Comparative Study of Topological and Linear Free Energy-Related Parameters for the Prediction of GC Retention Indices

L. BUYDENS *, D. COOMANS *, M. VANBELLE *, D. L. MASSART *ˣ, and R. VANDEN DRIESSCHE ‡

Abstract □ The different molecular connectivity indices were considered as to their capacity for describing GC retention indices of a data set consisting of molecules of different chemical families. $^1\chi$ describes best the chromatographic behavior on nonpolar stationary phases, whereas $^3\chi_p$ in combination with an electronic parameter (σ) yields the best results when using the polar stationary phases.

Keyphrases □ Molecular connectivity—correlation to GC retention indices, polar and nonpolar stationary phases, topological and linear free-energy parameters □ GC retention indices—correlation to molecular connectivity, polar and nonpolar stationary phases, topological and linear free-energy parameters □ Topology—correlation of GC retention indices and molecular connectivity, linear free energy parameters, polar and nonpolar stationary phases

The pharmacological action of drugs results from the interaction of the molecules with receptor sites. These interactions are often described in terms of structural parameters (1–3). GC retention indices, which result from the interaction of the molecules with the stationary phase, have been shown (4–8) to be related to the various structural parameters used to describe drug–receptor interactions. Therefore, the GC behavior of the molecules can be used as a simple model for methodological studies of structure–activity relationships.

Relating GC behavior to structural parameters such as the molecular connectivity ($^1\chi$) or the combination of $^1\chi$ with physicochemical parameters such as σ and π (5, 7), leads to satisfactory results when one studies molecules of a single chemical family such as the alcohols, alkanes, or the methyl esters of fatty acids (5, 7). For molecules of different families, especially using more polar stationary phases, the results show that the parameters used do not satisfactorily describe the GC behavior (5). Therefore, use of additional parameters that describe the variance unexplained by the other variables is necessary.

This paper investigates to what extent the additional higher-order terms of the molecular connectivity permit better results. For this purpose the higher-order terms are included in a statistical treatment together with the previously used structural parameters (5).

THEORETICAL

The parameters used are the valence molecular connectivity indices (hereafter called connectivity indices) the Wiener number (W), Hosoya's index (Z), Hansch's constant (π), and Hammett's constant (σ). The full definitions and calculations for these parameters can be found elsewhere (1–3, 9, 10).

To calculate the connectivity parameters the molecule is displayed in skeletal form. To each of the i atoms a value, δ_i, is assigned according to the difference between the number of valence electrons and the number of hydrogen atoms suppressed. An example using 2-methyl-1-butanol (I) is given in Fig. 1.

The respective connectivity indices are calculated by means of the following equations. For the zero-order term ($^0\chi$):

$$^0\chi = \sum_1^i (\delta_i)^{-1/2} \tag{Eq. 1}$$

where δ_i represent the δ-values of the i atoms. For the 2-methyl-1-butanol example, $^0\chi = (2/\sqrt{1}) + (2/\sqrt{2}) + (1/\sqrt{3}) + (1/\sqrt{5}) = 4.439$. For the first-order term ($^1\chi$):

$$^1\chi = \sum_i^N (\delta_i \delta_j)^{-1/2} \tag{Eq. 2}$$

where $\delta_{i,j}$ are the δ-values of the adjacent atoms i and j, and N is the number of bonds in the molecule. Using 2-methyl-1-butanol, $^1\chi = (1/\sqrt{1 \times 2})(2/\sqrt{2 \times 3}) + (1/\sqrt{1 \times 3}) + (1/\sqrt{2 \times 5}) = 2.417$. For the second-order term ($^2\chi$):

$$^2\chi = \sum^M (\delta_i \delta_j \delta_k)^{-1/2} \tag{Eq. 3}$$

where $\delta_{i,j,k}$ are the δ-values of the atoms of two adjacent bonds, and M is the number of adjacent bonds pairs in the molecule. For the 2-methyl-1-butanol example, $^2\chi = (1/\sqrt{1 \times 2 \times 3}) + (1/\sqrt{2 \times 3 \times 1}) + (1/\sqrt{1 \times 3 \times 2}) + (1/\sqrt{3 \times 2 \times 5}) + (1/\sqrt{2 \times 2 \times 3}) = 1.696$. For the third-order path term ($^3\chi_p$):

