Journal of Chromatography, 499 (1990) 333–344
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 706

MECHANISM OF RETENTION IN HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY ON POROUS GRAPHITIC CARBON AS REVEALED BY PRINCIPAL COMPONENT ANALYSIS OF STRUCTURAL DESCRIPTORS OF SOLUTES

ROMAN KALISZAN*
Department of Biopharmaceutics and Pharmacodynamics, Medical Academy, K. Marksa 107, 80-416 Gdansk (Poland)
KRZYSZTOF OŚMIAŁOWSKI
Department of Inorganic Chemistry, Medical Academy, K. Marksa 107, 80-416 Gdansk (Poland)
and
BARBARA J. BASSLER and RICHARD A. HARTWICK
Department of Chemistry, Rutgers University, Piscataway, NJ 08854 (U.S.A.)

SUMMARY

High-performance liquid chromatographic retention data derived for a set of non-congeneric aromatic solutes on a graphitic carbon stationary phase with hexane as the eluent were analysed. The capacity factors were quantitatively related to structural information extracted from nineteen molecular descriptors of solutes by multivariate analysis. The descriptors considered included molecular weight, molecular refractivity, topological indices, information content indices, quantum chemical indices and structural indices based on electron charge distribution within a molecule. The first three principal components obtained by factor analysis of nineteen structural descriptors appeared meaningful for the description of retention. Most important for retention were structural features reflecting abilities of solutes to participate in intermolecular interactions of the electron pair donor–acceptor and dipole–induced dipole type. The structural factor related to molecular size appeared less important for retention but decisive for the prediction of boiling points. Multivariate analysis of structural descriptors was demonstrated to provide systematic information useful for the explanation of the mechanism of a chromatographic process.

INTRODUCTION

Quantitative structure–retention relationships (QSRR) have been extensively studied in the last decade for three main purposes: explanation of the mechanism of retention, retention prediction and parameterization of the structures of solutes[1]. Most published QSRR were derived by means of a multiple regression analysis of a set of retention data (dependent variable) and various empirical, semiempirical and theoretical structural parameters assumed to be independent variables. Unfortunately, the independent (*i.e.*, explanatory) variables considered have often been interrelat-

0021-9673/90/$03.50 © 1990 Elsevier Science Publishers B.V.

ed by simple or multiple correlations. Sets of intercorrelated chemical data are unsuitable for multiple regression but can be the subject of multivariate analysis with factorial methods.

Factor analysis was first applied to chromatographic data in the early 1970s[2–4]. More recent examples of the application of multivariate analysis in chromatography may be found in refs. 5–15. The analyses reported were mostly aimed at the prediction of the retention of a particular set of solutes with various mobile–stationary phase systems. However, Wold and co-workers[14,15] used chromatographic data to extract some structural properties of amino acids. Little attention has been paid to the multivariate analysis of structural descriptors of solutes. In fact, Massart and co-workers[6,7] performed factor analysis of a set of topological and quantum chemical indices of a series of compounds but dit not relate the factors extracted quantitatively to the gas chromatographic retention indices studied.

It appeared soon after Knox *et al.*[16] introduced porous graphitic carbon as a high-performance liquid chromatographic (HPLC) stationary phase that the retention behaviour observed was different to that expected for a reversed-phase material. Our QSRR studies[17] demonstrated that specific, polar, size-independent solute–stationary phase interactions are decisive for retention. In this work we attempted to apply multivariate analysis in order to obtain some insight into the mechanism of HPLC retention on graphitic carbon.

EXPERIMENTAL

Materials

The test solutes selected for study are simple aromatic derivatives. The whole set of 20 solutes is diverse enough to avoid congenericity and includes various functionalities. On the other hand, a subgroup of phenol derivatives can be separated to allow specific comparisons.

For the sake of comparison, boiling point data were collected[18] for the solutes analysed and these data are given in Table I together with chromatographic and structural indices.

Chromatographic determinations

Solutes dissolved in heptane were chromatographed with neat heptane as the mobile phase on 30 cm × 1 mm I.D. columns slurry-packed with porous graphitic carbon material. A standard HPLC system was applied. Details of the analytical procedure have been given previously[17].

The column void volume was determined by the change in the refractive index of an injection of pentane in hexane. Logarithms of capacity factors (log k') are given in Table I.

Structural descriptors

Non-empirical structural descriptors were considered, *i.e.*, molecular parameters that can be calculated based exclusively on the structural formula of a solute. The total of nineteen descriptors determined included molecular weight (molwt), molecular refractivity (bondrefr), four topological indices, three information content indices, six quantum chemical indices and four modifications of quantum chemical data.

Molecular refractivity (bondrefr) was calculated as the sum of the bond refractivities for all pairs of connected atoms according to Iophphe[19]. A list of individual increments can be found in ref. 1, p. 96.

