arruda, E.S. Souza, B.S. Junkes, R.A. Yunes, V.E.F. Heinzen, Semi-empirical topological index to predict properties of halogenated aliphatic compounds, J. Chemom. 22 (2008) 186–194.
[31] E.S. Souza, B.S. Junkes, C.A. Kuhnen, R.A. Yunes, V.E.F. Heinzen, On a new semi-empirical electrotopological index for QSRR models, J. Chemom. 22 (2008) 378–384.
[32] E.S. Souza, B.S. Junkes, C.A. Kuhnen, R.A. Yunes, V.E.F. Heinzen, Modeling the semi-empirical electrotopological index in QSPR studies for aldehydes and ketones. J. Chemom. (in press).
[33] M. Karelson, V.S. Lobanov, A.R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR, Chem. Rev. 96 (1996) 1027–1043.
[34] Sadtler Standard GC Retention Index Library, Sadtler Research Laboratories, Philadelphia, PA, 1985, vols. 1–3.
[35] R.D.M.C. Amboni, B.S. Junkes, R.A. Yunes, V.E.F. Heinzen, Quantitative structure-odor relationships of aliphatic esters using topological indices, J. Agric. Food Chem. 48 (2000) 3517–3521.
[36] HyperChem for Windows, Release 7.01, 2002, Hypercube, Georgia, USA.
[37] W.B. Smith, Introduction to Theoretical Organic Chemistry and Molecular Modeling, VCW, New York, 1996, pp. 142–143.
[38] R.S. Mulliken, Electronic population analysis on LCAO-MO molecular ware functions, J. Chem. Phys. 23 (1955) 1833–1840.
[39] T. Bredow, K. Jug, Theory and range of modern semiempirical molecular orbital methods, Theor. Chem. Acc. 113 (2005) 1–14.
[40] O. Kikuchi, Systematic QSAR procedures with quantum chemical descriptors, Quant. Struct. Act. Relat. 6 (1987) 179–184.
[41] MicroCal Origin, version 5.0, 1997, Northampton, MA, USA.
[42] Tsar TM 3.3 for Windows, Oxford Molecular Ltd., 2000, Hunt Valle, MD, USA.
[43] M. Mihalic, N. Trinajstic, A graph-theoretical approach to structure-property relationships, J. Chem. Educ. 69 (1992) 701–712.