

CHROMSYMP. 1711

Correlation between chemical structure of non-congeneric solutes and their retention on polybutadiene-coated alumina

ROMAN KALISZAN*

Department of Biopharmaceutics and Pharmacodynamics, Medical Academy, K. Marksa 107, 80-416 Gdańsk (Poland)

and

KRZYSZTOF OŚMIAŁOWSKI

Department of Inorganic Chemistry, Medical Academy, K. Marksa 107, 80-416 Gdańsk (Poland)

SUMMARY

The high-performance liquid chromatographic capacity factors extrapolated to pure water were determined on a polybutadiene-coated alumina (PBCA) stationary phase for a selected set of non-congeneric solutes. The structures of the test solutes were characterized by means of calculated hydrophobicity parameters and several non-empirical molecular descriptors, including topological indices, information content indices and quantum chemical indices. Unique properties of PBCA for the chromatographic determination of the hydrophobicity of diverse compounds were demonstrated. Quantitative structure-retention relationship (QSRR) studies employing non-empirical structural descriptors showed that the mechanism of retention on PBCA was similar to that on ODS phases. For a quantitative description of retention, structural parameters reflecting the bulkiness of solutes (positive input) and their polarity (negative input) predominated. Principal component analysis of structural descriptors most often used in QSRR studies allowed the extraction of two principal factors that describe retention more adequately than the individual descriptors studied. Informative values of non-empirical structural indices for the evaluation of retention were compared.

INTRODUCTION

Quantitative structure-retention relationships (QSRR) can be exploited for the prediction of retention, the determination of the structural properties of solutes and obtaining some insight into the molecular mechanism of chromatographic separations.

A reliable and precise prediction of retention in a given chromatographic system is possible for closely related, congeneric solutes only. More chemically meaningful seem to be the two remaining applications of QSRR. Bearing in mind the second aim of QSRR, *i.e.*, the determination of structural properties of solutes, we recently

turned our attention to the new generation of high-performance liquid chromatographic (HPLC) stationary phases, such as the polymer-coated reversed-phase materials introduced by Schomburg and co-workers^{1,2}. Using polybutadiene-coated alumina (PBCA) as an HPLC stationary phase, we previously proposed a method for the quantitative characterization of the hydrophobicity of diverse sets of solutes including neutral, weakly basic and weakly acidic compounds³. The uniqueness of the approach consists in its applicability to organic bases, which can be chromatographed in undissociated form. In contrast to silica-based hydrocarbonaceous stationary phases, commonly used for hydrophobicity determinations by HPLC, PBCA can be operated over a wide range of pH, including alkaline environments. The other advantage of PBCA is the lack of free silanol sites, which are responsible for undesirable specific interactions of chemically bonded silicas with polar solutes.

It seemed of interest to undertake QSRR studies of the data derived on PBCA aimed at the analysis of the mechanism of retention at the molecular level. The results of QSRR studies could be compared with those obtained previously by standard reversed-phase HPLC employing ODS materials^{4,5}. Comparison with the QSRR derived for the porous graphitic carbon–heptane normal-phase system also appeared interesting⁶.

In designing the experiment, we took the following into consideration: the solutes selected for studies should be non-congeneric and include basic, neutral and acidic compounds, the number of solutes should be sufficient for meaningful statistics but manageable for retention measurements and structural analysis and the test solutes should be stable conformationally, planar if possible and their geometry should be well established. The last condition is important for obtaining unequivocal results for structural and especially quantum chemical calculations.

Most QSRR reported in the chromatographic literature were derived by means of multiple regression analysis. The dependent variable formed a set of retention data, whereas various empirical, semi empirical and non-empirical structural parameters were assumed as independent (explanatory) variables. Unfortunately, the explanatory variables applied were often mutually interrelated by simple or multiple correlations. Moreover, it happened regularly that too many explanatory variables (including their various transformations such as squares, reciprocals and exponentials) were surveyed to describe too few retention data by multiple regression, which led to chance correlations. All this means that a number of QSRR regression equations reported in the analytical literature are without physical relevance.

Intercorrelated chemical data sets, which are unsuitable for multiple regression, can be subjected to multivariate analysis with factorial methods. Factor analysis has been applied to chromatographic data since the early 1970s, but only occasionally. The analysis reported were carried out in order to predict changes in the retention of a given set of solutes accompanying the changes in mobile–stationary phase systems^{7–10}. Wold and co-workers^{11,12} applied multivariate statistics to extract structural properties of amino acids from their various retention data.

There are numerous structural descriptors of solutes used in QSRR studies¹³. Unfortunately, significant intercorrelation among them limits the applicability of multiple regression in QSRR. On the other hand, according to the present chemometric theory, as many relevant data as possible should be considered in structure–property relationship studies because this increases the probability of good character-

ization of solutes¹². The large data tables resulting from such an assumption can be analysed by multivariate methods to extract the systematic information contained in the data. A small set of abstract principal factors extracted from a large set of various structural descriptors of the solutes should be sufficient for the description of their HPLC data expressed as logarithms of capacity factors. Based on the above assumption, we have recently been able to describe the retention of a set of variously substituted benzene derivatives on porous graphitic carbon with *n*-heptane eluent by means of two principal components extracted from a set of 18 non-empirical structural descriptors of 20 solutes¹⁴.