The following indices based on molecular topology were considered: (i) indices of molecular shape (kappa values) of second and third order, κ^2 and κ^3 (kappa2 and kappa3), according to Kier[20]; (ii) the Kier and Hall[21] generalized molecular connectivity index of second order, χ^{2v} (chi2v); and (iii) the average distance sum connectivity index of Balaban[22] calculated according to Barysz *et al.*[23] (balaban). It should be mentioned that being in full agreement with mathematical topology, one can compare topological indices derived for closely congeneric compounds only, such as aliphatic hydrocarbons. The indices considered here are empirical modifications of graph theoretical indices and were introduced to provide the diversity of structures observed in real chemical systems.

The next group of structural indices considered were information indices of neighbourhood symmetry of zeroth order, ic0, first order, ic1, and second order, ic2 (ref. 24). These information content indices were calculated by means of the general information theory equation (Shannon's equation) from probabilities of finding equivalent atoms (ic0) or patterns of atoms in a given structural formula.

The separate group of non-empirical structural descriptors studied were quantum chemical indices. The indices were calculated for each solute by using the MNDO3 method[25] on molecular coordinate files obtained from the Cambridge Crystallographic Database[26]. Calculations were done on a VAX 780 computer with an array processor. The following quantum chemical indices were considered: total energy (etotal), heat of formation (heatf), energy of core–core repulsion (corer), energy of highest occupied molecular orbital (ehomo), energy of lowest unoccupied molecular orbital (elumo), dipole moment (dipolem) and eelctron excess charges on individual atoms.

The last group of molecular descriptors considered here were parameters derived from excess charge distribution within a molecule. These were submolecular polarity parameters, Δ (delta), electronic–geometric index, T^E (geoeli), and the local dipole index, D_L (diploc). The submolecular polarity parameter delta1 was calculated as the largest difference of excess charges for a pair of atoms in a molecule[27]. The second-order submolecular polarity parameter, delta2, was determined analogously as the second largest difference of excess charges. Determination of the local dipole index according to Kikuchi[28] is based on the equation

$$D_L = \Sigma|Q_A - Q_B|/N_{AB}$$

where Q_A and Q_B are net charges of atoms A and B which are bonded together. The summation is made over all pairs of bonded atoms and N_{AB} is the number of such pairs. The electronic–geometric index, T^E (geoeli), is based on excess charge distribution and on molecular geometry, as proposed previously[29].

Statistical analysis

Principal component analysis (PCA)[30,31] was performed on a set of nineteen structural descriptors of twenty solutes. The resulting principal component object scores were related to chromatographic data and boiling points by means of a stepwise multiple regression.

By PCA, a multi-dimensional parameter space may be reduced to a few significant components which concentrate the information distributed over many variables. In other words, all those original parameters which are interrelated by simple or multiple correlations are combined (linear combination) to a few orthogonal principal components.

Principal component loadings (vectors) and principal component object scores are usually displayed graphically on the plane spanned by two principal component axes. A vector representation of the original parameters in the principal component space is obtained by plotting the loading values of the principal components on the variable against each other.

In PCA, high loadings of almost all variables on the first principal component are usually observed, whereas the other PCs contain much less information. In such a situation it may be difficult to interpret the meaning of individual principal components and rotation of the principal coordinate system is often applied. The VARIMAX procedure was applied here using standard statistical software.

RESULTS AND DISCUSSION

For the set of twenty solutes studied, logarithms of HPLC capacity factors,

TABLE I

RETENTION DATA (log k'), BOILING POINTS AND NINETEEN NON-EMPIRICAL STRUCTURAL DESCRIPTORS FOR THE SET OF SOLUTES ANALYSED