$$^3\chi_p = \sum^p (\delta_i \delta_j \delta_k \delta_l)^{-1/2} \tag{Eq. 4}$$

where $\delta_{i,j,k,l}$ are the δ-values for the atoms of three adjacent bonds forming the butane skeleton (C—C—C—C) and p is the number of three adjacent bonds forming the butane skeleton in the molecule. Using 2-methyl-1-butanol, $^3\chi_p = (1/\sqrt{1 \times 2 \times 3 \times 1}) + (1/\sqrt{1 \times 2 \times 3 \times 2}) + (1/\sqrt{2 \times 3 \times 2 \times 5}) + (1/\sqrt{1 \times 3 \times 2 \times 5}) = 1.009$. For the third-order cluster term ($^3\chi_c$):

$$^3\chi_c = \sum^Q (\delta_i \delta_j \delta_k \delta_l)^{-1/2} \tag{Eq. 5}$$

where $\delta_{i,j,k,l}$ are the δ-values for the atoms of three adjacent bonds forming the isobutane skeleton, and Q is the number of isobutane groups in the molecule. For 2-methyl-1-butanol: $^3\chi_c = (1/\sqrt{2 \times 3 \times 2 \times 1}) = 0.289$.

The Wiener number (W) and Hosoya's index (Z) are also topological parameters. Formulas for their calculation can be found elsewhere (3, 11). Because the molecules considered herein are all aliphatic molecules the σ scale, derived from the hydrolysis of aliphatic acid esters was used (1, 2, 10).

EXPERIMENTAL

The different parameters were calculated for ~100 compounds[1] belonging to different families (alkanes, alkenes, alcohols, aldehydes, ketones, and esters). Retention indices (12) on four stationary phases of different polarity were used: squalane and SE-30 were used as the nonpolar phases, while ethylene glycol was of intermediate polarity and di-
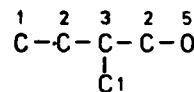

Figure 1—δ-Values for the atoms of 2-methyl-1-butanol (I). δ-Values of the respective atoms are given above (beside) the atoms.

---

[1] Furnished on request.

Table I—Correlation Coefficients Between Parameters

	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi_p$	$^3\chi_c$	W	Z	π	σ
$^0\chi$	1.000								
$^1\chi$	0.956	1.000							
$^2\chi$	0.812	0.688	1.000						
$^3\chi_p$	0.757	0.786	0.650	1.000					
$^3\chi_c$	0.348	0.134	0.789	0.250	1.000				
W	0.869	0.861	0.559	0.525	0.064	1.000			
Z	0.802	0.831	0.457	0.528	-0.040	0.950	1.000		
π	0.685	0.721	0.545	0.474	0.179	0.617	0.536	1.000	
σ	-0.335	-0.345	-0.309	-0.362	-0.410	-0.102	-0.029	-0.742	1.000

ethylene glycol succinate was used as the polar phase. The molecular connectivity indices were calculated with a microcomputer[2] using a program developed in-house.

Multiple regressions were calculated following a stepwise regression procedure using the SPSS program (13). The independent variables were introduced in the multiple regression equation only if they met certain statistical criteria. The variable that accounts for the largest amount of variance in the dependent variable (with the highest correlation coefficient) was used first in the regression equation. The variable that explains the greatest amount of variance not accounted for by the variables already in the equation entered the regression equation in the following step. Correlation studies between the different connectivity indices were also carried out using the SPSS program (13).

RESULTS AND DISCUSSION

To find the relationship between the different connectivity indices, correlation coefficients between these indices were calculated. The different structural features necessary to calculate all the connectivity indices are not present in all molecules. Therefore, some of the indices are not different from zero for all the molecules of the basic set. The correlation coefficients are calculated only for the molecules with a non-zero value for the indices. A subset was chosen to compare the different indices such that the indices compared are all different from zero for all the molecules of the subset. The resulting correlation matrix is shown in Table I.