The experiments described in this paper were designed to obtain some insight into the molecular mechanism of retention on PBCA under reversed-phase conditions, to evaluate the informative value of non-empirical structural descriptors most often used in QSRR studies and to compare the quality of QSRR derived by means of multiple regression and of multivariate analysis.

EXPERIMENTAL

Materials

The test solutes were selected to include various structures (Fig. 1). The set of 21 solutes is diverse enough to avoid congenericity and includes aromatic hydrocarbons, substituted benzene derivatives and heteroaromatic compounds. Bearing in mind the limitations of quantum chemicals method of calculation of structural indices, the solutes chosen for the study had conformations that can be unequivocally defined. The selection of rigid structures eliminates the possibility that the conformation of a solute interacting with the components of chromatographic phases differs from the conformation for which structural descriptors are determined. By choosing planar test solutes, we had in mind the feasibility of the numerical characterization of molecular shape.

The solutes chromatographed were of the highest available purity and originated from various sources.

Determination of retention parameters

The chromatographic system consisted of a single-piston reciprocating pump and a UV detector operating at 254 nm. A Rheodyne injection valve fitted with a 10- μ l sample loop was used. The HPLC column used was kindly supplied by Prof. R. A. Hartwick (Department of Chemistry, Rutgers University, Piscataway, NJ, U.S.A.). A stainless-steel column (150 \times 4.6 mm I.D.), was slurry packed with polybutadiene-coated Spherisorb A5Y using isopropanol as the slurry solvent and methanol as the packing solvent. The stationary phase was prepared according to the procedure of Schomburg and co-workers^{1,2}. Polybutadiene was immobilized on the alumina support with the help of a cross-linking reaction involving radical formation.

The mobile phase consisted of analytical-reagent grade methanol and a Britton-Robinson universal buffer prepared at pH 2.62 and 11.15. To provide a pH of 2.62, 150 ml of 0.2 *M* NaOH were added to 1000 ml of solution of 0.04 *M* CH₃COOH, 0.04 *M* H₃PO₄ and 0.04 *M* H₃BO₃. Buffer of pH 11.15 was prepared analogously by adding 825 ml of 0.2 *M* NaOH to 1000 ml of acid solution. The ionic strength of the buffers was adjusted to 0.2 by adding NaNO₃. The following metha-

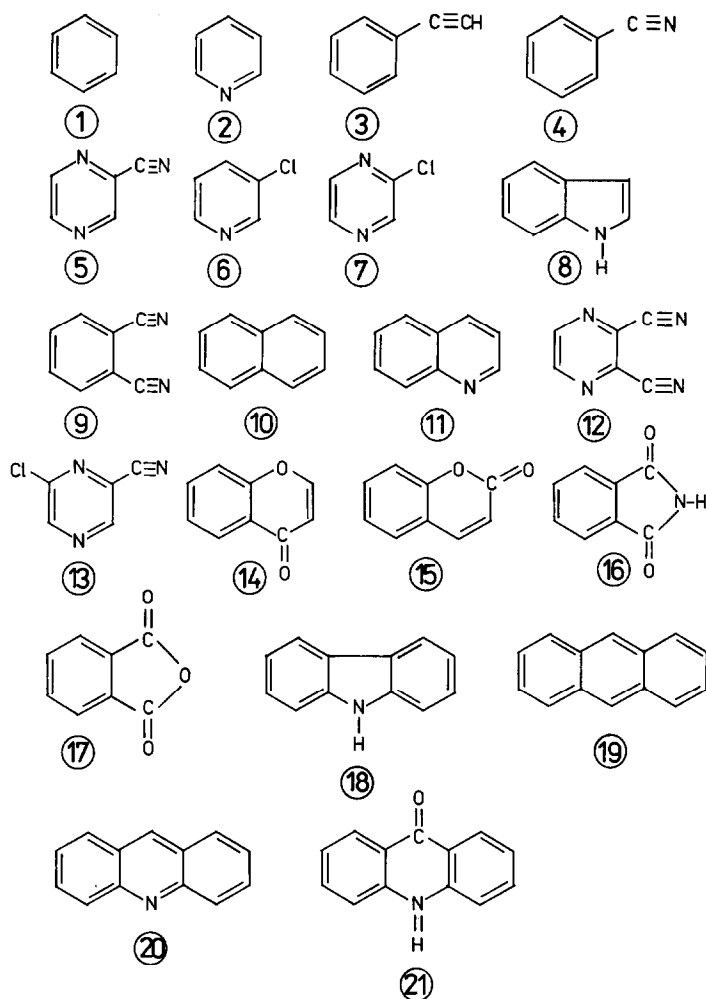


Fig. 1. Structural formulae of test solutes.

nol-buffer mixtures were used as mobile phases: 80:20, 50:50, 35:65, 20:80 and 0:100 (v/v). Before use, the eluents were filtered through 0.45- μ m nylon-66 filters. Alkaline buffer added to methanol in the proportion 20:80 (v/v) caused precipitation and this phase was discarded. Analytical-reagent grade chemicals were used.

The flow-rate was 1 ml/min. To calculate the capacity factors, k' , the solvent disturbance peak was used as a reference, observing specific precautions¹⁵.