No.	*Solutes*	*log k'*	*b.p. (K)*	*molwt*[a]	*bondrefr*[b]	*kappa2*[c]	*kappa3*[d]	*chi2v*[e]	*balaban*[f]	*ic0*[g]	*ic1*[h]
1	Phenol	0.431	454.7	94.11	27.768	1.755	1.016	1.336	8.640	1.314	1.914
2	*p*-Cresol	0.456	474.9	108.14	32.416	1.968	1.331	1.836	9.717	1.272	2.272
3	4-Ethylphenol	0.277	492	122.17	37.064	2.646	1.541	2.020	10.600	1.236	2.325
4	4-*n*-Propylphenol	0.312	505.6	136.19	41.712	2.194	1.229	2.416	11.365	1.207	2.275
5	2-Isopropylphenol	−0.171	486	136.19	41.712	2.843	1.834	2.517	11.992	1.207	2.275
6	2-*tert.*-Butylphenol	−0.419	494	150.22	46.360	2.666	1.808	3.767	13.351	1.183	2.209
7	4-*tert.*-Butylphenol	−0.022	512.5	150.22	46.360	2.666	2.054	3.797	13.064	1.183	2.209
8	Carvacrol	−0.194	510.7	150.22	46.360	3.055	1.808	3.001	13.192	1.183	2.259
9	Thymol	−0.148	506	150.22	46.360	3.055	1.808	3.217	13.262	1.183	2.259
10	Benzene	−1.060	353.2	78.11	26.184	1.606	0.845	1.155	7.330	1.000	1.194
11	Toluene	−0.791	383.6	92.14	30.832	1.783	1.038	1.411	8.435	0.997	1.533
12	*o*-Xylene	−0.433	417.4	106.17	35.480	1.994	1.101	2.084	9.661	0.991	1.658
13	*m*-Xylene	−0.666	412.1	106.17	35.480	1.994	1.353	2.158	9.596	0.991	1.658
14	*p*-Xylene	−0.658	411.3	106.17	35.480	1.994	1.353	2.155	9.542	0.991	1.658
15	Anisole	−0.317	428	108.14	32.616	2.468	1.331	1.517	9.711	1.272	1.774
16	Benzyl alcohol	−0.211	478.3	108.14	32.416	2.478	1.331	1.644	9.420	1.272	2.046
17	Chlorobenzene	−0.606	405	112.56	31.018	1.987	1.204	1.296	8.928	1.325	1.650
18	Acetophenone	0.283	475.6	120.15	35.618	2.444	1.389	1.922	10.770	1.264	1.993
19	Nitrobenzene	0.799	483.8	123.11	31.858	2.259	1.253	1.654	11.376	1.727	2.006
20	Methyl benzoate	0.236	472.6	136.15	37.232	3.129	1.662	1.858	12.034	1.392	2.192

[a] Molecular weight. [b] Sum of bond refractivities[20]. [c] Index of molecular shape of second order[19]. [d] Index of molecular shape of third order[19]. [e] Valence connectivity index of second order[21]. [f] Modified Balaban index[23]. [g,h,i] Information content indices of zeroth, first and second order, respectively[24]. [j] Total energy (eV). [k] Heat of formation (kcal). [l] Energy of core–core repulsion (eV). [m] Energy of the highest occupied molecular orbital (eV). [n] Energy of the lowest empty molecular orbital (eV). [o] Dipole moment calculated quantum chemically (D). [p] Submolecular polarity parameter of first order (electrons). [q] Submolecular polarity parameter of second order (electrons). [r] Local dipole index[28] (electrons). [s] Geometric electronic index[29] (electrons/Å^2).

boiling points and nineteen non-empirical structural descriptors are collected in Table I. The correlation matrix (not included) for all 21 data sets reflects high intercorrelations among most of the structural descriptors under study.

Stepwise multiple regression analysis of log k' data against the nineteen non-empirical structural descriptors of the solutes did not yield any statistically significant multi-parameter equation.

Intercorrelations among structural parameters limit the applicability of multiple regression analysis in QSRR studies. On the other hand, according to current chemometric theory, as many relevant data as possible should be used because this increases the probability of good characterization of the solutes[15]. The large data tables resulting from such an assumption can be subjected to multivariate analysis to extract the systematic information contained in the data. The set of structural data considered here consisted of the last nineteen columns in Table I.

Basing on the eigenvalues of the correlation matrix, we decided to extract five main factors from the data. The first factor accounted for 57% of the variance in structural data, the second for 21%, the third for 7%, the fourth for 6% and the fifth for 3% of the variance. Altogether, the principal components extracted accounted for about 95% of the variance in the structural data considered.

Subsequently, we calculated loadings (eigenvectors) of the five principal components extracted. The loadings were next subjected to VARIMAX rotation. The VARIMAX-rotated PC loadings, PCvr1 to PCvr5, are given in Table II. In Fig. 1 the loadings of two first principal components by individual structural descriptors are graphically depicted.