The results show that the parameters W and Z correlate best with the zero- and first-order terms of the molecular connectivity. The parameter σ, as expected, is completely different from the other variables, since its correlation coefficient with the other variables is nearly zero. The correlation between the connectivity indices decreases for the higher-order terms. $^3\chi_c$ does not correlate with the path connectivity term. These results, in agreement with those of Kier and Hall (3), suggest that the higher-order terms of the molecular connectivity contain different information which can indeed improve the regression equation.

Regression equations were calculated for the retention indices on the four stationary phases using these parameters. The equations are of the general form:

$$(RI)_i = a + bx_{1i} + cx_{2i} + dx_{3i} + \ldots + e_i \qquad \text{(Eq. 6)}$$

where RI is the retention index and $x_{1i}, \ldots, x_{3i}$ are the values for the different structural parameters for compound i. The results are given in Table II; the multiple correlation (MR) and the simple correlation (SR) coefficients are given. The multiple correlation coefficient is the correlation coefficient of the regression equation with all variables entered up to the considered step.
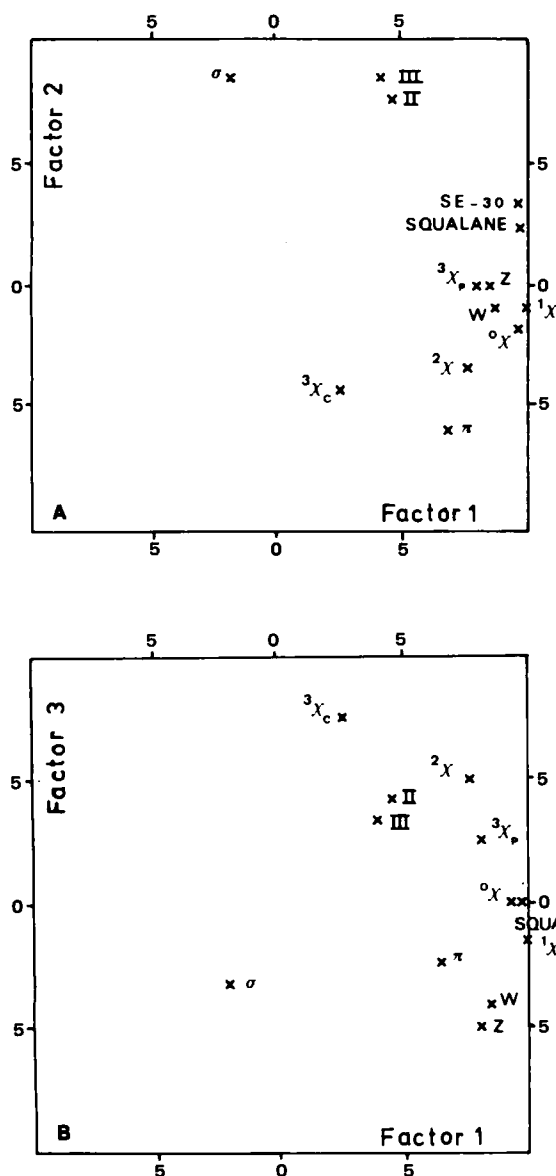

Figure 2—Factor analysis of the different structural parameters and the GC retention indices.

Table II—Correlation Coefficients of the Stepwise Regression Equations for the Total Data Set [a]

Stationary Phase	Parameter	Multiple Correlation	Simple Correlation
Nonpolar			
Squalane	$^1\chi$	0.91	0.91
	π	0.94	0.47
SE-30	$^1\chi$	0.84	0.84
	π	0.92	0.37
Polar			
Diethylene glycol succinate			
	σ	0.50	0.50
	$^3\chi_p$	0.68	0.25
	Z	0.71	0.07
	$^1\chi$	0.72	0.10
	$^0\chi$	0.74	0.04
	$^2\chi$	0.85	0.03
Polyethylene glycol			
	σ	0.37	0.37
	$^3\chi_p$	0.64	0.35
	W	0.67	0.07
	$^1\chi$	0.70	0.19
	$^2\chi$	0.72	0.138
	$^0\chi$	0.83	0.125

[2] Apple II computer.