For each solute the HPLC measurements were carried out in duplicate at both acidic and alkaline pH. Next, the $\log k'$ data were plotted against the volume fraction of methanol in the mobile phase for both pH values (see Fig. 2 for illustration). The linear part of the graph was extrapolated to zero content of methanol, yielding the value of capacity factor corresponding to pure water as the eluent, $\log k'_w$. Most of

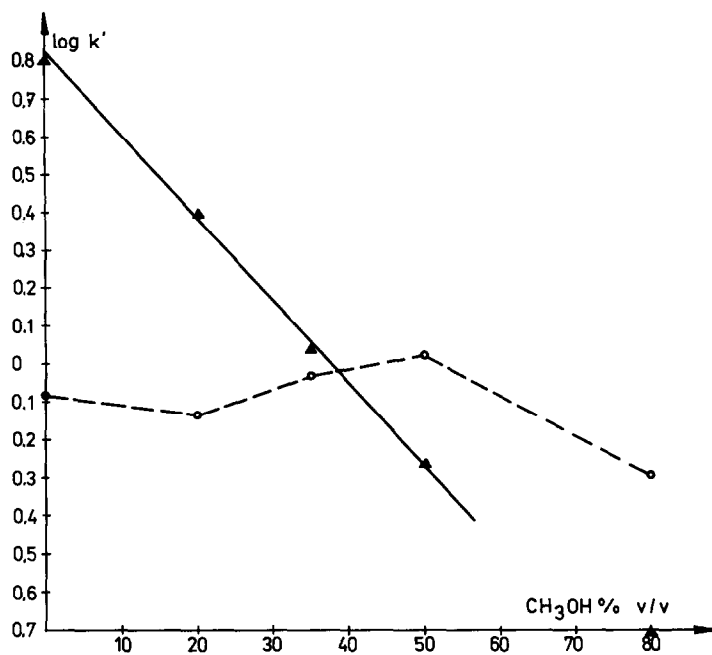


Fig. 2. Change in capacity factor with decreasing volume content of buffer in the eluent for quinoline. Solid line, best fit for the non-ionized solute (basic pH); broken line, retention of quinoline observed in the acidic eluent.

the compounds studied yielded measurable capacity factor data over a wide range of eluent compositions. In some instances, however, it was only possible to extrapolate linearly the retention observed for the two higher buffer contents studied. A similar approach has been applied by other workers¹⁶. Numerical values of $\log k'$ for non-ionized solutes extrapolated to pure aqueous eluent are given in Table I.

Structural descriptors

The HPLC system studied yields retention data reflecting the hydrophobicity of solutes. The reference scale for measurements of hydrophobicity, $\log P$, is provided by the *n*-octanol–water partitioning system. Experimentally determined $\log P$ data for 14 of 21 solutes studied were taken from ref. 17. For the full set of 21 compounds, the calculated $\log P$ data were obtained by applying the fragmental method of Hansch and Leo¹⁷. Calculated data were in good agreement with the data observed experimentally. The $\log P$ values are collected in Table I.

$\log P$ reflects ability of a compound to participate in the so-called hydrophobic interactions. Hydrophobic interactions are a complex net result of fundamental intermolecular interactions such as dispersive, inductive and dipole–dipole interactions. Thus, $\log P$ may be considered as a phenomenological structural parameter. The interpretation of empirical retention parameters in terms of other empirical or semi empirical parameters, such as $\log P$ data, is not very informative from the point of view of fundamental intermolecular interactions. More informative seem to be the

TABLE I

CAPACITY FACTORS EXTRAPOLATED TO PURE WATER, $\text{LOG } k'_w$ AND STRUCTURAL DESCRIPTORS OF THE SOLUTES NUMBERED AS IN FIG. 1

For explanation of the symbols of structural descriptors see text.

Solute No.	$\text{Log } k'_w$	<i>ind</i>	$\text{Log } P_{(\text{catet})}$	$\text{Log } P_{(\text{catet})}$	<i>molwt</i>	<i>bondrefr</i>	<i>etotal</i>	<i>ethomo</i>	<i>etumo</i>
1	1.0750	0	2.14	2.13	78.11	26.184	-47.0940	-0.5100	0.1463
2	-0.3400	1	0.65	0.65	79.10	24.464	-50.8780	-0.4655	0.1278
3	1.8600	0	2.53	2.53	102.14	33.350	-61.0640	-0.4760	0.1258
4	0.4324	0	1.56	1.56	103.12	30.624	-64.8855	-0.4899	0.1080
5	-0.3575	1		-0.68	105.09	27.180	-72.4109	-0.4653	0.0827
6	0.2197	1	1.38	1.36	113.55	29.296	-64.5365	-0.4333	0.1093
7	-0.1656	1		0.25	114.54	27.574	-70.0568	-0.4651	0.0816
8	1.0800	1	2.13	2.14	117.15	36.460	-73.7969	-0.4118	0.1180
9	0.2043	0		1.08	128.13	35.064	-82.6595	-0.4730	0.0958
10	2.5800	0	3.35	3.28	128.17	42.976	-74.5485	-0.4097	0.1127
11	0.8170	1	2.04	2.02	129.16	41.254	-80.8656	-0.4602	0.0688
12	-0.1504	1		-0.71	130.11	31.620	-90.1945	-0.4640	0.0768
13	-0.0220	1		0.04	139.54	32.014	-87.8378	-0.4726	0.0630
14	0.4608	0		1.10	146.13	39.346	-105.2470	-0.4168	0.0618
15	0.5210	0	1.39	1.40	146.15	39.266	-105.2660	-0.4204	0.0488
16	0.2410	0	1.15	0.97	147.13	36.964	-104.8130	-0.4339	0.0687
17	0.1640	0	-0.62	-0.72	148.12	34.984	-114.9990	-0.4909	0.0477
18	2.7550	1	3.59	3.52	167.21	53.252	-100.1160	-0.3572	0.0510
19	3.8500	0	4.45	4.44	178.23	59.768	-107.0620	-0.3684	0.0383
20	1.8500	1	3.40	3.40	179.22	58.046	-107.0290	-0.3584	0.0483
21	1.5150	1		2.99	195.22	56.476	-129.2860	-0.3924	0.0761