ic2[i]	*etotal*[j]	*heatf*[k]	*corer*[l]	*ehomo*[m]	*elumo*[n]	*dipolem*[o]	*delta1*[p]	*delta2*[q]	*diploc*[r]	*geoeli*[s]
2.661	1172.37	12.443	3336.19	−9.3347	−0.0245	1.55	0.444	0.347	0.1411	2.436
3.203	1328.68	10.938	4272.77	−9.2302	0.1309	1.52	0.441	0.361	0.1578	3.001
3.471	1484.60	18.407	5320.72	−9.0901	0.1767	1.36	0.435	0.342	0.1251	3.102
3.698	1639.92	39.683	6360.13	−9.1356	−0.0370	1.68	0.450	0.338	0.1045	3.080
3.482	1639.88	40.626	6723.87	−9.4610	0.0874	1.66	0.453	0.346	0.1056	3.472
3.152	1794.76	71.972	8189.62	−9.5373	0.0712	1.56	0.462	0.350	0.1028	4.217
2.992	1795.46	55.862	7919.10	−9.1947	0.1701	1.50	0.440	0.347	0.1236	4.219
3.593	1795.88	53.090	7828.34	−9.2813	0.0302	1.46	0.454	0.372	0.1133	4.091
3.593	1795.77	48.596	7920.80	−9.3971	0.0472	1.66	0.453	0.368	0.1038	4.000
2.661	848.65	249.257	2255.72	−7.4078	−0.4751	0.13	0.168	0.164	0.7960	1.506
2.386	1006.14	59.026	3319.61	−9.2088	0.1824	0.03	0.204	0.169	0.0851	1.727
2.419	1157.34	175.358	4451.23	−9.3617	0.2491	0.14	0.201	0.200	0.0796	2.318
2.530	1157.88	160.664	4343.00	−9.0636	0.0498	0.16	0.231	0.223	0.9170	2.440
2.197	1158.56	147.252	4298.57	−9.3161	0.2503	0.02	0.218	0.217	0.9440	2.309
2.858	1327.77	31.882	4321.22	−8.8705	0.2100	1.35	0.501	0.388	0.1901	3.540
3.031	1328.13	23.574	4293.52	−9.3306	0.2758	1.46	0.531	0.494	0.1655	3.488
2.459	1190.88	41.735	3192.70	−9.6014	−0.1082	1.43	0.126	0.124	0.5100	1.008
3.013	1455.63	33.040	5002.93	−9.8255	−0.3894	2.66	0.550	0.416	0.1327	2.936
2.700	1683.97	71.757	5139.94	−10.2590	−1.0730	4.88	0.842	0.819	0.2428	3.666
3.155	1777.78	−6.123	6161.55	−9.7408	−0.3392	1.76	0.764	0.716	0.2399	5.150

TABLE II

VARIMAX-ROTATED LOADINGS OF THE EXTRACTED PRINCIPAL COMPONENTS BY INDIVIDUAL STRUCTURAL DESCRIPTORS DENOTED AS IN TABLE I

Structural descriptor	*PCvr 1*	*PCvr 2*	*PCvr 3*	*PCvr 4*	*PCvr 5*
molwt	0.9411	0.1939	0.2309	0.0404	0.0422
bondrefr	0.9812	−0.1106	0.1058	0.0650	−0.0114
kappa2	0.7672	0.1657	0.2620	0.1219	0.4263
kappa3	0.9148	−0.0014	0.1633	0.0234	0.2406
chi2v	0.9570	−0.1417	−0.0115	0.0224	−0.0605
balaban	0.9476	0.2351	0.1739	0.0377	0.0993
ic0	−0.0039	0.8805	0.3678	−0.0455	0.1296
ic1	0.5796	0.2746	0.3280	0.6257	0.1320
etotal	0.8451	0.3942	0.2930	0.0901	0.1687
heatf	−0.1619	−0.2388	−0.8694	0.0240	−0.2526
corer	0.9688	0.0949	0.1654	0.0971	0.0826
ehomo	−0.3170	−0.4787	−0.4550	0.6322	−0.0343
elumo	0.1390	−0.8985	0.3036	−0.0758	0.1436
dipolem	0.1863	0.9190	0.2916	−0.0120	−0.0287
delta1	0.2626	0.7589	0.3085	0.0940	0.4708
delta2	0.1773	0.8080	0.1819	0.0074	0.5053
diploc	−0.2855	−0.1426	−0.8155	−0.1542	0.0290
geoeli	0.6868	0.3201	0.2065	0.1365	0.5749

The principal component scores for the set of twenty solutes studied, scaled to a variance of 1, obtained after VARIMAX rotation, are collected in Table III. For illustration the first two PC scores for the solutes considered are presented in Fig. 2.

As can be observed in Table II (see also Fig. 1), the first principal component, PCvr1, is loaded mostly by such structural descriptors as molecular refractivity (bondrefr), energy of core–core repulsion (corer), valence connectivity index of second order (chi2v), Balaban index (balaban), molecular weight (molwt), kappa parameter of third order (kappa3) and total energy (etotal). To a lesser extent PCvr1 is loaded by kappa2, electronic–geometric index (geoeli) and information content indices of first and second order (ici and ic2). It is evident that all these structural descriptors strongly loading PCvr1 reflect basically the size (bulkiness) of the solutes. In such a situation PCvr1 condenses information on molecular size.

The second principal component, PCvr2, is loaded predominantly by the following structural descriptors: dipole moment (dipolem), energy of LUMO (elumo), information content index of zeroth order (ic0) and the submolecular polarity parameters of the first and second order (delta1 and delta2). Hence it may be concluded that PCvr2 concentrates structural information related to the so-called molecular polarity. Specific, polar properties of chemical compounds are the result of electron distribution within a molecule. From the chromatographic point of view, such properties determine the ability of a solute to participate in intermolecular interactions with the stationary and/or mobile phase of the dipole–dipole, dipole–induced dipole and electron pair donor–acceptor type.