Table III—Correlation Coefficients of the Stepwise Regression Equation for the Data Set Without Outliers

Stationary Phase	Parameter	Multiple Correlation	Simple Correlation
Nonpolar			
Squalane[a]	$^1\chi$	0.94	0.94
	π	0.97	0.50
	σ	0.97	0.14
	Z	0.98	0.77
SE-30[b]	$^1\chi$	0.91	0.91
	π	0.96	0.42
	σ	0.98	0.11
	Z	0.98	0.76
Polar			
Diethylene glycol succinate[c]	σ	0.45	0.45
	$^3\chi_p$	0.76	0.41
	Z	0.82	0.09
	$^1\chi$	0.87	0.26
	π	0.89	−0.34
	$^0\chi$	0.90	0.21
	$^2\chi$	0.97	0.18
Polyethylene glycol[c]	$^3\chi_p$	0.52	0.52
	σ	0.73	0.28
	Z	0.78	0.14
	$^1\chi$	0.83	0.36
	π	0.85	−0.20
	$^0\chi$	0.86	0.30
	$^2\chi$	0.95	0.29

[a] $N = 48$. [b] $N = 45$. [c] $N = 44$.

For the nonpolar stationary phases, the topological parameter $^1\chi$ describes best the behavior; this is in agreement with earlier results (5). For the more polar stationary phases, the most important parameter in the earlier study was σ, describing electronic effects (5). In the present study σ also appeared first (Table II). This is not surprising, since the interaction with the more polar stationary phases are much more specific. The second most important parameter was the $^3\chi_p$ term. This means that once the electronic effects are filtered out of the total variance, $^3\chi_p$ describes best the residual steric influences on the chromatographic behavior of the more polar stationary phases.

As stated above, the higher-order terms of the molecular connectivity parameters cannot be computed for all 100 substances in the original data set. To derive the effect of the inclusion of $^3\chi_p$ in the set variables, the multiple regression was carried out with and without $^3\chi_p$ [i.e., with the variables of our earlier study (5)]. It is clear that $^3\chi_p$ does indeed improve the correlation coefficient, since the correlation coefficient obtained with the two first variables introduced in the regression equation increased from 0.57 (without $^3\chi_p$) to 0.68 (with $^3\chi_p$) for diethylene glycol succinate. The results for the polar phases are improved by including the higher-order term $^3\chi_p$. They are, however, still worse than those for the nonpolar stationary phases.

To see which effects are not explained by the parameters included in the regression equation, the molecules whose chromatographic behavior is not well described with the regression equation, i.e., the molecules which have a large residual (e in Eq. 6), can be examined. The greatest residuals are found for molecules that contain more than one functional group, such as an hydroxyl and a keto group or two keto groups. The results of the regression equations without these outliers are shown in Table III. For the nonpolar stationary phases, the results are satisfactory. For the most polar stationary phase (diethylene glycol succinate), the most important parameter is still the electronic parameter σ, followed by $^3\chi_p$. For the mediumly polar polyethylene glycol, the first parameter entered is now $^3\chi_p$, followed by σ. In both cases the results are improved by omitting the outliers.

$^1\chi$ is entered as the first topological parameter in the regression equation for the apolar stationary phases, while $^3\chi_p$ is used for the polar stationary phases. More information about the intercorrelation between the parameters and the retention indices can be obtained by factor analysis. Factor analysis reduces the total amount of variables to a few fundamental variables (factors or latent variables) which are a linear combination of the original variables. These new variables are chosen so that as much of the total variance as possible is explained. Up to this point, the retention indices were considered as dependent variables and π, σ, and the topological parameters as independent variables. This means that the retention indices were predicted or explained by means

of the structural parameters. However, both the retention indices and the structural parameters can be thought of as variables containing information about the structural aspects and, therefore, about possible interactions of the molecules. Factor analysis can show which variables (structural parameters and retention indices) contain analogous information (i.e., are related to each other).