non-empirical structural descriptors. Under the term non-empirical we understand here the molecular parameters that can be calculated exclusively on the basis of the structural formula of a solute. To calculate the following structural descriptors considered in this paper and given in Table I, only the structural formula of a solute is required.

Molecular weight is denoted here as *molwt*. Molecular refractivity (*bondrefr*) is calculated as the sum of the bond refractivities for all pairs of connected atoms according to Vogel *et al.*¹⁸. Thus, the refractivity increment for a C-H fragment is 1.676, for C-C 1.296, for C=C 4.17, for C-Cl 6.51, etc. A list of individual bond increments can be found in ref. 13 (p. 96). In fact, the *bondrefr* parameter is semi-empirical in nature but, in contrast to $\text{log } P$, its calculation and interpretation cause no ambiguities.

The following molecular topological indices were considered here: Wiener index, W (*wiener*)¹⁹, generalized molecular connectivity indices of first and second order, ${}^1\chi^v$ and ${}^2\chi^v$, (*chi1v* and *chi2v*) according to Kier and Hall²⁰, and indices of molecular shape of first and second order, κ^1 and κ^2 (*kappa1* and *kappa2*), proposed by Kier²¹. Certainly there are other empirically modified topological indices designed for heteroatom-containing compounds, but strong intercorrelations among the exist-

<i>dipolem</i>	<i>delta</i>	<i>wiener</i>	<i>chi1v</i>	<i>chi2v</i>	<i>ic0</i>	<i>ic1</i>	<i>ic2</i>	<i>kappal</i>	<i>kappa2</i>	<i>shape</i>
0.00	0.0115	27	2.0000	1.1547	1.0000	1.0000	1.0000	3.4116	1.6058	1.294
2.06	0.2467	27	1.8497	1.0246	1.3486	1.9717	2.8454	3.3442	1.5526	1.199
0.32	0.1752	64	2.4494	1.4880	0.9852	1.8703	2.4953	4.9275	2.1818	1.481
3.28	0.2490	64	2.3843	1.4778	1.2957	1.9143	2.2878	4.8590	2.1309	1.400
3.16	0.2545	64	2.0938	1.2025	1.5395	2.6635	2.8454	4.7222	2.0300	1.526
2.21	0.3485	42	1.9494	1.1712	1.6767	2.5503	3.0958	1.5953	1.9374	1.273
2.12	0.3011	42	1.8092	1.0281	1.8464	2.4464	2.8464	4.5270	1.8880	1.373
1.92	0.2485	79	2.8880	2.0649	1.2718	2.3050	3.1494	4.6211	1.6217	1.200
5.81	0.4025	117	2.7746	1.7692	1.3788	2.2359	2.5216	6.3120	2.6602	1.092
0.00	0.0605	109	3.4047	2.3472	0.9911	1.3921	1.8366	5.4822	2.1431	1.235
2.20	0.2713	109	3.2645	2.1961	1.2533	2.0192	2.7778	9.6300	4.4962	1.333
5.61	0.3306	117	2.4942	1.4772	1.4591	2.5850	2.5850	6.1840	2.5619	1.162
2.73	0.2902	88	2.2036	1.3388	1.7899	3.0581	3.2776	5.8102	2.3000	1.492
3.37	0.5837	140	3.3433	2.2810	1.3793	2.4922	3.2869	7.6943	2.3247	1.333
5.02	0.7228	145	3.3504	2.2923	1.3793	2.2628	3.0575	6.3098	2.3247	1.265
2.08	0.7716	139	3.2356	2.3288	1.6492	2.6250	2.2750	6.0949	1.9153	1.197
4.55	0.6440	139	3.1438	2.2226	1.4566	2.4402	2.7069	6.2443	2.0003	1.092
1.73	0.3839	220	4.4047	3.2162	1.2072	2.0958	2.6412	4.9873	1.6948	1.481
0.01	0.0447	280	4.8094	3.5465	0.9799	1.4834	2.1174	7.5714	2.8451	1.559
2.20	0.3752	280	4.6793	3.3738	1.1916	1.9558	2.5557	5.7595	2.8080	1.527
4.43	0.5495	335	4.4588	3.5453	1.3918	2.4183	3.1887	8.3842	3.0367	1.473

ing topological indices (and also molecular weight) are observed and some indices just duplicate others. The topological indices selected here are those most often applied for QSRR.