A separate discussion requires high loading of PCvr2 by ic0. The index ic0 reflects the diversity in the atom composition of a molecule. For the set of solutes

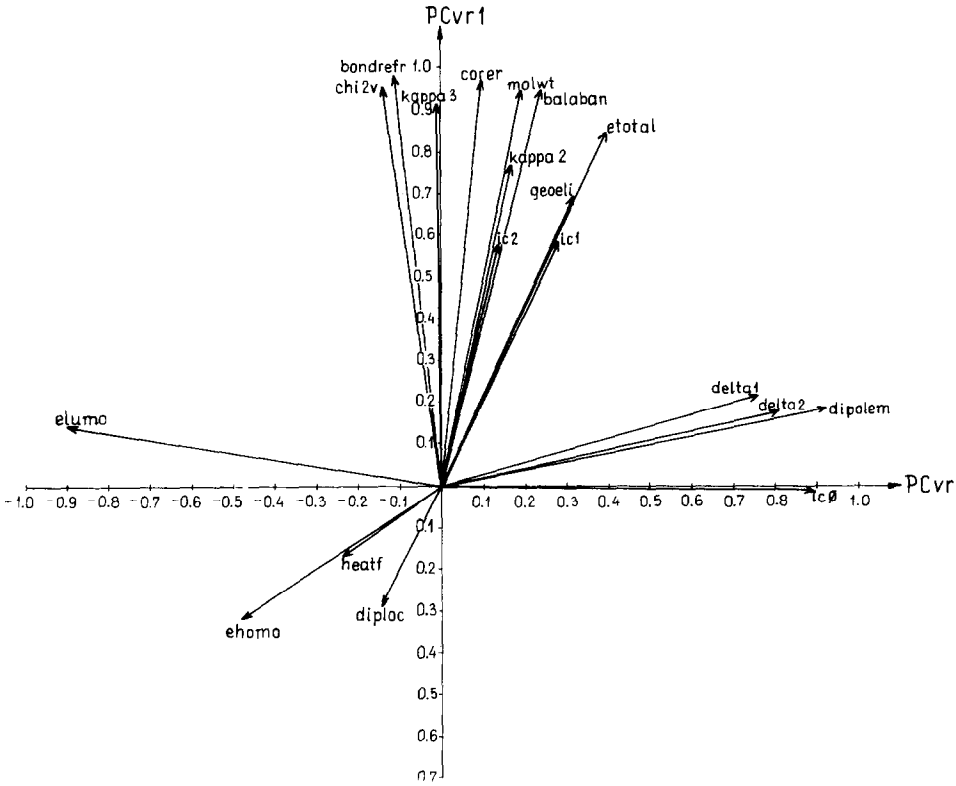

Fig. 1. VARIMAX-rotated laodings of two first principal components by individual structural descriptors denoted as in Table I.

studied it also reflects their molecular polarity as demonstrated by correlation of ic0 with dipole moment (r = 0.93) and with the submolecular polarities delta1 and delta2 (r = 0.81 and 0.82, respectively).

The third VARIMAX-rotated principal component, PCvr3, is highly loaded by the heat of formation (heatf) and the local dipole index (diploc). The next loading structural descriptor would be the energy of HOMO (ehomo), with a contribution of −0.4550. The parameter ehomo is one of the three structural descriptors most significantly loading PCvr4, the other two being ic1 and ic2. The loading values for these descriptors are 0.6322, 0.6257 and 0.6377, respectively. Loadings of PCvr4 by the remaining parameters are insignificant (<0.15). There is no parameter strongly loading PCvr5 but several descriptors provide contributions of about 0.5; geoeli, delta1, delta2 and kappa2. It may be speculated that PCvr5 contains some information on the molecular shape of the solutes.

The principal component scores after VARIMAX rotation for the solutes studied (Table III and Fig. 2) differentiate a subgroup of phenolic solutes from the remaining benzene derivatives. As illustrated in Fig. 2, phenols are distributed along the PCvr1 axis and are characterized by low values of PCvr2. The reverse situation is observed for non-phenolic compounds.

It seemed interesting to apply the structural information condensed in five prin-

TABLE III

VARIMAX-ROTATED PRINCIPAL COMPONENT SCORES FOR THE SOLUTES ANALYSED

The solutes are numbered as in Table I.