Factor analysis of the data set, without the outliers, resulted in the extraction of three important factors, explaining respectively 55, 20, and 13% of the total variance. Figure 2A shows that the topological parameters form a cluster near the cluster of the nonpolar stationary phases in the two-dimensional space of the first two factors. This cluster is rather distant from the cluster of the two polar stationary phases. The only parameter in the neighborhood of the polar phases is σ. These results are analogous to those of an earlier study (5). The topological parameters best describe the chromatographic behavior on the nonpolar stationary phases. In the plane of the first and third factor (Fig. 2B), however, the topological parameter $^3\chi_p$ is situated near the polar phases, which means that in the three-dimensional space the parameters σ and $^3\chi_p$ are most closely related to the polar stationary phases. The other topological parameters are still best related to the nonpolar phases. This shows that the steric effects determining the GC behavior are different for polar and nonpolar stationary phases. For the nonpolar phases, the lower-order connectivity terms are sufficient to account for these effects. For the more polar stationary phases, a better description of the molecules is necessary to describe the chromatographic behavior; this can be provided by means of the higher-order term $^3\chi_p$. It is, however, not the $^3\chi_c$ term that improves the regression equation. A reason for this could be the fact that $^3\chi_c$ describes only part of the molecule, while $^3\chi_p$ gives a more general idea of the whole molecule.

From these results it can be concluded that the GC behavior of molecules from different chemical families can be described satisfactorily by means of a general topological parameter on nonpolar stationary phases, since these interactions are quite nonspecific. For the more polar phases a higher-order term of the molecular connectivity, $^3\chi_p$, and an electronic parameter, σ, are required, since these interactions are quite specific (dipole–dipole interactions, hydrogen bonding, etc.). It should, however, be pointed out that these results are obtained for molecules containing only one functional group. The interpretation of the chromatographic behavior of molecules with more than one functional group is more difficult. A reason for this could be that the resultant electronic effect of two functional groups is not described sufficiently by σ. Electronic parameters, such as the dipole moment, could possibly improve these results. It is also known that two functional groups influence the partition coefficient of the molecules in a different way than each functional group separately (2, 14).

REFERENCES

(1) C. Hansch, "Drug Design," vol. 1, E. J. Ariens, Ed., Academic, New York, N.Y., 1971, chap. 2.

(2) Y. C. Martin, "Quantitative Drug Design: A Critical Introduction," Dekker, New York and Basel, 1978.

(3) L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Research," Academic, New York, N.Y., 1976.

(4) Y. Michotte and D. L. Massart, J. Pharm. Sci., **66**, 1630 (1977).

(5) L. Buydens and D. L. Massart, Anal. Chem., **53**, 1990 (1981).

(6) T. R. McGregor, J. Chromatogr. Sci., **17**, 314 (1979).

(7) L. B. Kier and L. H. Hall, J. Pharm. Sci., **68**, 120 (1979).

(8) J. S. Millership and A. D. Woolfson, J. Pharm. Pharmacol., **30**, 483 (1978).

(9) L. B. Kier and L. H. Hall, J. Pharm. Sci., **70**, 583 (1981).

(10) J. Shorter, "Correlation Analysis—Recent Advances," N. B. Chapman and J. Shorter, Eds., Plenum, New York, N.Y., 1978.

(11) G. L. Amidon and S. T. Anik, J. Pharm. Sci., **65**, 801 (1976).

(12) W. O. McReynolds, "Gas Chromatographic Retention Data," Preston Technical Abstracts Co., Illinois, 1966.

(13) N. H. Nie, C. H. Hull, J. G. Jenkins, K. Stenbrenner, and D. H. Bent, "SPSS-Statistical Package for the Social Sciences," 2nd ed., McGraw-Hill, New York, N.Y., 1975.

(14) R. F. Rekker, "Quantitative Structure-Activity Analysis," R. Franke and P. Oehme, Eds., Akademie-Verlag, Berlin, 1978.

ACKNOWLEDGMENTS