A separate group of structural indices form information indices of neighbourhood symmetry of zeroth order, *ic0*, first order, *ic1*, and second order, *ic2*²². These information content indices were calculated from probabilities of finding equivalent atoms or patterns of atoms in a given structural formula.

The molecular shape of the planar solutes studied was described by the parameter of shape, defined as the ratio of the longer to the shorter side of the rectangle of minimum area enveloping the structure drawn assuming standard van der Waals atomic radii²³.

A large group of non-empirical structural descriptors formed quantum chemical indices. Molecular parameters of orbitals were calculated with complete neglect of differential overlap, CNDO/2 method, using a standard program²⁴. The geometry assumed for calculation was based on standard crystallographic data. The following indices were calculated: total energy (*etotal*), energy of the highest occupied molecular orbital (*ehomo*), energy of the lowest unoccupied molecular orbital (*elumo*), dipole moment (*dipolem*) and electron excess charges on individual atoms. Energies

are given in Table I in electronvolts, dipole moments in debyes and excess charges in electrons. Based on charge distribution, the submolecular polarity parameter, Δ (delta), was determined as the maximal difference of electron excess charge for two atoms in the molecule⁴.

To differentiate between aromatic hydrocarbon derivatives and heterocyclic compounds, an indicator variable, *ind*, was introduced for which a value 1 was assigned in the case of benzene derivative and 0 for the remaining solutes.

Statistical analysis

Regression analysis was applied to relate retention data to calculated $\log P$ data.

Multiple regression was performed using non-empirical structural descriptors as explanatory variables and $\log k'_w$ as the dependent variable.

Principal component analysis (PCA) was carried out on a set of 16 non-empirical structural descriptors of 21 solutes. The calculation scheme proposed by Schaper and Kaliszan²⁵ was followed. The resulting principal component scores for individual solutes were related to the retention parameters by means of multiple regression analysis.

RESULTS AND DISCUSSION

The relationship between $\log k'_w$ extrapolated to pure water as eluent and $\log P$ calculated by the fragmental method is illustrated in Fig. 3. The relationship presented in Fig. 3 is described by the following regression equation:

$$\log k'_w = -0.052 + 0.208 (\log P)^2 \quad (1)$$

$$n = 21, s = 0.279, R = 0.9711, F = 314$$

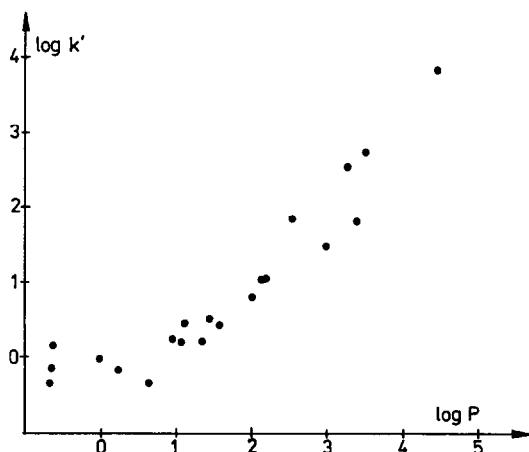


Fig. 3. Relationship between logarithms of capacity factors extrapolated to pure water as a mobile phase, $\log k'_w$, and logarithms of *n*-octanol-water partition coefficients calculated by the fragmental method of Hansch and Leo¹⁸, $\log P$.

where n is the number of compounds used for deriving the regression equation, s is the standard error of the estimate, R is the correlation coefficient and F is the value of the F -test. The statistical evaluation of eqn. 1 is satisfactory, especially as it includes a wide diversity of structures. The curvature of the $\log k'_w$ vs. $\log P$ plot is especially pronounced for highly hydrophilic solutes. The experimentally observed curvature of the relationship may be artificial to some extent. The retention of hydrophilic solutes on the relatively short column used is too low to obtain precise $\log k'_w$ data. The HPLC system applied was designed to prove the general validity of the approach and the differences in hydrophobicities of individual test solutes exceeded five $\log P$ units. When dealing with a set of less hydrophobic solutes, one can apply longer columns and/or a PBCA material of higher polymer coating. Nevertheless, the HPLC procedure applied here employing a 15-cm column packed with PBCA may be recommended for the evaluation of the hydrophobicity of compounds with $\log P$ above 0.5. The observation^{26,27} that in certain instances the calculated hydrophobicity parameters are more trustworthy than those derived experimentally by HPLC is also valid for PBCA when dealing with hydrophilic solutes.

Similar curvilinear relationships as expressed by eqn. 1 have been reported on ODS phases for barbiturates²⁸ and diols²⁹. The striking difference between PBCA and ODS is that on the former a single correlation between $\log k'_w$ and $\log P$ is obtained for a variety of solutes whereas on ODS separate relationships are reported for individual classes of compounds^{26,30,31}. Two correlation lines (for hydroxylated and non-hydroxylated solutes) were also reported on styrene-divinylbenzene copolymer stationary phase¹⁶.