Solute No.	*PCvr 1*	*PCvr 2*	*PCvr 3*	*PCvr 4*	*PCvr 5*
1	1.4096	−0.1692	1.0508	−0.1146	−0.1675
2	0.7181	0.0879	1.0513	0.8746	−0.0358
3	0.0823	0.3454	0.9058	1.1894	0.2208
4	−0.3910	−0.1595	0.6249	1.3567	−1.2177
5	−0.8082	0.1659	0.5300	0.4910	−0.0839
6	−1.7161	0.1006	−0.1624	−0.2756	−0.4645
7	−1.7017	0.3733	−0.2055	−0.2137	−0.0352
8	−1.4317	0.1615	0.0534	0.6297	0.0020
9	−1.5090	0.1269	0.1492	0.4699	−0.2106
10	1.4856	0.0749	−2.6331	2.5015	−0.5728
11	1.1533	1.1636	0.7842	−0.9680	−0.5168
12	0.2404	0.9928	−0.1926	−0.9950	−0.7277
13	0.0362	0.6940	−1.8420	−0.8304	0.3305
14	0.0419	0.9590	−1.6122	−1.5784	0.5789
15	0.7642	0.3651	0.4583	−0.0071	1.5090
16	0.7810	0.2429	0.8289	0.1247	1.5695
17	0.7581	−0.0330	0.4278	−1.3157	−1.8248
18	0.1020	−0.9377	0.4288	−0.3581	−0.6119
19	0.2569	−3.5829	−0.6015	−0.8545	−0.3883
20	−0.2719	−0.9712	−0.0440	−0.1264	2.6470

cipal components to correlation studies with HPLC data derived on graphitic carbon as stationary phase with hexane as the eluent. The stepwise regression analysis of log k' against PCvr1 to PCvr5 for the whole set of 20 solutes yielded the following equation:

$$\log k' = -0.145 - 0.323\ \text{PCvr2} + 0.259\ \text{PCvr3} \quad (1)$$

characterized by a correlation coefficient $R = 0.85$, standard error of estimate $s = 0.271$ and F-test value 22.2. The equation is significant at the $1.8 \cdot 10^{-5}$ level and the significance levels for the coefficients at the variables PCvr2 and PCvr3 are 0.0001 and 0.0005, respectively.

On analysing eqn. 1, one notes the absence of the molecular size-related PCvr1. As is well known, the bulkiness of chemical compounds determines their ability to participate in non-specific dispersive intermolecular interactions. The lack of significance of PCvr1 for a description of the retention of the whole set of diverse solutes means that the net London-type attraction of solutes and stationary phase is comparable to that between solutes and mobile phase molecules. The PCvr2 term in eqn. 1 can be interpreted as compromising the charge-transfer attractive interactions of the solutes and the graphitic carbon stationary phase on the one hand and the solute–hexane inductive interactions on the other. The contribution of PCvr3 to eqn. 1 is difficult to interpret in terms of the retention mechanism. Bearing in mind the input by the local dipole index to the third PC, one could consider PCvr3 as reflecting

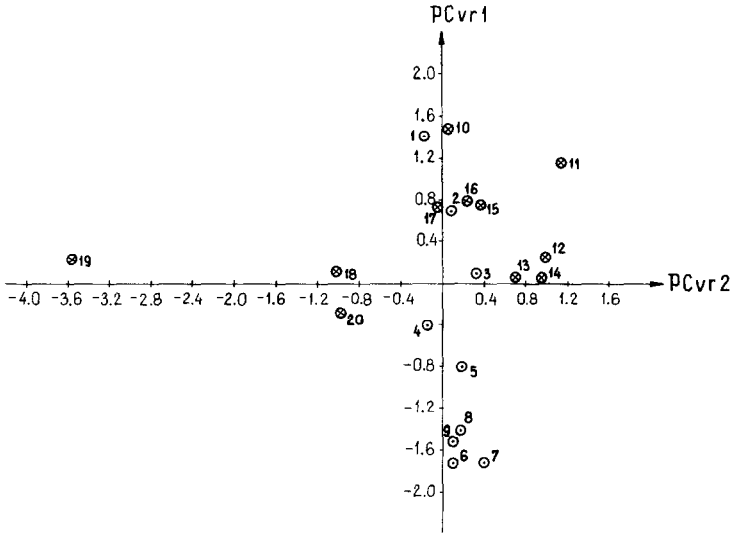

Fig. 2. Principal component scores after VARIMAX rotation for the solutes numbered as in Table I.

dipole–induced dipole interactions involving solute molecules and the molecules of both chromatographic phases.

Eqn. 1 may give some insight into the mechanism of separations in the graphitic carbon–hexane HPLC system. Its statistical value, however, is too low for the reliable prediction of retention. When phenolic derivatives are excluded from regression then the equation derived for the remaining eleven benzene derivatives has much increased predictive properties:

$$\log k' = -0.100 - 0.373\ \text{PCvr1} - 0.299\ \text{PCvr2} + 0.159\ \text{PCvr3} \quad (2)$$

Eqn. 2 is characterized by the statistical parameters $R = 0.977$, $s = 0.140$ and F-value $= 49.5$. The significance level for eqn 2 is $4.4 \cdot 10^{-5}$ and for the variables PCvr1, PCvr2 and PCvr3 the respective values are 0.001, 0.00001 and 0.002. The relationship between the measured log k' data and values calculated by eqn. 2 is illustrated in Fig. 3.