The single $\log k'_w$ vs. $\log P$ relationship may be due to insignificant interactions of solutes with the alumina support in the case of PBCA as opposed to relatively stronger interactions with silanol sites in the case of ODS. The present results support our previous observations of the unique properties of PBCA as a reversed-phase material for hydrophobicity determinations.

We also attempted to relate the determined capacity factors to non-empirical molecular structural descriptors of solutes. There are several reports on the importance of molecular size for the retention of congeneric solutes on ODS^{28,32-34}. Previously, we succeeded in obtaining retention data of benzene derivatives from reversed-phase HPLC on ODS by means of size-related and a polarity-related structural descriptors^{4,5}. In a two-parameter regression equation, the size-related descriptor was the quantum chemically calculated total energy (etotal), whereas polarity was quantified by the maximum difference of electron excess charges (δ).

Here we performed the multiple regression analysis of $\log k'_w$ data using the non-empirical structural descriptors given in Table I. The most meaningful regression equation found was

$$\log k'_w = -1.618 + 0.089 \text{ bondrefr} - 2.505 \delta \quad (2)$$

$$n = 21, s = 0.500, R = 0.9090, F = 42.8$$

The term bondrefr is significant at the level of at least 0.001 and δ at the level of 0.002. An equation of similar statistical value was obtained when using the molecular connectivity indices χ_{1v} or χ_{2v} instead of bondrefr. No statistically valid improve-

ment of eqn. 2 was obtained by introducing of additional structural descriptors considered. Only when an indicator variable, ind, was introduced did the quality of correlation increase (eqn. 3). The indicator variable was assigned a value of 1 for benzene derivatives and 0 for heterocyclic solutes.

$$\log k'_w = -1.272 + 0.089 \text{ bondrefr} - 2.648 \text{ delta} - 0.598 \text{ ind} \quad (3)$$

$$n = 21, s = 0.394, R = 0.9476, F = 49.8$$

Intercorrelations found among structural parameters limit the applicability of multiple regression analysis in QSRR studies. Such intercorrelated data, however, can be analysed by multivariate statistical methods. The set of structural data considered here consisted of the last 16 columns in Table I, *i.e.*, excluding $\log P$ and ind.

Two principal factors extracted from the structural data set appeared meaningful for the description of the retention of the solutes. The first factor accounted for 48.6% of the variance in structural data considered and the second factor for 25.2%.

Subsequently, the loadings (eigenvectors) of the two principal components extracted were calculated. The loadings were subjected to VARIMAX rotation to maximize the variance of squared loadings and thus facilitate interpretation of meaning of individual principal components. In Fig. 4 the loadings of the two principal compo-

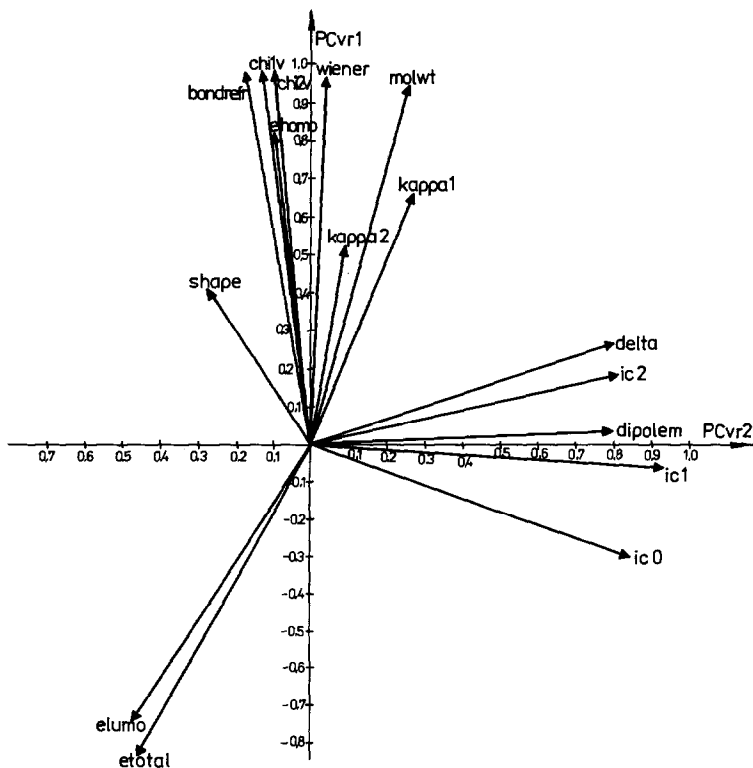


Fig. 4. VARIMAX-rotated loadings of two first principal components by individual structural descriptors denoted as in Table I.

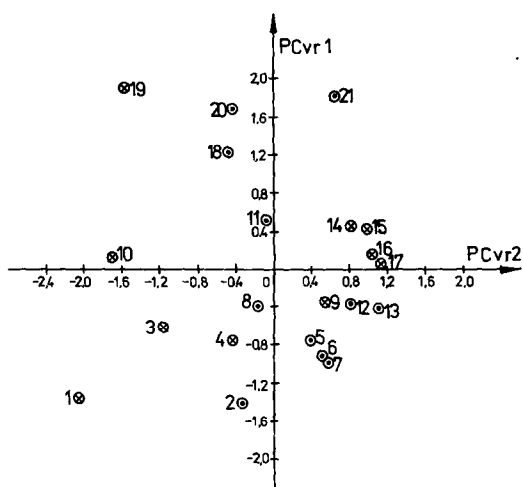


Fig. 5. Principal component scores after VARIMAX rotation for the solutes numbered as in Fig. 1.

nents by individual structural descriptors are depicted. The principal component scores for the set of 21 solutes studied, scaled to a variance of 1, obtained after VARIMAX rotation are presented in Fig. 5.

As shown in Fig. 4 the first principal component, PCvr1, is loaded mostly by such structural descriptors as chi1v , bondrefr , chi2v , wiener , molwt and etotal . All these structural descriptors reflect basically the size (bulkiness) of solutes. In such a situation PCvr1 condenses information on molecular size.

The second principal component, PCvr2, is loaded predominantly by the structural descriptors ic0 , ic1 , ic2 , delta and dipolem . Hence it can be concluded that PCvr2 concentrates structural information related to the so-called molecular polarity. The polar properties of chemical compounds are a function of the electron distribution within a molecule. 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 should be devoted to the high loadings of PCvr2 by the information content indices ic0 , ic1 and ic2 . These indices reflect the diversity in the atom composition of molecules. This is connected with specific, polar properties of solutes. In studies on structural descriptors of benzene derivatives we also observed a high loading by information content indices of the principal component reflecting the molecular polarity of solutes.

It seemed of interest to apply the structural information condensed in two principal components to correlation studies with liquid chromatographic data derived on PCBA. Eqn. 4 describes $\log k'_w$ in terms of non-rotated principal component scores scaled to a variance of 1, *i.e.*, PC1 and PC2:

$$\log k'_w = 0.8852 + 0.5942 \text{ PC1} - 0.9016 \text{ PC2} \quad (4)$$

$$n = 21, s = 0.380, R = 0.9485, F = 80.6$$

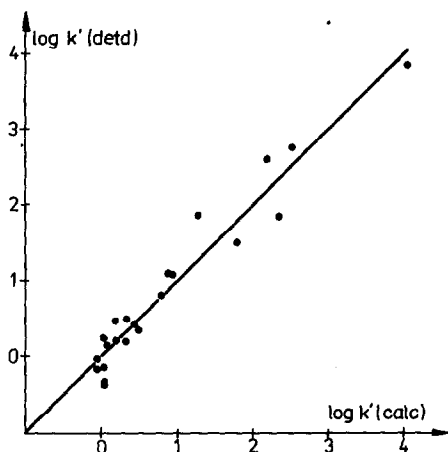


Fig. 6. Relationship between $\log k'$ (extrapolated to pure water) determined experimentally and calculated by eqn. 4.

Both the PC1 and PC2 terms are significant at at least the 0.001 significance level. The relationship between experimental $\log k'$ data and those calculated by eqn. 4 is illustrated in Fig. 6.

The statistical characteristics of eqn. 4 are better than those of eqn. 2, as indicated by the standard error, correlation coefficient and F -test. Qualitatively both equations are similar in that increasing solute bulkiness enhances retention whereas increasing polarity has the opposite effect. Eqn. 4 demonstrates that the correlation represented by eqn. 2 is not fortuitous.

Structural information extracted by principal component analysis of molecular descriptors appeared to be more precise and reliable than the respective information provided by individual structural parameters. Comparison of eqns. 4 and 3 suggests that the principal components extracted from the set of sixteen molecular descriptors contain information that allows the differentiation of benzene and heterocyclic derivatives. In other words, these factors contain information expressed explicitly by the indicator variable, *ind*, which was not included in multivariate analysis.

Analysing eqns. 2 and 4 from the point of view of the mechanism of reversed-phase HPLC on PBCA, one can conclude that non-specific, dispersive interactions of solutes with the stationary phase prevail over the analogous interactions with the mobile phase. This is indicated by the positive increment to eqns. 2 and 4 by structural parameters reflecting the ability of solutes to participate in dispersive (London-type) intermolecular interactions. On the other hand, the negative increment to eqns. 2 and 4 of the molecular parameters reflecting the ability of solutes to participate in specific, polar intermolecular interactions demonstrates that polar solute–mobile phase interactions are stronger than analogous solute–stationary phase interactions. Hence the mechanism of retention on PBCA is closely similar to that on ODS^{4,5} but basically different from that postulated for normal-phase HPLC on porous graphitic carbon^{6,14}.

The structural descriptors considered in this work can be conveniently determined by simple calculation procedures for any given structural formula. The ques-

tion is how much information about the properties of chemical compounds they contain. Principal component analysis of sixteen quantum chemical, information content and topological indices, together with molecular weight, molecular refractivity and molecular shape descriptors, allows the extraction of information useful for the determination of physico-chemical properties. Analysing loadings of principal components by individual structural descriptors (Fig. 4), one can conclude that total energy (etotal) and topological indices (wiener, chi1v and chi2v) provide information on the bulkiness of the solutes and have no advantages over established bulk measures such as molecular weight and refractivity.

Electron charge distribution calculated by standard CNDO/2 calculations provides information on the polar properties of solutes. Again, the submolecular polarity measure δ appeared to be a more appropriate descriptor of polar properties of solutes than the total dipole moment.