On comparing eqn. 1 and 2, one notes that in the latter the term PCvr1, related to molecular size, becomes significant. It may also be noted that PCvr1 scores for the solutes studied (Table III) attain higher values for smaller compounds. Thus, a negative coefficient for PCvr1 suggests that more bulky non-phenolic solutes will be more strongly retained in the HPLC system under study. This in turn can be interpreted as a result of weaker dispersive interactions between a solute and the solvent than the respective solute–stationary phase interactions. The terms PCvr2 and PCvr3 in eqn. 2 may be interpreted analogously as in eqn. 1 for a whole set of solutes. Again, the term PCvr2 predominates over others.

The following regression equation was derived for a subset of nine phenolic derivatives:

$$\log k' = 0.204 + 0.247\ \text{PCvr1} \quad (3)$$

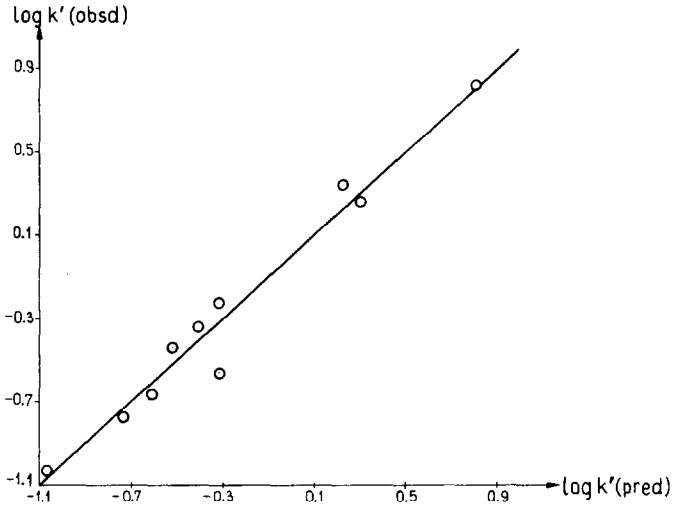

Fig. 3. Plot of observed values against values calculated from eqn. 2 of log k' for non-phenolic solutes.

For eqn. 3 the statistical characteristics are $R = 0.884$, $s = 0.158$ and F-value $= 25.0$. The significance levels for eqn. 3 and the variable PCvr1 are 0.002 and 0.001, respectively. As could be expected, similarity of the electronic properties of a closely congeneric group of phenols made the terms PCvr2 and PCvr3 insignificant for the retention differentiation. In such a situation, retention within a series depends on the ability of an individual member to participate in non-specific dispersive interactions. This does not mean that there are no polar intermolecular interactions present in the system. These interactions, as revealed by the general eqn. 1, may still be decisive for the retention of the whole homogeneous group of phenols studied. The contributions of polar interactions to the retention of individual phenols must be similar, however. The principal components extracted from the nineteen structural descriptors considered are not precise enough to account for changes in polar properties within a closely congeneric set of phenolic compounds. Hence the structural information contained in the non-empirical descriptors considered here does not suffice for the precise prediction of retention.

The same fundamental intermolecular interactions that determine chromatographic separations are also responsible for various physico-chemical properties of chemical compounds. These interactions, for instance, take place among molecules of a substance in a liquid state. It therefore seemed interesting to establish whether and to what extent the structural information extracted by PCA would be of value for the calculation of the boiling points of the solutes studied.

A stepwise regression analysis of boiling point data of the solutes, b.p., against the VARIMAX-rotated principal component scores yielded the following four-parameter regression equation:

$$\text{b.p.} = 457.9 - 33.2\ \text{PCvr1} - 15.2\ \text{PCvr2} + 23.0\ \text{PCvr3} + 9.2\ \text{PCvr4} \quad (4)$$

Eqn. 4 includes all twenty solutes studied and is characterized by $R = 0.944$, $s = 17.3$

and F-value = 30.9. The significance levels for eqn. 4 and the variables PCvr1, PCvr2, PCvr3 and PCvr4 are 0.0001, 0.00001, 0.001, 0.00001 and 0.03, respectively. The variable PCvr4 is hardly significant and thus the relationship described by eqn. 4 resembles eqn. 2. In both eqns. 2 and 4 the coefficients at PCvr1 and PCvr2 are negative whereas that at PCvr3 is positive. In eqn. 4, however, the variable PCvr1 predominates, reflecting strong dispersive attractions among molecules in liquid state. Also, the significance of PCvr3 is higher than that of PCvr2 in eqn. 4, in contrast to eqn. 2. Therefore, it may be concluded that in order to hold the solutes considered in the liquid state, dispersive intermolecular interactions are decisive, supplemented by specific polar interactions, whereas the reverse is true for HPLC retention in the graphitic carbon–hexane system.

The QSRR studies using principal components extracted from nineteen non-empirical structural descriptors provide evidence for the decisive role of specific, polar, electronic intermolecular interactions for the separation of solutes in a graphitic carbon–hexane HPLC system. This is in agreement with the general conclusions of our earlier work, in which we analysed retention on graphitic carbon from the point of view of the metallic character of the stationary phase and electron pair donor–acceptor interactions of solutes[17].