The CNDO/2 method is reliable as far as charge distribution is concerned but less reliable for orbital energy calculations. In such a situation, it is difficult to determine whether the ehomo and elumo descriptors fail to account for the ability of the solutes to participate in electron pair donor-acceptor interactions, or whether these interactions are of little importance for retention.

Further studies are required to explain the physical meaning of information content indices.

Neither the Kier indices of shape, kappa1 and kappa2, nor the shape parameter previously proposed in this laboratory, shape, appeared meaningful for the description of retention. Probably retention differences caused by differences in molecular shape are too subtle to be detectable for such a diverse set of solutes. Hence the question remains open of whether the shape parameters considered differentiate molecular properties or just structural formulae.

REFERENCES

- 1 U. Bien-Vogelsang, A. Deege, H. Figge, J. Kohler and G. Schomburg, *Chromatographia*, 19 (1984) 170.
- 2 G. Heinemann, J. Kohler and G. Schomburg, *Chromatographia*, 23 (1987) 435.
- 3 R. Kaliszán, R. W. Blain and R. A. Hartwick, *Chromatographia*, 25 (1988) 5.
- 4 R. Kaliszán, K. Ośmiałowski, S. A. Tomellini, S.-H. Hsu, S. D. Fazio and R. A. Hartwick, *Chromatographia*, 20 (1985) 705.
- 5 R. Kaliszán, K. Ośmiałowski, S. A. Tomellini, S.-H. Hsu, S. D. Fazio and R. A. Hartwick, *J. Chromatogr.*, 352 (1986) 141.
- 6 B. J. Bassler, R. Kaliszán and R. A. Hartwick, *J. Chromatogr.*, 461 (1989) 139.
- 7 J. F. K. Huber, C. A. M. Meijers and J. A. R. J. Hulsman, *Anal. Chem.*, 44 (1972) 111.
- 8 R. F. Hirsch, R. J. Gaydos and J. R. Chretián, *Anal. Chem.*, 52 (1980) 723.
- 9 M. N. Hasan and P. C. Jurs, *Anal. Chem.*, 55 (1983) 263.
- 10 G. Gullner, T. Cserháti, B. Bordás and M. Szógyi, *J. Liq. Chromatogr.*, 9 (1986) 1919.
- 11 S. Wold, L. Eriksson, S. Hellberg, J. Jonsson, M. Sjöström, B. Skageberg and C. Wikström, *Can. J. Chem.*, 65 (1987) 1814.
- 12 L. Eriksson, J. Jonsson, M. Sjöström and S. Wold, *Quant. Struct.-Act. Relat.*, 7 (1988) 144.
- 13 R. Kaliszán, *Quantitative Structure-Chromatographic Retention Relationships*, Wiley, New York, 1987.
- 14 R. Kaliszán, K. Ośmiałowski, B. J. Bassler and R. A. Hartwick, *J. Chromatogr.*, 499 (1989) 333.
- 15 J. H. Knox and R. Kaliszán, *J. Chromatogr.*, 349 (1985) 211.
- 16 S. Bitteur and R. Rosset, *J. Chromatogr.*, 394 (1987) 279.
- 17 C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.
- 18 A. I. Vogel, W. T. Cresswell, G. H. Jeffery and J. Leicester, *J. Chem. Soc.*, (1952) 514.

- 19 H. Wiener, *J. Phys. Chem.*, 52 (1948) 1082.
- 20 L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, 65 (1976) 1806.
- 21 L. B. Kier, *Acta Pharm. Jugosl.*, 36 (1986) 171.
- 22 R. Sarkar, A. B. Roy and P. K. Sarkar, *Math. Biosci.*, 39 (1978) 299.
- 23 R. Kaliszan, H. Lamparczyk and A. Radecki, *Biochem. Pharmacol.*, 28 (1979) 123.
- 24 P. A. Dobosh, *QCPE*, 11 (1969) 141.
- 25 K. -J. Schaper and R. Kaliszan, in E. Mutschler and E. Winterfeldt (Editors), *Trends in Medicinal Chemistry*, VCH, Weinheim, Berlin, 1989, pp. 125–139.
- 26 C. V. Eadsforth, *Pestic. Sci.*, 17 (1986) 311.
- 27 A. J. Leo, *J. Pharm. Sci.*, 76 (1987) 166.
- 28 M. J. M. Wells, C. R. Clark and R. M. Patterson, *J. Chromatogr. Sci.*, 19 (1981) 573.
- 29 D. Noël and P. Vangheluwe, *J. Chromatogr.*, 388 (1987) 75.
- 30 J. J. Sabatka, D. J. Minick, T. K. Schumaker, G. L. Hodgson, Jr., and D. A. Brent, *J. Chromatogr.*, 384 (1987) 349.
- 31 K. Miyake, N. Mizuno and H. Terada, *J. Chromatogr.*, 439 (1988) 227.
- 32 M. J. M. Wells, C. R. Clark and R. M. Patterson, *Anal. Chem.*, 58 (1986) 1625.
- 33 R. H. Rohrbaugh and P. C. Jurs, *Anal. Chem.*, 59 (1987) 1048.
- 34 E. J. Tierney, J. M. Bellama, G. Eng, F. E. Brinckman and R. B. Johannesen, *J. Chromatogr.*, 441 (1988) 229.