The principal component analysis of nineteen well established topological, information content and quantum chemical structural indices confirmed that these indices provide information mainly about the molecular size of chemical compounds. The information on electronic properties of the compounds which was condensed in PCvr2 and PCvr3 was also meaningful. Information contained in PCvr4 and PCvr5 is difficult to identify and has little relevance for the physico-chemical properties of the solutes considered here. The non-empirical molecular descriptors most often applied in QSRR studies still lack the full structural information required for the reliable prediction of molecular properties, including chromatographic retention. A further search for more specific structural parameters is required.

REFERENCES

1 R. Kaliszan, *Quantitative Structure–Chromatographic Retention Relationships*, Wiley, New York, 1987.
2 P. H. Weiner, C. J. Dack and D. G. Howery, *J. Chromatogr.*, 69 (1972) 249.
3 J. F. K. Huber, C. A. M. Meijers and J. A. R. J. Hulsman, *Anal. Chem.*, 44 (1972) 111.
4 J. F. K. Huber, E. T. Alderlieste, H. Harren and H. Poppe, *Anal. Chem.*, 45 (1973) 1337.
5 R. F. Hirsch, R. J. Gaydosh and J. R. Chrétien, *Anal. Chem.*, 52 (1980) 723.
6 L. Buydens, D. L. Massart and P. Geerlings, *Anal. Chem.*, 55 (1983) 738.
7 L. Buydens, D. Coomans, M. Vanbelle, D. L. Massart and R. Vanden Driessche, *J. Pharm. Sci.*, 72 (1983) 1327.
8 M. N. Hasan and P. C. Jurs, *Anal. Chem.*, 55 (1983) 263.
9 T. Cserháti, B. Bordás, L. Ekiert and J. Bojarski, *J. Chromatogr.*, 287 (1984) 385.
10 J. R. Chrétien, K. Szymoniak, C. Lion and J. K. Haken, *J. Chromatogr.*, 324 (1985) 355.
11 B. Walczak, M. Dreux, J. R. Chrétien, K. Szymoniak, M. Lafosse, L. Morrin-Allory and J. P. Doucet, *J. Chromatogr.*, 353 (1986) 109.
12 B. Walczak, J. R. Chrétien, M. Dreux, L. Morrin-Allory, M. Lafosse, K. Szymoniak and F. Membrey, *J. Chromatogr.*,, 353 (1986) 123.
13 K. Szymoniak and J. R. Chrétien, *J. Chromatogr.*, 404 (1987) 11.
14 S. Wold, L. Eriksson, S. Hellberg, J. Jonsson, M. Sjöström, B. Skageberg and C. Wikström, *Can. J. Chem.*, 65 (1987) 1814.
15 L. Eriksson, J. Jonsson, M. Sjöström and S. Wold, *Quant. Struct.-Act. Relat.*, 7 (1988) 144.

16 J. H. Knox, B. Kaur and G. R. Millward, *J. Chromatogr.*, 352 (1986) 3.
17 B. J. Bassler, R. Kaliszan and R. A. Hartwick, *J. Chromatogr.*, 461 (1989) 139.
18 R. C. West, M. J. Astle and W. H. Beyer (Editors), *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 1986–87.
19 B. V. Iophphe, *Refraktometricheskie Metody Khimii*, GNTIKhL, Leningrad, 1960.
20 L. B. Kier, *Acta Pharm. Jugosl.*, 36 (1986) 171.
21 L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, 65 (1976) 1806.
22 A. T. Balaban, *Chem. Phys. Lett.*, 89 (1982) 399.
23 M. Barysz, G. Jashari, R. S. Lall, V. K. Srivastava and N. Trinajstić, in R. B. King (Editor), *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983, pp. 222–230.
24 R. Sarkar, A. B. Roy and P. K. Sarkar, *Math. Biosci.*, 39 (1978) 299.
25 M. J. S. Dewar, *QCPE*, 11 (1975) 279 .
26 *Cambridge Crystallographic Database*, Cambridge Crystallographic Data Centre, Cambridge, 1979.
27 K. Ośmiałowski, J. Halkiewicz, A. Radecki and R. Kaliszan, *J. Chromatogr.*, 346 (1985) 53.
28 O. Kikuchi, *Quant. Struct.-Act. Relat.*, 6 (1987) 179.
29 K. Ośmiałowski, J. Halkiewicz and R. Kaliszan, *J. Chromatogr.*, 361 (1986) 63.
30 K.-J. Schaper and R. Kaliszan, in E. Mutschler and E. Winterfeldt (Editors), *Trends in Medicinal Chemistry*, VCH, Weinheim, Berlin, 1987, pp. 125–139.
31 J. R. Llinas and J. M. Ruiz, in G. Vernin and M. Chanon (Editors), *Computer Aids to Chemistry*, Ellis Horwood, Chichester, 1986, pp. 200–256.